



Komisja
Europejska

PROJEKTOWANIE I ZLECANIE EWALUACJI WPŁYWU OPARTYCH NA METODACH KONTRFAKTYCZNYCH

PRAKTYCZNY PRZEWODNIK DLA INSTYTUCJI ZARZĄDZAJĄCYCH
EFS

Zatrudnienie,
sprawy
społeczne
i włączenie
społeczne

KOMISJA EUROPEJSKA

Dyrekcja Generalna ds. Zatrudnienia, Spraw Społecznych i Włączenia Społecznego
Dyrekcja G – Fundusze, programowanie i wdrażanie
Dział G5 – Lepsze stanowienie prawa

Dane kontaktowe: Linda Adamaite

E-mail: EMPL-G5-UNIT@ec.europa.eu lub Linda.Adamaite@ec.europa.eu

*Komisja Europejska
B-1049 Bruxelles/Brussel*

**PROJEKTOWANIE I ZLECANIE
EWALUACJI WPŁYWU OPARTYCH
NA METODACH
KONTRFAKTYCZNYCH**

PRAKTYCZNY PRZEWODNIK DLA INSTYTUCJI
ZARZĄDZAJĄCYCH EFS

Tekst ukończono we wrześniu 2021 r.

Wydanie przewodnika „Projektowanie i zlecenie ewaluacji wpływu opartych na metodach kontrfaktycznych – praktyczny przewodnik dla instytucji zarządzających EFS” z 2021 r. zostało opracowane przez Jochena Kluvego (Humboldt-Universität zu Berlin) oraz Andrzeja Naldiniego i Marca Pompilięgo (Ismeri Europa). Jest to adaptacja wydania wytycznych z 2013 r. (ISBN 978-92-79-28238-6; DOI 10.2767/94454) przygotowanych przez Stephena Morrisa (Jednostka ds. Oceny i Badań Polityki, Manchester Metropolitan University), Hertę Tödtling-Schönhofer (Metis GmbH, Wiedeń) i Michaela Wisemana (George Washington Institute of Public Policy).

Niniejszy dokument został przygotowany dla Komisji Europejskiej, jednak przedstawione w nim opinie stanowią wyłącznie punkt widzenia jego autorów, a Komisja Europejska nie ponosi odpowiedzialności za żadne skutki wynikające z ponownego wykorzystywania niniejszej publikacji. Więcej informacji o Unii Europejskiej można znaleźć w internecie (<https://ec.europa.eu>).

Luksemburg: Urząd Publikacji Unii Europejskiej, 2021 r.

© Unia Europejska, 2021



Ponowne wykorzystanie dokumentów Komisji Europejskiej odbywa się na podstawie decyzji Komisji 2011/833/UE z dnia 12 grudnia 2011 r. w sprawie ponownego wykorzystywania dokumentów Komisji (Dz.U. L 330 z 14.12.2011, s. 39). Z wyjątkiem przypadków, w których stwierdzono inaczej, ponowne wykorzystywanie niniejszego dokumentu jest dozwolone na podstawie licencji Creative Commons Attribution 4.0 International (CC-BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>). Oznacza to, że ponowne wykorzystywanie jest dozwolone pod warunkiem odpowiedniego poinformowania o autorze oraz o wszelkich zmianach wprowadzonych w dokumencie.

Wykorzystanie lub zwielokrotnienie elementów, które nie należą do Unii Europejskiej, może wymagać uzyskania zgody bezpośrednio od uprawnionych.

Okładka: © Shutterstock, 2021

WERSJA PDF ISBN 978-92-76-40725-6

doi:10. 10.2767/02762

KE-02-21-992-EN-N

Spis treści

WPROWADZENIE: KONTEKST I CEL PRZEWODNIKA.....	5
ROZDZIAŁ 1. KONCEPCJA I PODEJŚCIA.....	12
1.1. ISTOTA EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH	12
1.2. DLACZEGO EWALUACJE KONTRFAKTYCZNE SĄ WAŻNE?	13
1.3. DLACZEGO EWALUACJE KONTRFAKTYCZNE SĄ TRUDNE Z TECHNICZNEGO PUNKTU WIDZENIA?	14
1.4. PRZEGLĄD PROJEKTÓW I PODEJŚĆ W ZAKRESIE EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH	16
1.5. W JAKI SPOSÓB EWALUACJA WPŁYWU OPARTA NA METODACH KONTRFAKTYCZNYCH MOŻE BYĆ OSADZONA W SZERSZYCH RAMACH EWALUACJI	17
ROZDZIAŁ 2. KWESTIE PRAKTYCZNE PRZY OPRACOWYWANIU EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH.....	25
2.1. WYBÓR INTERWENCJI, KTÓRE MAJĄ BYĆ PODDANE EWALUACJI WPŁYWU	28
2.1.1. <i>Nadanie priorytetu interwencjom na potrzeby ewaluacji wpływu</i>	30
2.1.2. <i>Wybór interwencji, które są odpowiednie do podejścia kontrfaktycznego</i>	32
2.2. PYTANIA EWALUACYJNE I ZMIENNE WYNIKOWE	37
2.2.1. <i>Jakie są założenia i cele interwencji?</i>	37
2.2.2. <i>Jaki jest cel ewaluacji?</i>	37
2.3. DANE DO OKREŚLENIA GRUPY KONTROLNEJ I POMIARU ZMIENNYCH WYNIKOWYCH	42
2.3.1. <i>Czy odpowiednie dane są lub mogą być dostępne?</i>	42
2.3.2. <i>W jaki sposób należy określić grupę „poddaną interwencji”?</i>	56
2.3.3. <i>Czynniki, które należy uwzględnić przy identyfikacji grupy kontrolnej</i>	57
2.3.4. <i>Jakiego rodzaju kwestie dotyczące danych należy przedstawić w programie ewaluacji?</i>	61
2.3.5. <i>Jakie są kluczowe ograniczenia w analizie danych i rezultatów?</i>	63
2.3.6. <i>Lista kontrolna służąca weryfikacji przygotowania i wykonalności ewaluacji wpływu opartej na metodach kontrfaktycznych</i>	66
2.4. METODA KONTRFAKTYCZNEJ EWALUACJI WPŁYWU, KTÓRA MA BYĆ ZASTOSOWANA	67
2.5. HARMONOGRAM I BUDŻET	68
2.5.1. <i>Jakie zasoby są dostępne?</i>	68
2.5.2. <i>Kiedy należy dokonać ewaluacji interwencji?</i>	73
2.6. WDROŻENIE EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH.....	75
2.6.1. <i>Wybór ewaluatora</i>	76
2.6.2. <i>Nadzór nad ewaluacją wpływu opartą na metodach kontrfaktycznych</i>	77
2.6.3. <i>Sprawozdawczość</i>	78
2.6.4. <i>Korzystanie z wyników</i>	79
ROZDZIAŁ 3. JAK WYBRAĆ ODPOWIEDNIĄ METODĘ DO PRZEPROWADZENIA EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH.....	82
3.1. RANDOMIZACJA – PODEJŚCIE EKSPERYMENTALNE	82
3.2. PROJEKTY NIERANDOMIZOWANE LUB QUASI-EKSPERYMENTALNE	86
3.2.1. <i>Grupa docelowa i grupa kontrolna bez randomizacji</i>	86
3.2.2. <i>Propensity score matching</i>	87
3.2.3. <i>Różnica w różnicach</i>	90
3.2.4. <i>Model regresji nieciągłej</i>	93

3.2.5. Zmienne instrumentalne	96
ROZDZIAŁ 4. DALSZY PROMOWANIE EWALUACJI WPŁYWU OPARTYCH NA METODACH KONTRFAKTYCZNYCH.....	103
4.1. POPRAWA POZIOMU ZROZUMIENIA WŚRÓD ZAINTERESOWANYCH STRON.....	103
4.2. BUDOWANIE ZDOLNOŚCI	104
4.3. POKONYWANIE BARIER PRAWNYCH	107
4.4. W KIERUNKU BARDZIEJ PROSPEKTYWNYCH PODEJŚĆ	108
4.5. ROZSZERZENIE ZAKRESU EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH.....	109
GLOSARIUSZE.....	114
AKRONIMY	114
DEFINICJE	115
ZAŁĄCZNIKI	124
ZAŁĄCZNIK 1. DODATKOWE INFORMACJE.....	124
ZAŁĄCZNIK 2. SUGEROWANY PROGRAM KURSU NA TEMAT EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH	128
ZAŁĄCZNIK 3. EWALUACJA WPŁYWU OPARTA NA METODACH KONTRFAKTYCZNYCH – PRZYKŁADY PODANE W PRZEWODNIKU	129

SPIS RAMEK

RAMKA 1. PRZYKŁAD EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH SŁUŻĄCEJ PORÓWNANIU KORZYŚCI I KOSZTÓW	23
RAMKA 2. EWALUACJA WPŁYWU OPARTA NA METODACH KONTRFAKTYCZNYCH OSADZONA W SZERSZYCH RAMACH.....	28
RAMKA 3. PYTANIA UŁATWIAJĄCE WYBÓR INTERWENCJI, KTÓRE MAJĄ BYĆ PODDANE EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH	29
RAMKA 4. NAJCZĘSTSZE RODZAJE INTERWENCJI I GRUP DOCELOWYCH WYBIERANYCH NA POTRZEBY EWALUACJI WPŁYWU OPARTYCH NA METODACH KONTRFAKTYCZNYCH DOTYCZĄCYCH EFS.....	32
RAMKA 5. DEFINIOWANIE GRUP KONTROLNYCH.....	36
RAMKA 6. PRZYKŁADY DANYCH WYKORZYSTYWANYCH DO EWALUACJI WPŁYWU OPARTYCH NA METODACH KONTRFAKTYCZNYCH	45
RAMKA 7. PRZYKŁADY ZINTEGROWANYCH BAZ DANYCH NA POTRZEBY EWALUACJI WPŁYWU OPARTYCH NA METODACH KONTRFAKTYCZNYCH	46
RAMKA 8. UNIJNE RAMY REGULACYJNE DOTYCZĄCE PRZETWARZANIA DANYCH OSOBOWYCH.....	49
RAMKA 9. OCHRONA I WYMIANA DANYCH.....	55
RAMKA 10. ZAGADNIENIA POLITYCZNE ZWIĄZANE Z PROGRAMEM SZKOLENIOWYM	60
RAMKA 11. INTERPRETACJA EFEKTÓW NETTO.....	60
RAMKA 12. NIEPEWNOŚĆ DOTYCZĄCA INTERPRETACJI WYNIKÓW.....	65
RAMKA 13. DOŚWIADCZENIA POLSKI ZWIĄZANE Z KONFERENCJAMI EWALUACYJNYMI	81
RAMKA 14. PRZYKŁAD BADANIA RANDOMIZOWANEGO PROJEKTU EFS DOTYCZĄCEGO MŁODZIEŻY	84
RAMKA 15. PRZYKŁAD EWALUACJI, W KTÓREJ ZASTOSOWANO PODEJŚCIE OPARTE NA PAROWANIU	89
RAMKA 16. PRZYKŁAD EWALUACJI, W KTÓREJ ZASTOSOWANO PODEJŚCIE OPARTE NA RÓŻNICACH.....	92
RAMKA 17. PRZYKŁAD EWALUACJI, W KTÓREJ ZASTOSOWANO PODEJŚCIE OPARTE NA REGRESJI NIECIĄGŁEJ	96
RAMKA 18. PRZYKŁAD BADANIA, W KTÓRYM PRZYJĘTO PODEJŚCIE OPARTE NA ZMIENNYCH INSTRUMENTALNYCH	98
RAMKA 19. PRZYKŁAD PROJEKTU MAJĄCEGO NA CELU WZMOCNIENIE KULTURY I ZDOLNOŚCI W ZAKRESIE EWALUACJI WPŁYWU OPARTEJ NA METODACH KONTRFAKTYCZNYCH	106
RAMKA 20. PRZYKŁADY EWALUACJI W OBSZARZE EDUKACJI	110
RAMKA 21. PRZYKŁAD OCENY WPŁYWU NA „MIĘKKIE WYNIKI” W NIEMCZACH.....	112

SPIS WYKRESÓW

WYKRES 1. RÓŻNE ZADANIA I RODZAJE EWALUACJI.....	19
WYKRES 2. ZOBRAZOWANIE PODEJŚCIA OPARTEGO NA MODELU MATRYCY LOGICZNEJ LUB „TEORII ZMIANY”	20
WYKRES 3. ZASADNICZA KOLEJNOŚĆ DZIAŁAŃ W RAMACH CIE	26
WYKRES 4. MINIMALNE WYKRYWALNE EFEKTY W PRÓBACH O RÓŻNEJ LICZEBNOŚCI	64
WYKRES 5. UPROSZCZONY HARMONOGRAM DOTYCZĄCY REZULTATÓW PROGRAMU SZKOLENIOWEGO	74
WYKRES 6. DWUGRUPOWA KONTROLA RANDOMIZOWANA	83
WYKRES 7. STYLIZOWANY PROJEKT QUASI-EKSPERYMENTALNY Z GRUPĄ OBJĘTĄ INTERWENCJĄ I GRUPĄ KONTROLNĄ	87
WYKRES 8. ILUSTRACJA PODEJŚCIA OPARTEGO NA WARTOŚCIACH <i>PROPENSITY SCORE</i>	89
WYKRES 9. ILUSTRACJA PODEJŚCIA OPARTEGO NA RÓŻNICACH	92
WYKRES 10. ILUSTRACJA PODEJŚCIA OPARTEGO NA REGRESJI NIECIĄGŁEJ	94
WYKRES 11. ILUSTRACJA PODEJŚCIA OPARTEGO NA ZMIENNYCH INSTRUMENTALNYCH	97

SPIS TABEL

TABELA 1. ZALECANA TREŚĆ PROGRAMU EWALUACJI	27
TABELA 2. RODZAJE I ŹRÓDŁA DANYCH.....	44
TABELA 3. STRUKTURA KOSZTÓW PODSTAWOWYCH CIE	72
TABELA 4. PODSTAWOWE INFORMACJE, KTÓRE NALEŻY ZAMIEŚCIĆ W DOKUMENTACJI PRZEDSTAWIAJĄCEJ CIE	79
TABELA 5. PORÓWNANIE NAJWAŻNIEJSZYCH CECH GŁÓWNYCH PODEJŚĆ DO CIE	100
TABELA 6. CHARAKTERYSTYKA CIE NA PODSTAWIE PRZYKŁADÓW PODANYCH W NINIEJSZYM PRZEWODNIKU.....	129

Wprowadzenie: kontekst i cel przewodnika

Okres programowania 2021–2027 rozpoczyna się dramatycznymi doświadczeniami związanymi z pandemią COVID-19 i spowodowanym nią kryzysem gospodarczym. Unia Europejska (UE) zwiększyła swój wkład finansowy i swoje działania polityczne na rzecz odbudowy gospodarek krajowych i zwiększenia zatrudnienia. W tych ramach Europejski Fundusz Społeczny Plus (EFS+) odgrywa kluczową rolę w zapewnianiu powszechnej pomocy bezrobotnym, z naciskiem na osoby młode i kobiety, a także we wspieraniu interwencji mających na celu zwalczanie ubóstwa dzieci, promowanie lepszego kształcenia i włączenia społecznego słabszych grup społecznych w całej UE. Różnorodne cele realizowane za sprawą EFS+ i zapisana w nim potrzeba osiągnięcia szybkich rezultatów w obszarach zatrudnienia i włączenia społecznego wymagają efektywnej alokacji zasobów. Coraz większe znaczenie w kształtowaniu polityki mają podejścia oparte na dowodach, a podstawowym instrumentem kierowania polityką publiczną jest ewaluacja.

W okresie 2014–2020 Komisja Europejska (KE) wspierała zdolności ewaluacyjne państw członkowskich i instytucji zarządzających (IZ) oraz promowała stosowanie ewaluacji wpływu opartej na metodach kontrfaktycznych (CIE). Pozytywne wyniki stwierdzone w ewaluacji wpływu opartej na metodach kontrfaktycznych są namacalnym dowodem na to, że efekty EFS wykraczają poza to, co można by osiągnąć w przypadku braku funduszu. Znaczną liczbę ewaluacji wpływu opartych na metodach kontrfaktycznych zaplanowano początkowo w planach ewaluacji, a następnie przeprowadzono podczas wdrażania programów EFS. W wielu przypadkach podczas przygotowania i realizacji ewaluacji wpływu opartych na metodach kontrfaktycznych napotkano trudności lub ewaluacje te pozostawały odosobnionymi doświadczeniami i nie zostały uwzględnione w ramach systematycznej ewaluacji.

Niniejszy przewodnik jest przeznaczony dla instytucji zarządzających (IZ) i innych organów odpowiedzialnych za wdrażanie interwencji i programów finansowanych z EFS+ i ma na celu pomoc w planowaniu, projektowaniu i zleceniu ewaluacji wpływu opartych na metodach kontrfaktycznych. Uwzględniono w nim wcześniejsze doświadczenia i przedstawiono praktyczne porady dotyczące niektórych kluczowych kwestii, które należy wziąć pod uwagę przy opracowywaniu ewaluacji wpływu opartej na metodach kontrfaktycznych. Przewodnik stanowi aktualizację poprzednich wytycznych na lata 2014–2020 i jest w większym stopniu ukierunkowany na omówienie kwestii problematycznych, jakie pojawiały się podczas realizacji ewaluacji wpływu opartej na metodach kontrfaktycznych w praktyce. Niemniej jednak omówiono w nim aspekty metodologiczne, które również w miarę możliwości uproszczono i uzupełniono zaktualizowanymi przykładami wybranych z szeregu ewaluacji dotyczących EFS z lat 2014–2020.

Ewaluacje wpływu oparte na metodach kontrfaktycznych pozwalają odpowiedzieć na kluczowe pytania, które umożliwiają podejmowanie decyzji politycznych opartych na dowodach, mianowicie: jakie są efekty

Przewodnik dla instytucji zarządzających opisujący zaktualizowane doświadczenia z lat 2014–2020

CIE odpowiada na pytanie „co działa?”

przyczynowe interwencji i „co działa?”. Za ich pomocą można zdobyć dowody, które pozwolą sprawdzić, czy interwencje finansowane z EFS rzeczywiście przyczyniają się do zmiany sytuacji uczestników i pozwalają uzyskać planowane osiągnięcia. Dobrze wykonane ewaluacje wpływu oparte na metodach kontrfaktycznych dostarczają dowodów na „efekt netto”¹ lub wpływ interwencji, umożliwiając decydom wykluczenie alternatywnych wyjaśnień zmian sytuacji uczestników lub osiągnięć, które mogły zostać zaobserwowane. Gdy ewaluacje wpływu oparte na metodach kontrfaktycznych dostarczają oszacowania obecności i wielkości efektu netto, z konieczności zawierają one miarę niepewności, w zależności od poprawności metodologicznej i dostępnych informacji. Rodzaj dowodów dostarczanych przez ewaluacje wpływu oparte na metodach kontrfaktycznych umożliwia decydom ocenę skuteczności interwencji, dokonywanie porównań między interwencjami i ocenę ich względnych wyników. Ponadto dowody te dostarczają ważnych danych na potrzeby analizy kosztów i korzyści lub analizy efektywności kosztowej.

Przewodnik został opublikowany w czasie bezprecedensowych wyzwań dla EFS+. Biorąc pod uwagę ogromny wzrost środków UE na inwestycje i zatrudnienie przewidziany w pakiecie Next Generation EU, niezwykle ważne jest, aby decydenci dokonywali pomiaru efektów interwencji, za które odpowiadają, i starali się je zrozumieć. Fundusze publiczne należy przeznaczyć na bardziej produktywnie i skuteczne interwencje w celu przyspieszenia odbudowy i zmniejszenia nierówności społecznych. W związku z tym obowiązkiem osób odpowiedzialnych za wypłacanie środków z EFS+ jest uzasadnienie swoich wyborów poprzez wykazanie, że wybrane przez nich interwencje są skuteczne i zapewniają wartość dla obywateli. Najlepszym sposobem na osiągnięcie tego jest przeprowadzenie większej liczby wysokiej jakości ewaluacji wpływu opartych na metodach kontrfaktycznych.

EFS+ to główny europejski instrument na rzecz wspierania zatrudnienia i włączenia społecznego. W okresie programowania 2014–2020 z EFS przeznaczono prawie 125 mld EUR na aktywną politykę rynku pracy, politykę edukacyjną i politykę włączenia społecznego realizowane za pośrednictwem programów operacyjnych (PO) w 28 państwach członkowskich. Jak określono w rozporządzeniu (UE) nr 1303/2013 w sprawie przepisów ogólnych, ewaluacje „przeprowadza się w celu poprawy jakości projektowania i wdrażania programów, jak również w celu analizy ich skuteczności, efektywności i wpływu”.

W okresie programowania 2021–2027 w dalszym ciągu prowadzona będzie analiza wykonania i rezultatów². Będzie to wymagało wzmocnienia obecnych systemów oraz zdolności monitorowania i ewaluacji, w tym ustaleń dotyczących gromadzenia danych. Plany ewaluacji pozostaną obowiązkowe, a dodatkowy nacisk zostanie położony na ewaluację wpływu. Ponieważ dostępnych jest wiele metod pomiaru wpływu operacji

Zorientowanie na rezultaty i wysokiej jakości ewaluację.

¹ Efekt netto lub wpływ interwencji to różnica między wynikiem całkowitym (lub brutto) a tym, co zostałoby osiągnięte w przypadku braku interwencji. Efekt netto może być również ujemny, gdy interwencja jest mniej skuteczna niż dynamika rynku. Więcej informacji na ten temat można znaleźć na s. 9.

² ROZPORZĄDZENIE PARLAMENTU EUROPEJSKIEGO I RADY (UE) 2021/1060 z dnia 24 czerwca 2021 r. ustanawiające wspólne przepisy dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego Plus, Funduszu Spójności, Funduszu na rzecz Sprawiedliwej Transformacji i Europejskiego Funduszu Morskiego, Rybackiego i Akwakultury, a także przepisy finansowe na potrzeby tych funduszy oraz na potrzeby Funduszu Azylu, Migracji i Integracji, Funduszu Bezpieczeństwa Wewnętrznego i Instrumentu Wsparcia Finansowego na rzecz Zarządzania Granicami i Polityki Wizowej.

wspieranych z EFS+, instytucje zarządzające muszą zdecydować, która metoda lub jaka kombinacja metod jest najodpowiedniejsza do spełnienia wymogów regulacyjnych. Rygorystyczna kwantyfikacja wpływu interwencji obejmuje również scenariusze kontrfaktyczne.

Ważną cechą nowych przepisów jest koncentracja na dobrym wykonaniu i zorientowanie na rezultaty. Strategie i techniki przeprowadzania wysokiej jakości ewaluacji są niezbędne do zdobycia zasadniczej wiedzy umożliwiającej wykazanie wszystkim państwom członkowskim, które interwencje „działają”, a które nie. Niezbędne jest zwiększenie jakości ewaluacji i opracowanie wiarygodnych dowodów potwierdzających zapewnienie wartości dodanej.

Zasadniczo rozpoczęcie gromadzenia dowodów potwierdzających skuteczność interwencji politycznych jest proste. Wymogi obejmują:

- identyfikację problemu, który należy rozwiązać;
- identyfikację instrumentów, które należy zastosować w celu rozwiązania problemu;

wzór, który łączy instrumenty z rezultatami.

Dla ewaluacji systemu lub instrumentu finansowania podstawowe znaczenie ma stosowanie jasnych i wymiernych wskaźników dotyczących zarówno wkładu interwencji, jak i produktów i rezultatów związanych z danym problemem. Powszechną praktyką jest wyznaczanie celów zarówno w odniesieniu do produktów, jak i rezultatów oraz porównywanie rzeczywistych osiągnięć z wartościami docelowymi. W celu obserwacji wkładu interwencji, jej produktów i rezultatów w czasie oraz pozyskania informacji zwrotnych na temat zarządzania stosuje się monitorowanie. Leżąca u podstaw tych praktyk teoria interwencji często wskazuje na rezultaty pośrednie, które mogą również stać się przedmiotem monitorowania. To jednak dopiero początek – ustalenie, czy dana interwencja działa, nie jest proste.

Potrzeba rezultatów...

Konieczne jest uzupełnienie obecnej praktyki w zakresie ewaluacji podejściami, które generują znacznie mocniejsze dowody potwierdzające efekt netto lub wpływ interwencji. Mierzenie osiągnięć jest kwestią odpowiedzialności za wykorzystane fundusze. Ewaluacja wpływu oparta na metodach kontrfaktycznych pozwala udzielić odpowiedzi na fundamentalne pytanie o to, czy interwencja jest skuteczna. Choć ewaluacja wpływu oparta na metodach kontrfaktycznych polega na ustaleniu związku przyczynowego między interwencjami a rezultatami, konieczne może być zastosowanie również ewaluacji opartej na teorii i ewaluacji procesu w celu zidentyfikowania podstawowych mechanizmów przyczynowo-skutkowych i zapewnienia, aby wpływ osiągnięty w jednym miejscu stanowił podstawę dowodową do tego, by powielić daną politykę w innym miejscu.

...i dowodów potwierdzających efekt netto

W okresie programowania 2014–2020 wszystkie państwa członkowskie i instytucje zarządzające przyjęły plan ewaluacji opisujący cele, działania, elementy organizacyjne i produkty ewaluacji. W planach ewaluacji przewidziano kilka rodzajów ewaluacji, takich jak bieżące ogólne ewaluacje programów, a także ewaluacje tematyczne mające na celu udzielenie odpowiedzi na konkretne pytania ewaluacyjne lub skupienie się na konkretnych osiach lub priorytetach inwestycyjnych PO. Analiza planów ewaluacji dotyczących 177 programów EFS na koniec 2018 r.³ wykazała, że

Doświadczenia państw członkowskich z CIE

³ Zob. [Ismeri Europa – Ecorys – Institute for Employment Studies, 2019.](#)

w tym okresie programowania oczekuje się przeprowadzenia około 132 ewaluacji wpływu opartych na metodach kontryfaktycznych, co wskazuje na postęp i większy nacisk na metody kontryfaktyczne w okresie programowania 2014–2020 w porównaniu z okresem 2007–2013⁴.

Według repozytorium projektów Punktu Informacyjnego ds. Oceny⁵ do czerwca 2021 r. zakończono 1 795 ewaluacji, z czego 1 001 dotyczyło programów w ramach EFS/Inicjatywy na rzecz zatrudnienia ludzi młodych (YEI) lub EFS/YEI i EFRR (odpowiednio 675 i 326). Spośród 1 001 ewaluacji obejmujących EFS 323 stanowiły ewaluacje wpływu, a w przypadku 106 z nich zastosowano podejście kontryfaktyczne: było tak w przypadku 82 z 234 ewaluacji wpływu dotyczących programów EFS/YEI i w przypadku 24 z 89 ewaluacji wpływu dotyczących programów EFS/YEI i EFRR⁶.

Coraz większe wykorzystanie CIE

W pierwszych latach obecnego okresu programowania ewaluacje kontryfaktyczne programów EFS/YEI lub EFRR koncentrowały się na poprzednim okresie programowania, natomiast od 2018 r. prawie wszystkie ewaluacje wpływu oparte na metodach kontryfaktycznych dotyczą oceny efektów interwencji z lat 2014–2020⁷. Z 27 państw członkowskich 20 przeprowadziło ewaluacje kontryfaktyczne programów, w tym EFS, co również świadczy o większym rozpowszechnieniu tego podejścia w porównaniu z przeszłością.

Chociaż podejścia metodologiczne do ewaluacji wpływu opartych na metodach kontryfaktycznych różnią się w poszczególnych państwach członkowskich, podejście kontryfaktyczne było najczęściej wdrażane w celu oceny wpływu aktywnej polityki rynku pracy (szkolenia, środki zachęty, pomoc w poszukiwaniu pracy, krótkie praktyki zawodowe), ale w znacznie mniejszym stopniu w innych dziedzinach, jak np. interwencje mające na celu wsparcie instytucji oświatowych i uczniów, oraz w odniesieniu do polityki związanej z kwestiami społecznymi, jak np. interwencje mające na celu walkę z ubóstwem lub wykluczeniem społecznym.

Pod względem metodologicznym najczęściej stosowana jest metoda *propensity score matching*, natomiast inne metody są mniej powszechne. Ewaluacje wpływu oparte na metodach kontryfaktycznych często skupiają się na efektach krótko- lub średnioterminowych (6 lub 12 miesięcy), natomiast efekt długoterminowy jest rzadziej oceniany.

Analiza ewaluacji wpływu opartych na metodach kontryfaktycznych przeprowadzona w ramach projektu Punktu Informacyjnego ds. Oceny wykazała następujące ograniczenia: osoby w grupach kontrolnych nie były wystarczająco podobne do osób z grup objętych wsparciem, zaobserwowano niską wiarygodność danych, na przykład pochodzących z badań *ex post*, przeprowadzono wywiady wymagające od respondentów dokładnego przywoływania zdarzeń, przyjęto niewielką liczebność próby dającą statystycznie nieistotne wyniki itp. W niektórych przypadkach sprawozdania z ewaluacji nie zawierają wystarczających informacji na temat etapów, wyborów i ograniczeń metodologicznych, które to informacje są

Wciąż jednak występują pewne ograniczenia

⁴ Zob. [Bratu C. i in., 2014](#).

⁵ W ramach projektu Punktu Informacyjnego ds. Oceny, finansowanego przez DG REGIO i DG EMPL, gromadzone są informacje na temat ewaluacji publikowanych w internecie od dnia 1 stycznia 2015 r. na stronach internetowych IZ. Identyfikację ewaluacji przeprowadza sieć ekspertów krajowych i są one podsumowywane i oceniane pod względem jakości i wiarygodności ustaleń. https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/

⁶ 90 ze 106 ewaluacji zostało przeanalizowanych i podsumowanych w ramach projektu Punktu Informacyjnego ds. Oceny.

⁷ 29 z 33 ewaluacji wpływu opartych na metodach kontryfaktycznych przeprowadzonych do 2018 r. dotyczy poprzedniego okresu programowania.

niezbędne do tego, by móc informować innych – a nie tylko osoby, które zleciły sprawozdania – o efektach zbadanych środków i tym samym zapewnić im szerszą wiedzę na temat efektów.

Podsumowując, mimo postępów poczynionych w ostatnich latach przeprowadzenie ewaluacji wpływu opartych na metodach kontrfaktycznych wymaga fachowej wiedzy technicznej i woli politycznej. W niniejszym przewodniku przedstawiono argumenty przemawiające za stosowaniem ewaluacji wpływu opartymi na metodach kontrfaktycznych i określono niektóre kwestie, którymi IZ muszą się zająć, aby zapewnić udane wdrożenie tych ewaluacji. Poza praktycznymi aspektami ewaluacji wpływu opartych na metodach kontrfaktycznych w przewodniku skoncentrowano się również na szerszych kwestiach, którymi być może trzeba będzie się zająć, aby zapewnić wyższą jakość ewaluacji wpływu. Przewodnik zawiera szereg podstawowych zaleceń, które – mamy nadzieję – będą pomocne dla IZ, ale jednym z głównych założeń jest pokazanie, że przeprowadzanie ewaluacji wpływu opartych na metodach kontrfaktycznych wymaga starannego planowania procesu przygotowywania danych (szczególnie danych administracyjnych), określania jasnych założeń i celów itp., aby uniknąć potencjalnych problemów na etapie wdrażania lub niskiej jakości ewaluacji.

Przewodnik dla praktyków

Przewodnik zawiera również przykłady ewaluacji i praktyk, aby ułatwić czytelnikowi zrozumienie całego procesu. Chociaż głównym źródłem informacji na potrzeby omówienia przykładów i praktyk przedstawionych w niniejszym przewodniku są ewaluacje zebrane w ramach projektu Punktu Informacyjnego ds. Oceny, wzięto również pod uwagę inne źródła, w szczególności ewaluacje przeprowadzone przez Wspólne Centrum Badawcze (Centrum Badań nad Ewaluacją Wpływu (CRIE)) oraz, w mniejszym stopniu, publikacje naukowe. W miarę możliwości przykłady i doświadczenia opisane w przewodniku odnoszą się do interwencji w ramach Europejskiego Funduszu Społecznego.

Przewodnik składa się z czterech rozdziałów.

W **rozdziale 1** omówiono charakter ewaluacji wpływu opartych na metodach kontrfaktycznych i wyjaśniono, dlaczego są one ważne. Rozdział ten zawiera wstępny przegląd podejść polegających na ewaluacji wpływu opartej na metodach kontrfaktycznych, z naciskiem na rozróżnienie między podejściami eksperymentalnymi i quasi-eksperymentalnymi. Przedstawiono w nim również techniczne i praktyczne problemy, z którymi należy się zmierzyć podczas prowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych. Omówiono w nim też, w ujęciu ogólnym, rodzaje zagadnień politycznych, które mogą być uwzględnione w ewaluacji wpływu opartej na metodach kontrfaktycznych, oraz związek między metodami kontrfaktycznej ewaluacji wpływu a innymi podejściami do ewaluacji (na przykład: podejścia oparte na teorii, ewaluacja procesu i analiza efektywności).

Struktura odzwierciedlająca etapy ewaluacji

W **rozdziale 2** omówiono szereg kwestii, które IZ powinny rozważyć podczas projektowania i wdrażania ewaluacji wpływu opartych na metodach kontrfaktycznych. W przewodniku opisano niektóre z kluczowych wyzwań, jakie często stają przed osobami opracowującymi ewaluacje wpływu oparte na metodach kontrfaktycznych, oraz przedstawiono zalecenia dotyczące sposobów sprostania tym wyzwaniom. Omówione zagadnienia posłużą jako wskazówki dla osób, które będą zlecać oparte na metodach kontrfaktycznych ewaluacje wpływu interwencji finansowanych z EFS.

Ewaluacje wpływu oparte na metodach kontrfaktycznych dostarczają solidnych dowodów skuteczności funduszy, ale tylko pod warunkiem, że są dobrze zaplanowane i odpowiednio wykonane. W tym celu przy zleceniu ewaluacji IZ muszą zająć się pewnymi kluczowymi kwestiami. To, w jaki sposób i w jakiej kolejności dokładnie IZ będą rozpatrywać te kwestie, będzie podyktowane aspektami praktycznymi i rozwiązaniami instytucjonalnymi na miejscu w państwie członkowskim. W niniejszym przewodniku podkreślono niektóre z tych ważnych kwestii i zwrócono na nie uwagę IZ. Główne kwestie wymieniono zgodnie z sześcioma etapami procesu ewaluacji: 1) wybór operacji, które mają być ocenione; 2) określenie pytań ewaluacyjnych i zmiennych wynikowych; 3) analiza danych w celu określenia grupy kontrolnej i pomiaru zmiennych wynikowych; 4) wybór metody kontrfaktycznej ewaluacji wpływu; 5) określenie harmonogramu i budżetu; 6) wdrożenie ewaluacji. Szczególną uwagę poświęcono dostępności i ochronie danych, która to kwestia może poważnie ograniczyć możliwość przeprowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych.

W rozdziale 3 omówiono etap 4, wymieniony powyżej (wybór metody kontrfaktycznej ewaluacji wpływu) i skoncentrowano się na charakterystyce głównych metod stosowanych w analizach kontrfaktycznych. W szczególności przedstawiono pięć metod: metodę eksperymentalną lub randomizację, metodę *propensity score matching*, metodę różnicy w różnicach, model regresji nieciągłej i zmienne instrumentalne. Celem tej sekcji przewodnika nie jest szkolenie czytelników w zakresie korzystania z metod ewaluacji wpływu opartej na metodach kontrfaktycznych, ale uświadomienie czytelnikom mocnych i słabych stron każdej metody oraz wskazanie, w jakiej sytuacji dana metoda jest bardziej odpowiednia niż inna.

Rozdział 4 dotyczy szerszych zagadnień związanych z rozwojem w dalszej perspektywie. Obejmują one potrzebę rozwijania zdolności do skutecznego przeprowadzania ewaluacji wpływu opartych na metodach kontrfaktycznych, zarówno w ramach IZ (decydenci i urzędnicy), jak i wśród badaczy, społeczności akademickich i firm konsultingowych państw członkowskich. W sekcji tej omówiono również potrzebę zmierzenia się z barierami prawnymi w zakresie dostępu do danych i aktualizacji ewaluacji wpływu opartych na metodach kontrfaktycznych dotyczących programów EFS+ w celu zapewnienia bardziej zaawansowanego projektowania umożliwiającego rozszerzenie zakresu polityki włączenia społecznego i polityki edukacyjnej, szacowania efektów „miękkich wyników” oraz wdrażania podejść polegających na metaewaluacji.

Podsumowując, w niniejszym przewodniku: 1) przedstawiono argumenty przemawiające za stosowaniem ewaluacji wpływu opartych na metodach kontrfaktycznych, 2) określono ważne etapy skutecznej realizacji ewaluacji wpływu opartych na metodach kontrfaktycznych oraz 3) określono ewaluacje wpływu oparte na metodach kontrfaktycznych jako istotną część krajobrazu EFS+. Nadrzędnym celem jest zwiększenie wkładu EFS+ w dobrobyt obywateli Europy.

Autorzy pragną podziękować za cenne wsparcie i pomoc udzielone przez członków Działu G5 – Lepsze Stanowienie Prawa DG ds. Zatrudnienia podczas przygotowania niniejszej publikacji, w szczególności Lindzie Adamaite, Jeannette Monier i Marii José Cueto Faus. Autorzy pragną również podziękować Andrei Pisanowi i Ernestowi Belisariowi za ich wkład

w sekcję poświęconą ochronie danych oraz Pawłowi Hessowi za jego wsparcie i wkład.

Rozdział 1. Koncepcja i podejścia

W tym rozdziale poruszono podstawowe kwestie dotyczące charakteru i celu podejścia kontrfaktycznego w ocenie efektu przyczynowego interwencji. Mówiąc konkretnie, wyjaśniono w nim istotę ewaluacji wpływu opartej na metodach kontrfaktycznych, szczególnie w odniesieniu do rodzajów operacji współfinansowanych z EFS+. Przeanalizowano w nim również związek między podejściami kontrfaktycznymi a innymi metodami ewaluacji oraz omówiono, dlaczego ewaluacje wpływu oparte na metodach kontrfaktycznych są ważne – szczególnie w obecnych czasach. Ponadto omówiono kwestie polityczne, które mogą być poruszone w ewaluacjach wpływu opartych na metodach kontrfaktycznych, jak również przedstawiono krótki przegląd niektórych głównych metod kontrfaktycznych istotnych dla oceny interwencji współfinansowanych z EFS+, które następnie szczegółowo opisano w rozdziale 3.

1.1. Istota ewaluacji wpływu opartej na metodach kontrfaktycznych

Ewaluacje wpływu oparte na metodach kontrfaktycznych mają pomóc w określeniu efektu netto lub wpływu interwencji. Wyróżnia je cel, jakim jest potwierdzenie, że dana interwencja przynosi określony rezultat, tzn. że konkretny rezultat można przypisać wyłącznie danej interwencji. W ramach ewaluacji wpływu opartych na metodach kontrfaktycznych osiąga się to dzięki wyodrębnieniu interwencji i wykluczeniu alternatywnych wyjaśnień zaobserwowanego rezultatu.

Wykluczenie alternatywnych wyjaśnień jest możliwe przede wszystkim dzięki zastosowaniu „scenariusza kontrfaktycznego”, czyli odpowiedzi na pytanie: „Co by się stało – pod względem interesujących nas wyników – gdyby interwencja nie została zrealizowana?”. Aby dobrze zrozumieć pojęcie scenariusza kontrfaktycznego i uprościć tę kwestię, można zastosować przykład osoby bezrobotnej uczestniczącej w programie szkoleniowym, którego celem jest zwiększenie zatrudnienia. Na potrzeby określenia wpływu szkolenia na daną osobę zgodnie z podejściem kontrfaktycznym przewiduje się wykorzystanie dwóch potencjalnych rezultatów⁸. Pierwszym z nich jest status zatrudnienia uczestnika szkolenia po odbyciu szkolenia. Jest to rezultat widoczny dla uczestnika szkolenia. Drugim potencjalnym rezultatem jest status zatrudnienia tego uczestnika szkolenia w przypadku, gdyby nie brał on udziału w programie szkoleniowym, przy założeniu, że wszystkie inne zmienne pozostają takie same. W tych okolicznościach ten drugi rezultat określa się mianem rezultatu kontrfaktycznego. Wpływ szkolenia na danego uczestnika szkolenia określa się jako różnicę między rezultatami zaobserwowanymi a kontrfaktycznymi. Jest to efekt lub wpływ przyczynowy netto szkolenia na daną osobę. Jediną różnicą między okolicznościami lub uwarunkowaniami, które były podstawą zaobserwowanych i kontrfaktycznych rezultatów, jest uczestnictwo danej osoby w szkoleniu. Dlatego każda różnica

*Scenariusz
kontrfaktyczny*

⁸ Bardziej szczegółowe omówienie modelu przyczynowości „potencjalnych wyników” można znaleźć w opracowaniu [Holland P., 1986](#).

między tymi dwoma rezultatami musi stanowić wpływ szkolenia na status zatrudnienia danej osoby.

W rzeczywistości nie obserwujemy i nie możemy zaobserwować rezultatów kontrfaktycznych dla osób, do których skierowana jest interwencja. Głównym celem ewaluacji wpływu opartej na metodach kontrfaktycznych jest jednak dostarczenie przekonujących szacunków dotyczących rezultatów kontrfaktycznych w odniesieniu do grup osób lub przedsiębiorstw, na które wpływ mają interwencje współfinansowane z EFS. Wpływ ten wyraża się zatem na przykład jako różnice w średnich wartościach lub proporcjach między średnimi wartościami zaobserwowanymi a „szacowanymi” wartościami kontrfaktycznymi. W większości sytuacji przeprowadzenie ewaluacji wpływu opartych na metodach kontrfaktycznych ma na celu porównanie rezultatów interwencji (środka lub instrumentu) dla tych podmiotów (osób, MŚP itp.), które z niej skorzystały, z sytuacją grupy nieobjętej interwencją. W terminologii ewaluacji wpływu opartej na metodach kontrfaktycznych wskazuje się grupę „poddaną interwencji” lub „objętą interwencją” jako osobną od grupy „kontrolnej”, która powinna być jak najbardziej podobna pod każdym względem (z wyjątkiem przeprowadzanych interwencji) do grupy poddanej interwencji. To właśnie na podstawie grupy kontrolnej uzyskuje się oszacowania rezultatów kontrfaktycznych, ze szczególnym uwzględnieniem różnic w cechach – zaobserwowanych i niezaobserwowanych – między tymi dwoma grupami. Możliwe jest również porównanie wielu różnych interwencji poprzez poddanie kwalifikujących się jednostek szeregowi wariantów interwencji (np. inne interwencje finansowane z EFS lub interwencje finansowane z innych źródeł), utworzenie wielu grup objętych interwencją i porównanie rezultatów tych poszczególnych grup lub porównanie ich z rezultatami grupy kontrolnej niepoddanej interwencji.

Definiowanie grup objętych interwencją i...

W przypadku gdy grupa kontrolna nie jest poddana żadnej interwencji, pytanie ewaluacyjne, na które należy odpowiedzieć, brzmi: „Jaki jest wpływ przeprowadzenia interwencji w stosunku do braku pomocy lub wsparcia?”. Z drugiej strony, gdy rezultaty interwencji będącej przedmiotem zainteresowania są porównywane z rezultatami innej interwencji, pytanie ewaluacyjne brzmi: „Jaki jest wpływ przeprowadzenia rozważanej interwencji w porównaniu z zastosowaniem dobrze określonego alternatywnego rozwiązania?”. W wielu przypadkach ewaluacja wpływu oparta na metodach kontrfaktycznych może być zaprojektowana tak, aby odpowiedzieć na jedno z tych podstawowych pytań. Decyzja dotycząca tego, na które pytanie odpowiedzieć, zależy od priorytetów decydentów i praktycznych ograniczeń projektowych.

... grup kontrolnych

W przypadkach, w których dokonuje się porównania między dwoma różnymi metodami interwencji, powinien istnieć między nimi jasno określony kontrast, co ma znaczenie z punktu widzenia kształtowania polityki.

1.2. Dlaczego ewaluacje kontrfaktyczne są ważne?

Przeprowadzenie ewaluacji wpływu opartych na metodach kontrfaktycznych dostarcza ważnych informacji dotyczących efektów netto lub wpływu interwencji. Takie ewaluacje dostarczają szacunków dotyczących wielkości wpływu, ich znaku (pozytywnego lub negatywnego) oraz statystycznych miar niepewności. Pomagają potwierdzić lub odrzucić domniemany związek przyczynowy między interwencją a rezultatami, który leży u podstaw projektu interwencji. Te zmierzone efekty można wykorzystać do realizacji szeregu

celów: wykazania przejrzystości i rozliczalności w wydatkowaniu zasobów podatników; podejmowania decyzji politycznych w oparciu o jasne dowody; wyciągania wniosków na podstawie wyników odnotowanych w różnych regionach danego państwa lub w różnych państwach członkowskich.

Osobom odpowiedzialnym za interwencje i dbającym o to, aby ich programy nadal otrzymywały finansowanie, powinno zależeć na promowaniu ewaluacji wpływu opartych na metodach kontrfaktycznych w celu wykazania, że ich programy zapewniają odpowiedni stosunek wartości do ceny i przynoszą wymierne korzyści uczestnikom, a także całemu społeczeństwu. Dowody pozyskane w ramach ewaluacji wpływu opartych na metodach kontrfaktycznych będą miały szczególne znaczenie dla osób odpowiedzialnych za alokację zasobów. IZ chętnie wykażą, że ich programy rzeczywiście „działają”. Aby zrobić to w sposób przekonujący, IZ będą musiały zlecić wysokiej jakości ewaluację wpływu oparte na metodach kontrfaktycznych.

Przejrzystość i rozliczalność

Te cechy ewaluacji wpływu opartych na metodach kontrfaktycznych są źródłem ważnych informacji dla decydentów, których zadaniem jest alokacja zasobów do różnych interwencji. Decyzje dotyczące finansowania potencjalnych interwencji podejmowane są w kontekście limitów zasobów. W tym kontekście decydenci potrzebują solidnych dowodów potwierdzających wpływ programu i jego efektywność kosztową, aby mogli jak najlepiej wykorzystać dostępne zasoby. W ocenie względnej efektywności interwencji efekty netto oszacowane w ramach ewaluacji wpływu opartych na metodach kontrfaktycznych można przedstawić bardziej szczegółowo, badając efektywność kosztową interwencji lub przeprowadzając pełną analizę kosztów i korzyści.

Uzasadnienie decyzji o alokacji zasobów

Ważną dodatkową wiedzę można również uzyskać dzięki porównaniu efektów netto podobnych operacji realizowanych na różnych terytoriach. Jeżeli w wyniku ewaluacji wpływu opartych na metodach kontrfaktycznych wykryte zostaną znaczące różnice w efektach tych operacji, najprawdopodobniej oznacza to, że jakość operacji znacznie się różni lub że niektóre operacje bardziej pasują do danego kontekstu społeczno-gospodarczego niż inne. Porównania te – powszechnie nazywane „metaewaluacjami” lub „metaanalizami” – pozwalają na większe uogólnienie wyników ewaluacji wpływu opartych na metodach kontrfaktycznych i stanowią ważne badania w odniesieniu do środków z zakresu polityki (np. szkoleń i usług zintegrowanych itp.)⁹. Do przeprowadzenia tych porównań zgodnie z kryteriami naukowymi niezbędne są odpowiednia liczba ewaluacji wpływu opartych na metodach kontrfaktycznych i ich systemowe wdrożenie. W związku z tym podstawowym warunkiem częstszego stosowania i przydatności tych porównań przy podejmowaniu decyzji politycznych jest dobra koordynacja na poziomie krajowym lub unijnym.

Porównanie operacji w różnych regionach lub państwach członkowskich

1.3. Dlaczego ewaluacje kontrfaktyczne są trudne z technicznego punktu widzenia?

Istnieje wiele podejść, które można opisać jako „niemiarodajne” próby oszacowania wpływu interwencji. W niniejszym przewodniku omówiono je w celu zilustrowania złożoności nieodłącznie związanych z ewaluacjami wpływu

Wychodzenie poza uproszczone podejścia

⁹ Zob. na przykład [Card, D., Kluve, J. i Weber, A., 2017.](#)

opartymi na metodach konfaktycznych i nie stanowią odniesienia do rzeczywistej praktyki ewaluacyjnej.

Jednym z takich podejść jest sytuacja, w której decydent zamierza ocenić wpływ programu szkoleniowego dla bezrobotnych, porównując dochód uczestników szkolenia po szkoleniu z dochodem wszystkich bezrobotnych, którzy nie uczestniczyli w programie. Następnie decydent przypisuje zaobserwowaną różnicę w dochodach między uczestnikami programu a osobami nieuczestniczącymi w programie programowi szkoleniowemu.

Nie jest to poprawna strategia identyfikacji wpływu szkolenia na dochody, ponieważ osoby nieuczestniczące w szkoleniu mogą różnić się w istotny sposób od uczestników szkolenia, a różnice te mogą wpływać na rezultaty – dlatego takie podejście nie wyklucza alternatywnych wyjaśnień jakichkolwiek zaobserwowanych różnic w dochodach. Na przykład uczestnicy szkolenia mogą mieć większe wrodzone zdolności niż osoby nieuczestniczące w szkoleniu. Innymi słowy, do udziału w programie szkoleniowym zgłaszają się osoby bezrobotne o wyższych zdolnościach. Zdolność wpływa nie tylko na decyzję o uczestnictwie, ale także na rezultaty – osoby bezrobotne o wyższym poziomie wrodzonych zdolności częściej mają wyższe dochody niż osoby o niższych zdolnościach. W rezultacie jakakolwiek zaobserwowana różnica w dochodach między grupą objętą interwencją a grupą kontrolną może wynikać zarówno z programu szkoleniowego, jak i wrodzonych różnic w zdolnościach, a zatem osoba przyjmująca takie podejście może „przecenić” wpływ interwencji.

Jeśli zdolności nie można zmierzyć, a różnice we wrodzonych zdolnościach między dwiema grupami nie mogą być brane pod uwagę przy szacowaniu wpływu poprzez porównanie wyników dwóch grup, to uznaje się, że szacowany wpływ programu szkoleniowego jest obarczony **błędem selekcji**. Aby rozwiązać ten problem, ewaluatorzy starają się zebrać jak najwięcej informacji na temat ważnych czynników, które wpływają na decyzję o uczestnictwie i wynikające z tego rezultaty. Na podstawie tych danych wybiera się odpowiednią grupę kontrolną spośród osób nieuczestniczących – tj. grupę, która jest w możliwie największym stopniu podobna do grupy objętej interwencją pod względem tych czynników – i przeprowadza się analizy statystyczne uwzględniające te czynniki. Ewaluatorzy często przyjmują przy tym, że te możliwe do zaobserwowania czynniki decydują o wyborze uczestników programu. Takiego „założenia dotyczącego identyfikacji” zasadniczo nie da się łatwo uzasadnić i to ewaluator musi przedstawić przekonujące argumenty, świadczące o tym, że założenie to jest wiarygodne w danym kontekście, na podstawie znajomości czynników instytucjonalnych i teorii behawioralnej.

*Przeciwdziałanie
błędowi selekcji*

Inne „niemiarodajne” podejście może polegać na obserwowaniu przez decydenta dochodów uczestników przed szkoleniem i po szkoleniu, a następnie przypisaniu zmiany sytuacji przed i po w tym zakresie interwencji szkoleniowej. W istocie podejście to opiera się na założeniu, że przy braku interwencji średni dochód pozostaje niezmienny.

*Sytuacja „przed i
po”*

Również w tym przypadku prawie nigdy nie będzie to prawidłowa strategia mierzenia wpływu szkolenia na dochód, chyba że można w sposób uzasadniony założyć czasową stabilność. Wynika to z faktu, że dochody uczestników szkolenia nieuchronnie zmieniają się w czasie z powodów całkowicie niezwiązanych ze szkoleniem. Na przykład często można zaobserwować, że zarobki uczestników szkolenia spadają przed

uczestnictwem w szkoleniu, częściowo z powodu czynników przejściowych. W wielu przypadkach odbicie nastąpi niezależnie od interwencji szkoleniowej¹⁰. W takim niemiernym podejściu polegającym na ocenie wpływu szkolenia na podstawie różnicy między zarobkami bezpośrednio przed rozpoczęciem programu a zarobkami po jego zakończeniu lekceważy się fakt, że w wielu przypadkach zarobki i tak by wzrosły.

Aby dostosować takie projekty, wymagana jest miara scenariusza kontrfaktycznego – tj. miara tego, jak zmieniłyby się dochody uczestników szkolenia w przypadku braku interwencji szkoleniowej. Na przykład taki rezultat kontrfaktyczny można uzyskać na podstawie badania starannie dobranej grupy kontrolnej niepoddanej interwencji szkoleniowej, której dochody obserwuje się w tym samym przedziale czasowym co dochody uczestników szkolenia. Często przyjmuje się wówczas występowanie wspólnych tendencji, tj. zakłada się, że w przypadku braku interwencji tendencja w dochodach wśród uczestników szkolenia i członków grupy kontrolnej byłaby taka sama.

Ograniczenia tych „niemiernych” podejść skłaniają do poszukiwania bardziej przekonujących metod ewaluacji. Jak zasugerowano powyżej, bardziej przekonujące metody są jednak trudniejsze do wdrożenia z technicznego punktu widzenia. W dalszej sekcji tego rozdziału przedstawiono krótki zarys niektórych konkretnych podejść do ewaluacji wpływu opartej na metodach kontrfaktycznych, które prawdopodobnie będą miały największe znaczenie w kontekście EFS.

1.4. Przegląd projektów i podejść w zakresie ewaluacji wpływu opartej na metodach kontrfaktycznych

Koncentrując się na efektach interwencji dla uczestników, rezultaty kontrfaktyczne szacuje się zwykle na podstawie danych zgromadzonych od grup osób nieuczestniczących w interwencji, które są podobne do uczestników interwencji poddawanej ewaluacji. W tabeli 1 na końcu tego rozdziału przedstawiono krótki przegląd podejść, niektóre ich zalety i ograniczenia oraz podstawowe rodzaje danych, których podejścia te wymagają.

W przypadku projektów ewaluacji wpływu opartych na metodach kontrfaktycznych rozróżnia się przede wszystkim podejście eksperymentalne i quasi-eksperymentalne. Podejście eksperymentalne jest powszechnie określane jako „badanie randomizowane z grupą kontrolną” (ang. *randomised control trial*, RCT), a czasami także jako „eksperyment społeczny”.

Podejście eksperymentalne uważa się za złoty standard wśród metod kontrfaktycznych stosowanych do ewaluacji efektów interwencji. Podejście to można testować i modyfikować w stosunkowo krótkim czasie i w większości przypadków stanowi ono optymalne rozwiązanie. Dobry projekt ewaluacji wpływu powinien umożliwiać uzyskanie szacunkowych rezultatów kontrfaktycznych, które nie będą obciążone błędem systematycznym. W wielu zastosowaniach podejście eksperymentalne można uznać za dające takie szacunki nieobciążone błędem systematycznym. Przy omawianiu podejść do ewaluacji wpływu opartej na metodach kontrfaktycznych często

*Projektowanie
randomizowane –
złoty standard*

¹⁰ Zjawisko to powszechnie nazywa się „Ashenfelter's Dip” (spadek Ashenfeltera) od nazwiska ekonomisty, który jako pierwszy je omówił. Zob. [Ashenfelter O., 1978.](#)

pożądane jest rozpoczęcie od nakreślenia podejścia eksperymentalnego. Wynika to z tego, że metody quasi-eksperymentalne mają zasadniczo naśladować optymalne założenia eksperymentalne.

Podczas omawiania projektów ewaluacji wpływu opartej na metodach kontrfaktycznych kluczowe cechy każdego podejścia są przedstawione w możliwie najprostszy sposób, tak aby wyjaśnić podstawowe zasady. W rzeczywistości zastosowania tych metod mogą być znacznie bardziej złożone, a kwestie takie jak niezgodność – tj. sytuacja, gdy osoby wskazane do udziału w interwencji nie uczestniczyły w niej – mogą znacznie zwiększyć liczbę wyzwań, z jakimi trzeba się będzie zmierzyć.

Istnieje szeroki wachlarz podejść, w których zasadniczo dąży się do naśladowania procesu randomizacji. Określa się je mianem **podejść quasi-eksperymentalnych**. W niniejszym przewodniku nie sposób dokonać przeglądu wszystkich takich podejść ani przedstawić kompletnego szczegółowego opisu technicznego każdego z nich. W szerszym kontekście jednak metody quasi-eksperymentalne, które z największym prawdopodobieństwem mogą być wykorzystane w kontekście EFS, to: 1. metoda *propensity score matching*; 2. metoda różnicy w różnicach; 3. metoda regresji nieciągłej; 4. zmienne instrumentalne. Przegląd głównych podejść i ich względnych zalet przedstawiono w Tabeli 5 w rozdziale 3; ich prezentacja jest znacznie uproszczona w celu podkreślenia kluczowych zasad każdego podejścia. Wykaz literatury dotyczącej metod quasi-eksperymentalnych przedstawiono w załączniku 1.

*Projektowanie
nierandomizowane
lub quasi-
eksperymentalne*

1.5. W jaki sposób ewaluacja wpływu oparta na metodach kontrfaktycznych może być osadzona w szerszych ramach ewaluacji

Ewaluacje kontrfaktyczne pozwalają odpowiedzieć na pewne pytania dotyczące efektów przyczynowych interwencji. Podejścia te są ograniczone do zakresu, w jakim mogą umożliwić udzielenie odpowiedzi na inne pytania dotyczące interwencji. Pomocne jest rozróżnienie między pytaniami ewaluacyjnymi dotyczącymi **wyjaśnienia przyczynowego** a tymi dotyczącymi **opisu przyczynowego**. Ewaluacje wpływu oparte na metodach kontrfaktycznych mają na celu **opisanie** konsekwencji interwencji. Takie metody są mniej odpowiednie do **wyjaśniania** mechanizmów i kontekstów, dzięki którym powstają relacje przyczynowe. To rozróżnienie jest ważne, ponieważ pomaga wyjaśnić szczególną rolę ewaluacji wpływu opartej na metodach kontrfaktycznych¹¹.

*Wyjaśnienie
przyczynowe i opis
przyczynowy*

Dobrze zaprojektowana ewaluacja wpływu oparta na metodach kontrfaktycznych będzie dla decydenta źródłem informacji o tym, czy interwencja doprowadziła do zmiany rezultatów, do której została zaprojektowana. Dzięki niej decydent będzie mógł zebrać dowody na wielkość ewentualnego wpływu lub efektu, określić, czy wpływ był pozytywny czy negatywny, a także ustalić miarę niepewności. Z kolei jeśli chodzi o wyjaśnienie, dlaczego i w jaki sposób doszło do wpływu mierzonego za pomocą ewaluacji, ewaluacje wpływu oparte na metodach kontrfaktycznych nie są już tak skuteczne. W przypadku gdy interwencja nie przyniosła

*Czego decydenci
mogą się
dowiedzieć z CIE, a
czego nie mogą*

¹¹ Zob. Shadish W.R., Cook T.D. i Campbell D.T., 2002 oraz [Stern E. i in., 2012](#).

żadnego wpływu często ciężko jest wręcz ustalić powód takiej sytuacji na podstawie ewaluacji wpływu opartej na metodach kontrfaktycznych.

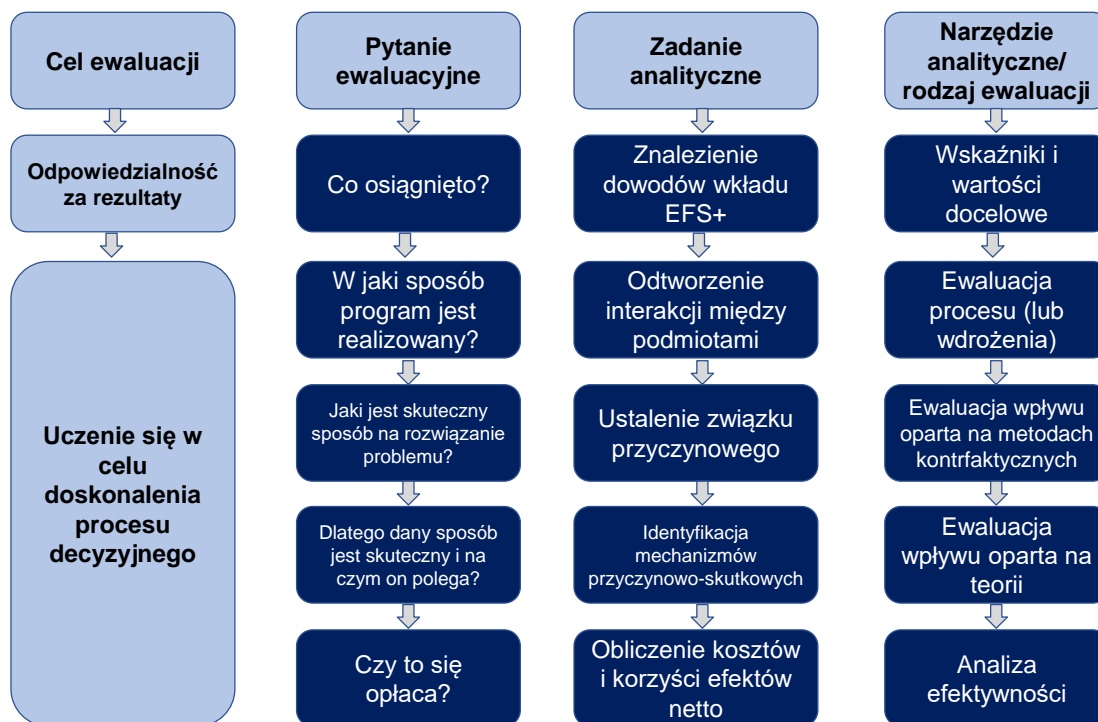
W większości organów decyzyjnych zainteresowane strony zadające pytania o opis przyczynowy i o wyjaśnienie przyczynowe kierują się zazwyczaj różnymi interesami i perspektywami. Zarządzający programami i praktycy zajmujący się programami mają tendencję do skupiania się na pytaniach o wyjaśnienie przyczynowe. Podmioty odpowiedzialne za przydzielanie zasobów i decydenci wyższego szczebla odpowiedzialni za ustalanie budżetu zwykle skupiają się na pytaniach o opis przyczynowy. W praktyce rozróżnienie między wyjaśnieniem przyczynowym a opisem przyczynowym może być dość niewyraźne. W niektórych okolicznościach ewaluacje wpływu oparte na metodach kontrfaktycznych mogą dostarczyć wyjaśnienia, z czego wynika określony wpływ, na przykład poprzez zbadanie wpływu interwencji na ważne podgrupy. Konieczne jest jednak uważne rozważenie rodzajów pytań, jakie stawiają zainteresowane strony w odniesieniu do interwencji, i wybranie odpowiedniego podejścia, aby odpowiedzieć na każde z nich. W przypadkach, w których podstawowym pytaniem jest, czy interwencja działa, w wielu okolicznościach ewaluacja wpływu oparta na metodach kontrfaktycznych będzie odpowiednim wyborem. W przypadkach, w których podstawowym pytaniem jest, jak interwencja działa, bardziej odpowiednie są metody ewaluacji opartej na teorii i ewaluacji procesu.

Te różne rodzaje pytań i celów podsumowano na poniższym wykresie.

Powyższe rozważania prowadzą do wniosku, że ewaluacje wpływu oparte na metodach kontrfaktycznych należy opracowywać w postaci planu ewaluacji. Taki plan ewaluacji musi obejmować różne formy ewaluacji ukierunkowane na znalezienie odpowiedzi na różne pytania zadawane przez różne zainteresowane strony w obszarze polityki. W praktyce plan ewaluacji rzadko kiedy, o ile w ogóle, będzie zawierał ewaluację wpływu opartą na metodach kontrfaktycznych bez ewaluacji procesu.

Z myślą o ewaluacji stosuje się szeroką gamę podejść, które służą wielu różnym celom. Kluczowym problemem jest to, w jaki sposób można połączyć te podejścia, tak by posłużyły do promowania uczenia się polityki. Łączenie w odpowiedni sposób różnych rodzajów ewaluacji – z różnymi celami w ramach cyklu programowania – jest prawdziwym wyzwaniem w tej dziedzinie. Jak już omówiono i jak pokazano na kolejnym wykresie, ewaluacja wpływu oparta na metodach kontrfaktycznych, ewaluacja procesu i podejścia oparte na teorii uzupełniają się nawzajem.

Wykres 1. Różne zadania i rodzaje ewaluacji



Źródło: na podstawie [Martini A., 2009.](#)

Solidna strategia ewaluacji powinna składać się z następujących elementów:

- ewaluacja wpływu oparta na teorii,
- ewaluacja procesu (lub wdrożenia),
- ewaluacja wpływu oparta na metodach kontrfaktycznych (CIE),
- analiza efektywności.

Łączenie rodzajów ewaluacji

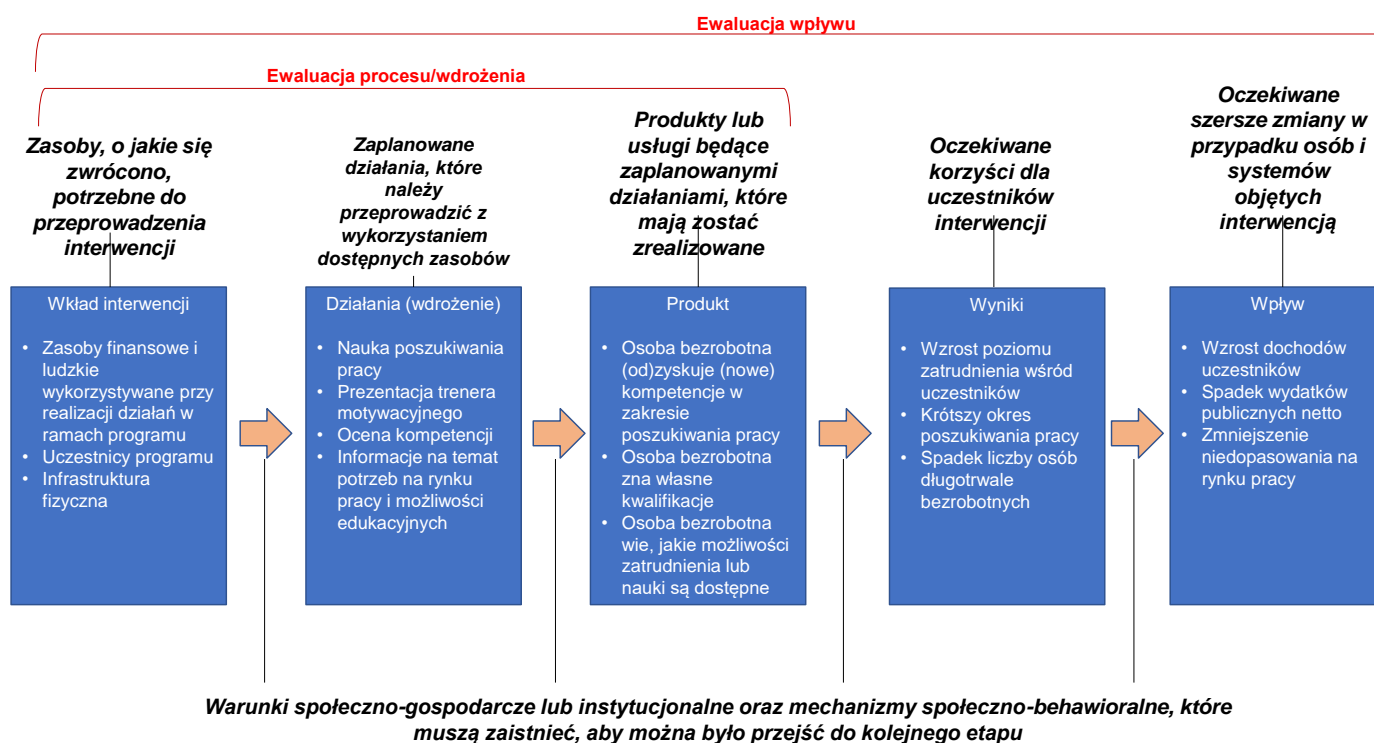
W niniejszym przewodniku szczegółowo omówione zostaną jedynie podejścia kontrfaktyczne do ewaluacji wpływu. W kontekście ewaluacji wpływu opartej na metodach kontrfaktycznych podejścia oparte na teorii są sposobem na zrozumienie założeń projektowych leżących u podstaw interwencji.

Ewaluacje oparte na teorii są stosowane w niektórych okolicznościach do opisanego zamierzonego działania interwencji i sprawdzenia, czy można zaobserwować przewidzianą w teorii interwencji zmianę w rezultatach. W tym sensie podejścia oparte na teorii można wykorzystać do oceny wpływu, aby udzielić odpowiedzi na pytanie „w jaki sposób” wygenerowano wpływ, ale mogą one również posłużyć do zbadania wpływu interwencji w przypadkach, gdy przeprowadzenie ewaluacji wpływu opartych na metodach kontrfaktycznych nie jest możliwe. Kolejny wykres przedstawia stylizowany przykład teorii zmiany dotyczący interwencji mającej na celu poprawę służb zatrudnienia. Ewaluacja oparta na teorii pozwala zbadać, czy dowody potwierdzają kolejność etapów przyczynowych, od nakładu do wpływu, zgodnie z założeniami teorii zmiany, a także ustalić, w jakich warunkach i za pomocą jakich mechanizmów społecznych było to możliwe. Szczegółowy opis wykorzystania podejść opartych na teorii do określania wpływu wykracza poza zakres tego dokumentu.

Ewaluacja oparta na teorii odnosi się do „teorii zmiany”

W kontekście ewaluacji wpływu opartej na metodach kontrfaktycznych ewaluacja oparta na teorii uwzględnia sposób planowania i projektowania interwencji oraz planowany sposób jej działania. Zasadniczo podejście to obejmuje współpracę z zainteresowanymi stronami interwencji nad opracowaniem wspólnego opisu „teorii zmiany” leżącej u podstaw interwencji, co przedstawiono w sposób uproszczony na następnym wykresie. Wszystkie interwencje realizowane są w oparciu o logikę programu, która łączy wkład interwencji i działania w jej ramach z jej produktami, rezultatami pośrednimi, a następnie rezultatami bardziej długoterminowymi. W konsekwencji sformułowanie teorii zmiany jest również ważnym wsparciem dla ewaluacji wpływu opartych na metodach kontrfaktycznych – ułatwia rekonstrukcję procesu wdrożenia i jego ewentualnego wpływu na produkty i wyniki, a także identyfikację najbardziej reprezentatywnych zmiennych wynikowych, które należy sprawdzić za pomocą analizy kontrfaktycznej.

Wykres 2. Zobrazowanie podejścia opartego na modelu matrycy logicznej lub „teorii zmiany”



Źródło: na podstawie [W.K. Kellog Foundation, 2004](#) i [Bredgaard T., 2015](#).

Ewaluacja oparta na teorii może łączyć się z ewaluacją wpływu opartą na metodach kontrfaktycznych na wiele użytecznych sposobów. Jasno sformułowana teoria zmiany (lub logika interwencji) może posłużyć do opracowania projektu ewaluacji wpływu opartej na metodach kontrfaktycznych. Dobrze wyrażona teoria zmiany może być dla osoby odpowiedzialnej za projektowanie ewaluacji wpływu źródłem informacji dotyczących między innymi następujących aspektów:

... i przydaje się do CIE

- Które rezultaty są ważne i wymagają pomiaru?
- Jaki może być prawdopodobny znak i wielkość wpływu interwencji?
- Kto jest zamierzoną grupą docelową i jak wybrać grupę kontrolną?

- Ile czasu potrzeba na urzeczywistnienie efektów i rezultatów programu?
- Jakie dane mogą być wymagane do pomiaru uczestnictwa w interwencji?
- Na ile wiarygodna jest grupa kontrolna jako miara scenariusza kontrfaktycznego?

Opracowanie teorii zmiany może również pomóc w identyfikacji potencjalnych nieoczekiwanych efektów, które można wziąć pod uwagę przy projektowaniu ewaluacji wpływu opartej na metodach kontrfaktycznych. W pewnym stopniu jasno sformułowana teoria zmiany może również pomóc ewaluatorowi zinterpretować rezultaty ewaluacji wpływu opartej na metodach kontrfaktycznych. Jednak jeśli chodzi o interpretację, wielu informacji może też dostarczyć **ewaluacja procesu**.

Ewaluacja procesu

Ewaluacja procesu w kontekście ewaluacji wpływu opartej na metodach kontrfaktycznych ma dwa cele. Pierwszy to ocena „wierności”, drugi to ocena różnicy między doświadczeniami osób należących do grupy objętej interwencją i grupy kontrolnej.

Za pomocą oceny „wierności” bada się, w jakim stopniu interwencja – w formie, w jakiej została zrealizowana – jest zgodna ze swoim projektem. Ewaluacja procesu polega na rozważeniu, jakie usługi faktycznie udostępniono uczestnikom interwencji. Czy odpowiadają one zamierzeniom teorii interwencji? W przypadku występowania różnic pod względem realizacji w różnych miejscach, z czego takie różnice wynikają? Większość interwencji uwzględnia zarówno logikę zarządzania, jak i efektu:

Ocena wierności

- Logika zarządzania dotyczy oczekiwanej reakcji organów wdrażających w odniesieniu do zasad programu i zachęt.
- Logika efektu dotyczy oczekiwanej reakcji osób, do których skierowana jest interwencja, z uwzględnieniem faktycznej realizacji interwencji.

Analiza procesu pod kątem wierności dostarcza informacji o tym, co faktycznie zostało osiągnięte za sprawą interwencji, a zatem co faktycznie przyczyniło się do zaobserwowanych efektów. Dostarcza ona również ważnych informacji zwrotnych na potrzeby zarządzania projektami.

Szczególnie ważna w kontekście ewaluacji kontrfaktycznej jest ocena różnic. Ewaluatorzy – podobnie zresztą jak autorzy niniejszego przewodnika – często koncentrują się na wpływie interwencji. Przed oceną wpływu na rezultaty należy jednak ocenić wpływ na wkład interwencji, czyli zbadać faktycznie wynikającą z interwencji różnicę między możliwościami oferowanymi grupie objętej interwencją i grupie kontrolnej. W zasadzie każdą ewaluację wpływu opartą na metodach kontrfaktycznych można „odwrócić do góry nogami” i wykorzystać grupę objętą interwencją jako grupę kontrolną do oceny rezultatów w odniesieniu do osób należących do grupy, którą przed tą zamianą przyjęto za grupę kontrolną. Oznacza to, że doświadczenia grupy kontrolnej należy poznać równie dobrze co doświadczenia grupy objętej interwencją, ponieważ to właśnie różnica między wkładem interwencji w odniesieniu do tych dwóch grup stanowi – w rozumieniu ewaluacji wpływu opartej na metodach kontrfaktycznych – przyczynę różnic w rezultatach.

Różnica między grupą poddaną interwencji a grupą kontrolną

Wracając ponownie do programu szkoleniowego, można sobie wyobrazić dwie zupełnie różne sytuacje wyjściowe. W jednym przypadku program szkoleniowy jest realizowany w ogólnym kontekście, w którym szkolenia tego typu nie są dostępne z innych źródeł. Grupa kontrolna po prostu radzi sobie

bez takiego szkolenia. Możliwa jest jednak odmienna sytuacja, w której istnieją jakieś inne szkolenia. Mogą one być oferowane na przykład przez firmy specjalizujące się w przygotowaniu zawodowym. W takim przypadku analiza procesu musi obejmować, w możliwym zakresie, ocenę różnicy w uczestnictwie w szkoleniach między grupą objętą interwencją a grupą kontrolną, a nie tylko zakładać, że wszystkie aspekty interwencji są poza zasięgiem grupy kontrolnej.

Chociaż ewaluację procesu można zlecać całkowicie niezależnie od innych form ewaluacji, ich znaczenie zarówno dla zarządzania, jak i ewaluacji wpływu opartej na metodach kontrfaktycznych sprawia, że planowanie ewaluacji procesu i ewaluacji wpływu powinno odbywać się jednocześnie.

CIE wymaga ewaluacji procesu

Dobra analiza procesu może pomóc ocenić wierność interwencji, a ewaluację procesu zapewniają wyjaśnienie przyczynowe interwencji. Bez dobrze zaprojektowanej ewaluacji procesu często trudno jest w pełni zinterpretować wyniki ewaluacji wpływu opartej na metodach kontrfaktycznych lub oszacować koszty niezbędne do oceny korzyści i kosztów, kiedy już dostępne są szacunki wpływu.

Jak zauważono powyżej, kolejnym dodatkowym wkładem, jaki ewaluacja procesu może wnieść w interpretację ustaleń poczynionych w ramach ewaluacji wpływu, jest uwzględnienie kontekstu, w którym zrealizowano interwencję. Zrozumienie kontekstu jest ważne, ponieważ pozwala poznać warunki powodzenia interwencji oraz daje pewien obraz tego, w jakim stopniu interwencja mogłaby przynieść podobne efekty, gdyby została wdrożona w innym miejscu, na innym obszarze geograficznym lub w innym czasie. Jest to szczególnie ważne przy omawianiu możliwości przenoszenia podejść politycznych i propagowaniu dobrych praktyk w zakresie uczenia się i wymiany w wymiarze transnarodowym. Analiza procesu pozwala zwiększyć zaufanie do tego, co określa się mianem **trafności zewnętrznej** wyników ewaluacji.

W większości zastosowań analiza efektywności obejmuje ocenę efektywności kosztowej lub pełną analizę kosztów i korzyści.

Określanie wskaźników efektywności kosztowej

Analiza efektywności kosztowej obejmuje porównanie kosztów interwencji z jej efektami lub wpływem, które zostały określone w ramach ewaluacji wpływu opartej na metodach kontrfaktycznych. Mówiąc prościej, wskaźnik efektywności kosztowej określa się przez podzielenie wpływu interwencji – wyrażonego w jednostkach miary lub jednostkach standaryzowanych – przez koszt netto realizacji interwencji na jednostkę poddaną interwencji.

Wskaźnik efektywności kosztowej programu szkoleniowego, którego celem jest pomoc bezrobotnym w znalezieniu pracy, może pomóc określić wysokość środków na uczestnika potrzebnych, aby ułatwić mu wyjście z bezrobocia i znalezienie zatrudnienia.

Analiza kosztów i korzyści służy porównaniu korzyści z kosztem netto

Wskaźnik efektywności kosztowej jest ważnym miernikiem dla osób odpowiedzialnych za alokację zasobów na poszczególne programy. Wartości uzyskane z szeregu różnych interwencji umożliwiają osobom odpowiedzialnym za alokację zasobów dokonywanie względnych osądów co do tego, które interwencje zapewniają lepszy stosunek wartości do ceny.

Zamiast wyrażać efekty programu w jednostkach miary lub jednostkach standardowych, w **analizie kosztów i korzyści** próbuje się dokonać monetyzacji wpływu oszacowanego w ramach ewaluacji wpływu opartej na

metodach kontrfaktycznych i porównać uzyskane wartości z kosztami netto interwencji. Celem analizy kosztów i korzyści jest ustalenie, czy korzyści programu wyrażone w formie pieniężnej przewyższają jego koszty netto. W analizie kosztów i korzyści typowego programu szkoleniowego EFS porównuje się korzyści, jakie interwencja zapewnia uczestnikom programu, władzom i szerszej rozumianemu społeczeństwu, z kosztami netto interwencji. Jeśli chodzi o uczestników, korzyści z programu (zwykle lepsze szanse na zatrudnienie i zwiększone zarobki) określa się za pomocą ewaluacji wpływu opartej na metodach kontrfaktycznych. Tak określone korzyści zostaną pomniejszone o wartość podatków zapłaconych przez uczestników i innych kosztów zatrudnienia w celu uzyskania korzyści netto. Z perspektywy władz korzyści z interwencji będą wynikać z dodatkowych dochodów podatkowych i zmniejszonych świadczeń społecznych, natomiast to właśnie władze ponoszą większość kosztów interwencji. Korzyści dla społeczeństwa jako całości uzyskuje się przez zsumowanie korzyści dla uczestników i władz oraz pomniejszenie ich o sumę kosztów dla uczestników i władz.

Analiza efektywności kosztowej oraz analiza kosztów i korzyści wciąż nie są zbyt powszechnie stosowane w ewaluacjach dotyczących Europejskiego Funduszu Społecznego. Analizy te są jednak bardzo przydatne w podejmowaniu decyzji, czy dana interwencja powinna być ponownie finansowana w przyszłości, lub w identyfikacji najbardziej skutecznej interwencji spośród zestawu podobnych interwencji (zob. przykład w poniższej ramce).

Oszacowanie wpływu w ramach ewaluacji wpływu opartej na metodach kontrfaktycznych jest kluczowym składnikiem zarówno analizy efektywności kosztowej, jak i analizy kosztów i korzyści. W przypadku tej pierwszej oszacowanie wpływu stanowi miarę skuteczności, natomiast w przypadku tej drugiej jest kluczowym źródłem szacowania korzyści pieniężnych. Oczywiście jest również, że zarówno badania efektywności kosztowej, jak i analiza kosztów i korzyści wymagają gromadzenia dokładnych danych na temat kosztów, z których można uzyskać koszty netto. Takie działania są zwykle określane jako analiza kosztów. W niektórych złożonych ewaluacjach prowadzonych metodami mieszanymi analizy kosztów często stanowią zintegrowany element ewaluacji procesu, w którym instrumenty badawcze mogą być dostosowane na potrzeby zbierania ważnych danych na temat kosztów.

Ramka 1. Przykład ewaluacji wpływu opartej na metodach kontrfaktycznych służącej porównaniu korzyści i kosztów

Przykład zastosowania podejścia kontrfaktycznego do oszacowania korzyści i kosztów przedstawiono w badaniu [Bazzoli M. i in., 2018](#). Badacze skoncentrowali się na programach szkolenia zawodowego przeprowadzonych we włoskiej Prowincji Autonomicznej Trydentu w latach 2010–2011, obejmujących ponad 300 godzin zajęć szkoleniowych. Ewaluacji poddano dwie grupy kursów: te finansowane ze środków prowincji i te finansowane z Europejskiego Funduszu Społecznego (EFS), w których brało udział odpowiednio 954 i 205 uczestników.

W celu wdrożenia analizy kosztów i korzyści przeprowadzono następujące działania: 1) autorzy ocenili wpływ kursów szkoleniowych na prawdopodobieństwo znalezienia pracy przez uczestników w ciągu trzech lat od rozpoczęcia kursu, stosując przy tym metodę *propensity score matching*¹²; 2) oszacowano wpływ

¹² Dokładniej mówiąc, autorzy zastosowali blokowanie z wykorzystaniem estymatora korekty regresji.

kursów na dochody brutto do końca 2013 r.; 3) autorzy oszacowali wielkość dodatkowych zwrotów fiskalnych (wynikających z wpływu na zarobki) oraz oszczędności w środkach publicznych wynikające ze zmniejszenia liczby osób pobierających zasiłki dla bezrobotnych; 4) porównano korzyści i koszty kursów.

Zastosowano szereg administracyjnych zbiorów danych: a) dane z monitorowania dotyczące uczestników i ich cech; b) dane z rejestrów bezrobotnych z publicznej służby zatrudnienia (*centri per l'impiego*, czyli urzędów pracy) w celu identyfikacji grupy kontrolnej; c) dane z bazy danych COB, archiwum obowiązkowych powiadomień firm o umowach o pracę, przesyłanych do publicznych służb zatrudnienia i wykorzystywanych do identyfikacji statusu zatrudnienia w grupie poddanej interwencji i grupie kontrolnej, zarówno przed udziałem w kursie szkoleniowym, jak i po udziale w nim; d) dane z archiwów dochodów podatkowych w celu obliczenia zarobków osób fizycznych oraz e) dane z Włoskiego Instytutu Zabezpieczeń Społecznych (INPS) w celu uzyskania informacji na temat zasiłków dla bezrobotnych otrzymywanych przez osoby fizyczne.

Po 36 miesiącach w grupie poddanej interwencji, czyli uczestniczącej w kursach finansowanych ze środków prowincji, prawdopodobieństwo zatrudnienia było o około 5 punktów procentowych wyższe niż w grupie kontrolnej, podczas gdy wpływ kursów EFS był znacznie wyższy, o około 28 punktów procentowych. W ciągu 3 lat po interwencji osoby, które uczestniczyły w kursach finansowanych ze środków prowincji zarabiały średnio o 2 250 EUR rocznie więcej niż grupa kontrolna, natomiast osoby uczestniczące w kursach EFS zarabiały o 4 106 EUR rocznie więcej niż grupa kontrolna. Dane te pozwoliły autorom również na oszacowanie korzyści dla administracji publicznej w postaci zwiększonych wpływów podatkowych i zmniejszonych wydatków na świadczenia społeczne. Osoby uczestniczące w kursach finansowanych ze środków prowincji zapłaciły o 126 EUR podatku dochodowego więcej niż grupa kontrolna za każdy analizowany rok, podczas gdy szacunkowa kwota w przypadku osób z grupy uczestniczącej w kursie EFS wyniosła 318 EUR rocznie. Jeśli chodzi o zmniejszenie zasiłków dla bezrobotnych wypłacanych przez administrację publiczną, wpływ kursów był znikomy, najprawdopodobniej dlatego, że wartość pieniężna zasiłków dla bezrobotnych zależała od okresu zatrudnienia przed uzyskaniem statusu bezrobotnego, a wielu uczestników było młodych z niewielkim doświadczeniem zawodowym.

Koszty interwencji wyniosły 4 800 EUR na uczestnika w przypadku kursów finansowanych ze środków prowincji i 14 500 EUR w przypadku kursów EFS. Autorzy porównali również koszty i korzyści na poziomie indywidualnym dla obu rodzajów kursów; średnio stwierdzono, że jeśli wziąć pod uwagę okres 2010–2013, koszty były wyższe niż korzyści¹³.

¹³ Inne „podobne” badanie znajduje się w [Lammers M. and Kok L., 2021](#).

Rozdział 2. Kwestie praktyczne przy opracowywaniu ewaluacji wpływu opartej na metodach kontrfaktycznych

W tym rozdziale omówiono kwestie praktyczne, które należy rozważyć podczas przygotowań do ewaluacji. Warto z niego korzystać przy planowaniu działań ewaluacyjnych, podejmując decyzję o tym, w przypadku których interwencji należy zastosować podejście polegające na ewaluacji wpływu opartej na metodach kontrfaktycznych, oraz w celu określenia kluczowych kwestii, które należy rozważyć podczas projektowania ewaluacji wpływu opartej na metodach kontrfaktycznych.

*Przygotowywanie
CIE*

Przyjmuje się, że sytuacja wyjściowa to taka, w której zarządzający programem w ramach IZ (lub kierownik instytucji pośredniczącej (IP) odpowiedzialnej za wdrażanie interwencji w ramach EFS) rozważa, które interwencje należy poddać ewaluacji, i ustala odpowiednie strategie wdrożenia ewaluacji wpływu opartej na metodach kontrfaktycznych. Przyjmuje się również, że urzędnicy w IZ nie będą sami przeprowadzać ewaluacji, ale będą zlecać usługi ewaluacyjne zewnętrznym ekspertom. Chociaż ewaluacja wpływu oparta na metodach kontrfaktycznych zostanie przeprowadzona przez wykonawcę, IZ (lub IP) będzie musiała zaplanować ewaluację i przygotować się do niej przed zleceniem jej wykonania.

W planie ewaluacji należy określić strategię ewaluacji, w tym poszczególne rodzaje ewaluacji, opisane w poprzednim rozdziale.

Plany ewaluacji są obowiązkowe dla wszystkich programów i muszą być zatwierdzone przez komitet monitorujący najpóźniej rok po zatwierdzeniu programu¹⁴. Plany te muszą być sporządzone na początku okresu programowania i muszą zawierać ustalenia dotyczące procesu ewaluacji (zarządzanie ewaluacjami i związek między ewaluacją a monitorowaniem), rzeczywiste działania związane z ewaluacją (np. orientacyjny wykaz ewaluacji, które mają być przeprowadzone, zakres każdej oceny, główne zagadnienia, niezbędne dane, potencjalne wykorzystanie, orientacyjny harmonogram, struktury zarządzania), harmonogram ewaluacji, budżet całkowity oraz budowanie zdolności do przeprowadzenia ewaluacji.

Plany ewaluacji mają zazwyczaj charakter ogólny, natomiast planowanie ewaluacji wpływu opartej na metodach kontrfaktycznych wymaga bardziej szczegółowych przygotowań. Najlepiej, aby nastąpiło to w momencie sporządzania planu ewaluacji, przy czym niektóre kwestie szczegółowe można dopracować na późniejszym etapie. IZ/IP muszą jednak mieć świadomość, że nawiązanie kontaktów z zainteresowanymi stronami i inne ustalenia niezbędne do gromadzenia danych związanych z interwencją rzadko bywa łatwe i wymaga planowania z dużym wyprzedzeniem.

W niniejszym przewodniku skupiono się na sposobach opracowania programu ewaluacji dla konkretnych interwencji kwalifikujących się do

*Opracowanie
programu ewaluacji*

¹⁴Plany ewaluacji są wymagane na mocy art. 44 ust. 5 i 6 rozporządzenia (UE) 2021/1060. Jeden plan może obejmować planowane ewaluacje więcej niż jednego programu, ale wszystkie programy muszą być objęte planem ewaluacji.

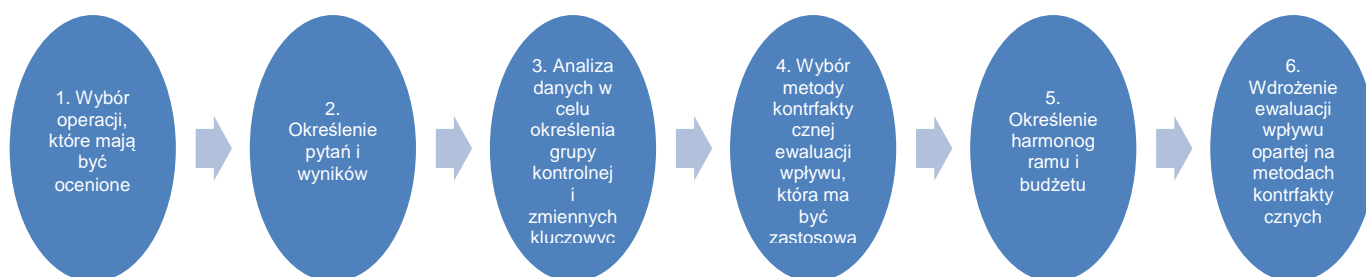
ewaluacji wpływu opartej na metodach kontrfaktycznych. Ten program powinien być częścią planu ewaluacji lub ewentualnie może być ustanowiony jako operacyjne działanie następcze wynikające z planu ewaluacji.

dla konkretnych interwencji

Nie wszystkie interwencje finansowane z Europejskiego Funduszu Społecznego mogą być poddane ewaluacji kontrfaktycznej. Decydenci muszą zdecydować, na czym należy się skoncentrować. Konieczne będzie przeprowadzenie procesu wyboru interwencji, które mają być poddane ewaluacji wpływu. W niniejszym przewodniku wskazano pewne aspekty, które IZ będą musiały wziąć pod uwagę przy wyborze odpowiednich interwencji. Ponadto głównym celem niniejszego przewodnika jest pomoc osobom odpowiedzialnym za zlecenie ewaluacji wpływu opartych na metodach kontrfaktycznych w przeanalizowaniu niektórych wyzwań związanych z opracowaniem skutecznej ewaluacji wpływu, a tym samym w opracowaniu programów ewaluacji dla poszczególnych rodzajów ewaluacji wpływu opartych na metodach kontrfaktycznych, które biorą pod uwagę.

W niniejszym przewodniku zakłada się, że po wybraniu interwencji, które mają być poddane ewaluacji wpływu opartej na metodach kontrfaktycznych, IZ będą musiały opracować program ewaluacji dla każdej wybranej interwencji. Termin „program” jest tu używany w celu odróżnienia tego działania od formalnych „planów” ewaluacji wymaganych na mocy rozporządzenia (UE) 2021/1060 w sprawie wspólnych przepisów na okres programowania 2021–2027. W szczególności termin „program” odnosi się do zestawu standardowych działań niezbędnych do zdefiniowania i wdrożenia ewaluacji wpływu opartej na metodach kontrfaktycznych, które to działania należy przygotować przed uruchomieniem programu, jak pokazano na poniższym wykresie.

Wykres 3. Zasadnicza kolejność działań w ramach CIE



Programy te będą stanowiły podstawowy dokument przy zleceniu ewaluacji wpływu opartych na metodach kontrfaktycznych i zapewnią podstawy umożliwiające wykonawcom przeprowadzenie rygorystycznej i wartościowej analizy. W dalszej części rozdziału omówiono kwestie, z którymi należy się zmierzyć podczas planowania ewaluacji. Programy ewaluacji muszą być dostosowane do specyficznych okoliczności, w jakich interwencja występuje. Nie można spekulować na temat tego, jakie dokładnie będą te okoliczności. W związku z tym w niniejszym przewodniku omówiono kwestie, które a) powinny być uwzględnione w programach lub b) powinny zachęcić do przeanalizowania wyzwań, którym trzeba będzie sprostać w związku z programami.

Kwestie, które należy rozważyć podczas planowania ewaluacji

Po zapoznaniu się z niektórymi kwestiami, które należy wziąć pod uwagę przy rozważaniu, które interwencje można poddać ewaluacji wpływu opartej na metodach kontrfaktycznych, oraz ustaleniu, czy przeprowadzenie takiej ewaluacji jest w ogóle możliwe przy użyciu dostępnych rodzajów danych, należy zwrócić uwagę na niektóre z kluczowych zagadnień, które należy rozważyć przy opracowywaniu programu ewaluacji. Taki program należy sporządzić na piśmie przed zleceniem przeprowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych lub szerszej analizy ewaluacyjnej, tak aby móc przygotować specyfikację istotnych warunków zamówienia oraz wyznaczyć wykonawcę. Zasadniczą treść programu ewaluacji przedstawiono w tabeli poniżej.

Tabela 1. Zalecana treść programu ewaluacji

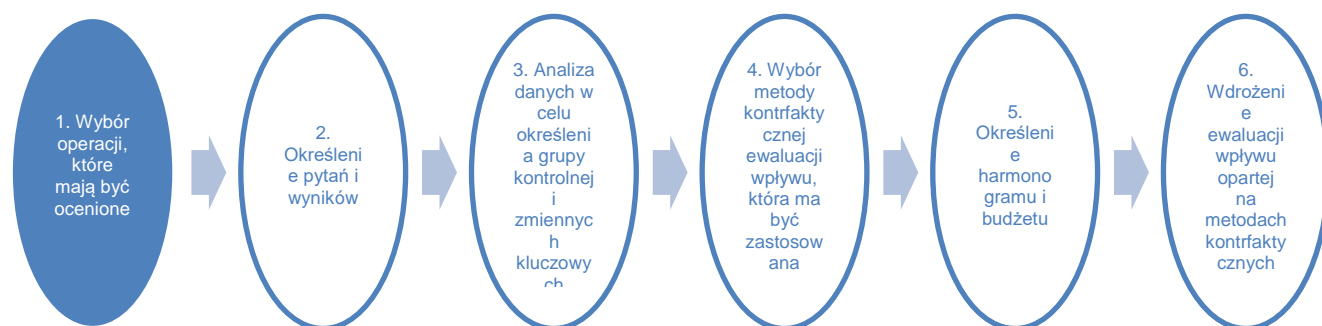
Główne etapy przygotowania i wdrożenia ewaluacji wpływu opartej na metodach kontrfaktycznych	Treść
1. Wybór operacji, które mają być ocenione	<ul style="list-style-type: none"> – Operacje lub rodzaje operacji w ramach EFS (jeśli to możliwe, należące do wcześniej określonej typologii), które mają być oceniane w drodze ewaluacji wpływu opartej na metodach kontrfaktycznych. – Krótki opis logiki działania wybranych operacji (cele, główne kryteria kwalifikowalności i populacja docelowa, rodzaje pomocy, model wdrożenia, przybliżone daty uruchomienia i zakończenia, orientacyjny budżet).
2. Określenie pytań ewaluacyjnych i zmiennych wynikowych	<ul style="list-style-type: none"> – Pytania ewaluacyjne. – Oczekiwane funkcjonowanie operacji („teoria zmiany”) i zmienne wynikowe (np. status zatrudnienia, zmiany w zarobkach, poziom ubóstwa, średnie wyniki egzaminów itp.).
3. Analiza danych w celu określenia grupy kontrolnej i pomiaru zmiennych wynikowych	<ul style="list-style-type: none"> – Dane administracyjne (np. rejestry bezrobotnych, rejestry podatkowe, administracyjne dane ubezpieczeniowe, rejestr szkół lub baza danych studentów itd.) lub inne dane (np. badania, duże zbiory danych), które należy wykorzystać. - Zmienne bazy danych lub badania, które mają być wykorzystane do ewaluacji wpływu opartej na metodach kontrfaktycznych. - Szereg czasowy koniecznych zmiennych. - Główne zasady i kwestie dotyczące dostępu do danych (bezpośrednia dostępność dla IZ, konieczność zawarcia umowy z innymi administracjami, przepisy i ograniczenia dotyczące prywatności).
4. Wybór metody kontrfaktycznej ewaluacji wpływu	<ul style="list-style-type: none"> – Możliwa metoda kontrfaktycznej ewaluacji wpływu, którą należy zastosować (można ją szczegółowo opisać na późniejszym etapie, ale wykorzystanie badania randomizowanego z grupą kontrolną lub innych metod należy wskazać z wyprzedzeniem, aby wspierać spójny proces ewaluacji).
5. Określenie harmonogramu i budżetu	<ul style="list-style-type: none"> – Harmonogram ewaluacji wpływu opartych na metodach kontrfaktycznych, główne cele pośrednie: a) decyzja o wyborze podejścia eksperymentalnego lub quasi-eksperymentalnego, b) szczegółowe pytania ewaluacyjne, c) przygotowanie SIWZ, d) wskazanie ewaluatora, e) przygotowanie danych, f) analiza danych, g) sprawozdania pośrednie i końcowe, h) walidacja wyników, i) rozpowszechnianie rezultatów i wniosków. – Budżet dostępny na ewaluację wpływu opartą na metodach kontrfaktycznych.
6. Wdrożenie ewaluacji	<ul style="list-style-type: none"> – Wybór ewaluatora. – Nadzór nad wdrożeniem ewaluacji wpływu opartej na metodach kontrfaktycznych. – Sporządzanie sprawozdań z ewaluacji wpływu opartej na metodach kontrfaktycznych. – Rozpowszechnianie rezultatów (główne zainteresowane strony, które należy zaangażować, główne narzędzia).

Ramka 2. Ewaluacja wpływu oparta na metodach kontrfaktycznych osadzona w szerszych ramach

Wiele przeprowadzonych w państwach członkowskich ewaluacji wpływu opartych na metodach kontrfaktycznych dotyczących interwencji finansowanych z EFS jest osadzonych w szerszych ramach ewaluacji:

- w Niemczech w ramach PO Bund EFS na lata 2014–2020 ewaluacja kontrfaktyczna programu na rzecz integracji osób długotrwale bezrobotnych na rynku pracy finansowanego w ramach PI 9.i, jest częścią szerszej strategii ewaluacyjnej przewidującej roczne sprawozdania okresowe, w których analizowane są również zagadnienia związane z realizacją interwencji. W latach 2017–2021 sporządzono cztery sprawozdania¹⁵. To samo dotyczy ewaluacji środków w ramach EFS wspierających integrację osób długotrwale bezrobotnych w Badenii-Wirtembergii, w przypadku której to ewaluacji analiza kontrfaktyczna nastąpiła po bardziej jakościowej analizie, skoncentrowanej na ocenach interwencji dokonanych przez uczestników¹⁶.
- W Piemoncie (Włochy) ewaluacja kontrfaktyczna dotycząca wpływu bonów dla osób w trudnej sytuacji, finansowanych w ramach PI 8.i oraz 9.i, na zatrudnienie była jednym z etapów bardziej kompleksowej bieżącej ewaluacji. W dwóch sprawozdaniach wstępnych przeanalizowano kwestie związane z wdrażaniem i subiektywne spostrzeżenia uczestników (kwiecień 2018 r. i luty 2019 r.), natomiast w dwóch kolejnych sprawozdaniach opublikowanych w lipcu 2019 r. i pod koniec 2020 r. skupiono się na wpływie na zatrudnienie przy zastosowaniu podejścia kontrfaktycznego¹⁷.
- W Marche (Włochy) ewaluację wpływu opartą na metodach kontrfaktycznych w celu zbadania wpływu interwencji na rzecz długotrwale bezrobotnych w 2020 r. przeprowadzono w formie analizy wpływu skoncentrowanej tematycznie po bardziej ogólnej analizie wpływu operacji w ramach EFS ukierunkowanych na osoby bezrobotne w 2019 r. (sprawozdanie z praktyki)¹⁸.
- W Polsce w ramach programu operacyjnego Wiedza Edukacja Rozwój na lata 2014–2020 w okresie od końca 2015 r. do maja 2020 r. zlecono przeprowadzenie szeregu ewaluacji (8 sprawozdań¹⁹) w celu przeanalizowania wsparcia w ramach EFS i Inicjatywy na rzecz zatrudnienia ludzi młodych z różnych perspektyw i przy zastosowaniu różnych metod analitycznych (np. analiz jakościowych za pośrednictwem badań i wywiadów, modeli makro, działań w terenie oraz podejścia kontrfaktycznego).

2.1. Wybór interwencji, które mają być poddane ewaluacji wpływu



¹⁵ Zob. [Boockmann B. i in., 2017](#) - [Boockmann B. i in., 2018](#) - [Boockmann B. i in., 2019](#) oraz [Boockmann B. i in., 2021](#). W 2019 r. i 2021 r. zastosowano ewaluację wpływu opartą na metodach kontrfaktycznych.

¹⁶ Zob. [Hunger K. i Sattler K., 2017](#) oraz [Scheller F. i Seidel K., 2020](#). W 2020 r. zastosowano ewaluację wpływu opartą na metodach kontrfaktycznych.

¹⁷ Zob. [Pomatto G., 2017](#) – [Pomatto G., 2019](#) – [Poy S., 2019](#); [Poy S., 2020](#). W 2019 r. i 2020 r. zastosowano ewaluację wpływu opartą na metodach kontrfaktycznych.

¹⁸ Zob. [Pompili M., Giorgetti I., 2020](#) - [Pompili M., Giorgetti I., 2020a](#)

¹⁹ Zob. [Instytut Badań Strukturalnych - Imapp - IQS, 2015](#) – [Baran J. i in., 2016](#) – [Baran J. i in., 2017](#) - [Baran J. i in., 2018](#) - [Baran J. i in., 2018a](#) – [Palczyńska M. i in., 2019](#) - [Kalinowski H., 2020](#) - [Kalinowski H. i in., 2020](#). W 2017 r. i 2020 r. zastosowano ewaluację wpływu opartą na metodach kontrfaktycznych.

Wybór interwencji, które mają być poddane ewaluacji wpływu, wymaga przeprowadzenia trzech głównych etapów:

Kryteria wyboru interwencji

1. Należy zidentyfikować kwestie strategiczne.
2. Po określeniu priorytetów strategicznych należy ocenić, czy poszczególne interwencje odpowiadają podstawowym wymogom podejścia konfaktycznego oraz w jakim stopniu są one innowacyjne lub wnoszą znaczący wkład do bazy wiedzy.
3. Należy jasno określić dostępność lub potencjalną dostępność rodzajów danych niezbędnych do przeprowadzenia ewaluacji wpływu opartej na metodach konfaktycznych. Ta trzecia kwestia okazała się dotychczas główną przeszkodą w przeprowadzaniu konfaktycznych ewaluacji interwencji w ramach EFS i zasługuje na szczególną uwagę.

Ramka 3. Pytania ułatwiające wybór interwencji, które mają być poddane ewaluacji wpływu opartej na metodach konfaktycznych

Ewaluacja wpływu oparta na metodach konfaktycznych nie jest odpowiednia dla wszystkich interwencji, a przeprowadzenie jej w przypadku każdej interwencji z zasady nie jest efektywne kosztowo. Instytucje zarządzające muszą zdecydować, w jaki sposób rozdzielić zasoby, aby osiągnąć największe korzyści. Oczekuje się, że wybory te zostaną uwzględnione w planie ewaluacji, a przy planowaniu ewaluacji wpływu opartych na metodach konfaktycznych należy uwzględnić trzy główne elementy: a) priorytety strategiczne, b) wykonalność ewaluacji wpływu opartej na metodach konfaktycznych oraz c) dostępność niezbędnych danych.

Na **strategię ewaluacji** wpływ mają: skala, rozwój polityki, obszary niepewności i zapotrzebowanie na wiedzę. Instytucje zarządzające powinny zadać sobie następujące pytania:

- Czy duże kwoty środków przeznaczonych na daną interwencję sprawiają, że uzasadnienie wydatków jest szczególnie ważne? Takie interwencje są stosunkowo łatwe do zidentyfikowania, ponieważ otrzymują one większość środków przeznaczonych na każdy cel szczegółowy (jak określono w art. 2 rozporządzenia w sprawie EFS+ 2021/1057).
- Czy środek jest przedmiotem procesu reformy i czy wyniki ewaluacji mogą posłużyć do krytycznego przeglądu działań w tym zakresie? Takie interwencje są związane z niedawnymi reformami polityki pracy, polityki edukacyjnej lub polityki społecznej; chociaż nie przeznaczają się na nie znacznych środków pieniężnych, to jednak mają kluczowe znaczenie dla powodzenia reformy.
- Czy interwencja jest innowacyjna i jest testowana w ramach pilotażu lub badania przed przeprowadzeniem jej na większą skalę? Choć na takie interwencje nie przeznaczają się raczej ogromnych środków, wymagają one wczesnej oceny w celu podjęcia decyzji, czy dany eksperyment należy kontynuować i rozszerzać, czy też należy go zakończyć.
- Czy interwencja koncentruje się na obszarach, które wymagają dodatkowych dowodów skuteczności? Do tej grupy należą interwencje, które w przeszłości nie były poddane ewaluacji lub których ostatnie ewaluacje przeprowadzono tak dawno, że wymagają aktualizacji. W programach EFS+ występuje mnóstwo takich interwencji, ponieważ ewaluacje wpływu oparte na metodach konfaktycznych dopiero niedawno zaczęły być powszechnie stosowane, a „efekty netto” są nieznanne.

Wykonalność odnosi się zarówno do cech interwencji, jak i do okoliczności, w których są one wprowadzane. Osoby odpowiedzialne za planowanie powinny być w stanie udzielić twierdzących odpowiedzi na następujące pytania:

- Czy oddziaływanie interwencji jest odrębne, czy wyróżnia się na tle innych i czy jest wystarczająco jednolite?
- Czy porównanie między grupą objętą interwencją a grupą kontrolną jest wystarczająco znaczące, aby można było zmierzyć wpływ interwencji?
- Czy populacja docelowa jest duża i dobrze zdefiniowana?

- Czy teoria, która łączy interwencję z zamierzonymi wynikami, jest logicznie spójna?
- Czy w populacji docelowej można jednoznacznie zidentyfikować grupę objętą interwencją?
- Czy wielkość grupy objętej interwencją jest wystarczająca?
- Czy można zidentyfikować wiarygodną grupę kontrolną?
- Czy różnica pomiędzy doświadczeniami grupy objętej interwencją i grupy kontrolnej może utrzymywać się przez wystarczająco długi okres, aby można było zmierzyć wpływ?

Najważniejsze są **dane**. Istotą ewaluacji wpływu opartej na metodach kontrfaktycznych jest pomiar, a pomiar wymaga informacji ilościowych zarówno na temat grupy objętej interwencją, jak i grupy kontrolnej, oraz kontekstu, w którym prowadzona jest ewaluacja. To, jakie dane są potrzebne, zależy zwykle od teorii interwencji i strategii zastosowanej do ustalenia scenariusza kontrfaktycznego. Wybierając interwencje w ramach ewaluacji wpływu opartej na metodach kontrfaktycznych, planująca ewaluację instytucja zarządzająca musi postawić następujące pytania:

- Co koniecznie należy wiedzieć o członkach grupy docelowej i grupy kontrolnej?
- Co koniecznie należy wiedzieć o charakterze interwencji, jaka jest faktycznie realizowana w grupie objętej interwencją?
- Czy grupa kontrolna została poddana jakimkolwiek innemu oddziaływaniu? Czy dostępne są dane na ten temat?
- Jakie dane są dostępne ze źródeł administracyjnych i innych?
- Czy dostępne są dane opisujące karierę zawodową poszczególnych osób?
- Czy zindywidualizowane dane z różnych źródeł mogą być łączone?

Więcej szczegółów na temat tych kwestii omówiono w dalszej części tego rozdziału.

2.1.1. Nadanie priorytetu interwencjom na potrzeby ewaluacji wpływu

Przed nadaniem priorytetu konkretnym interwencjom na potrzeby ewaluacji wpływu opartej na metodach kontrfaktycznych należy rozważyć szersze kwestie strategiczne. Należy skupić się na wyborze tych interwencji, w przypadku których ewaluacje wpływu gwarantują największy zysk w postaci wiedzy o tym, co działa. Korzyści płynące z dobrze zaprojektowanych, starannych ewaluacji wynikają nie tylko dla instytucji zarządzającej i państwa członkowskiego, które je zlecają, ale również dla innych państw członkowskich i ich instytucji zarządzających, innych zainteresowanych stron oraz Komisji.

Wkład w uzasadnienie wydatków

Biorąc pod uwagę, że ewaluacje wpływu oparte na metodach kontrfaktycznych koncentrują się na kwestiach, które mają kluczowe znaczenie dla decydentów, w szczególności tych odpowiedzialnych za decyzje o alokacji zasobów, zasadne jest ukierunkowanie działań w zakresie ewaluacji wpływu na programach i interwencjach, które wymagają szczególnie wielu zasobów. Im więcej czasu i innych zasobów pochłania dany program lub interwencja, tym ważniejsze jest ustalenie, czy jest on skuteczny, a zatem czy generowane korzyści przewyższają poniesione koszty. Kosztowne interwencje, które nie przynoszą wartości społecznej ani ekonomicznej, mogą wymagać ponownego rozważenia, podczas gdy inne, które wykazują wartość dodaną, mogą zasługiwać na zwiększone finansowanie i większą uwagę.

Ukierunkowanie na interwencje wymagające wielu zasobów

Wyniki przeprowadzonych niedawno ewaluacji interwencji w ramach EFS finansowanych w okresie programowania 2014–2020 wykazały, że konieczne są strategiczne korekty i większa koncentracja na kluczowych celach polityki. Interwencje dotyczące zatrudnienia i mobilności pracowników okazały się mniej skuteczne w przypadku osób starszych i osób bardziej oddalonych od rynku pracy; grupy te wymagają nowych i bardziej skutecznych instrumentów²⁰. Chociaż interwencje w zakresie polityki społecznej wykazują duże zróżnicowanie pod względem kosztów na uczestnika i rodzaju działania, zazwyczaj nie przeprowadza się systematycznej analizy kosztów i korzyści²¹. Interwencje w zakresie kształcenia i szkolenia wykazały pozytywne rezultaty, jednak ograniczona liczba badań wpływu uniemożliwia pełną ocenę ich długoterminowych efektów²². Ewaluacje wpływu oparte na metodach kontrfaktycznych oferują możliwość przeglądu interwencji w celu określenia najbardziej skutecznych i efektywnych podejść dla danych grup docelowych, a tym samym maksymalizacji stosunku wartości do ceny nowych programów EFS+.

Wkład interwencji w proces reform

Interwencje, które stanowią kluczowy element szerszego programu reform, często przyciągają znaczne finansowanie. Fakt, że interwencja w ramach EFS ma centralne znaczenie dla strategii włączenia społecznego lub stanowi kluczowy element aktywnego programu rynku pracy, sprawia, że w naturalny sposób poświęca się jej większą uwagę.

Interwencje przyczyniające się do innowacji w zakresie polityki

Innowacyjne i odkrywcz

Oczywistymi kandydatami do ewaluacji wpływu opartej na metodach kontrfaktycznych są nowe i innowacyjne interwencje pilotażowe. Testowanie efektów interwencji za pomocą pilotażu lub badania niewątpliwie wymaga starannej ewaluacji. Przeprowadzenie dobrze zaprojektowanej ewaluacji wpływu opartej na metodach kontrfaktycznych jest tym ważniejsze, jeżeli istnieje wyraźne zobowiązanie do zwiększenia skali lub szerszego wdrożenia interwencji, w przypadku gdy zostanie ona uznana za udaną.

Wkład na rzecz nauki

Skupienie uwagi i zasobów na konkretnych obszarach objętych programem i specyficznych interwencjach w tych obszarach jest szczególnie uzasadnione, gdy istnieje niewiele lub nie ma żadnych dowodów na to, co sprawdza się w danym obszarze polityki. Może na przykład panować prawdziwa niepewność co do danej polityki w przyszłości oraz ryzyko nadmiernego polegania na dowodach, które mogą nie mieć bezpośredniego odniesienia (np. dowody pochodzące z innych krajów).

Przedstawienie dowodów

Ewaluacje wysokiej jakości mogą być uznane za dobra publiczne. Korzyści, jakie te ewaluacje generują w zakresie uczenia się, rozciągają się na różne zainteresowane strony, nie tylko te związane z daną IZ. W związku z tym ważne jest, aby rozważyć, które zainteresowane strony mogą skorzystać na proponowanej ewaluacji wpływu. Mogą to być instytucje pośredniczące (IP) lub agencje zajmujące się interwencjami w ramach tego samego programu, inne IZ lub IP w danym państwie członkowskim bądź agencje i instytucje zajmujące się funduszami krajowymi lub regionalnymi. Inną oczywistą

²⁰ [Fondazione G. Brodolini, Metis GmbH, Applica, Ockham IPS \(2020\).](#)

²¹ [ICF, Cambridge Econometrics and Eurocentre \(2020\).](#)

²² [Ecorys, Ismeri Europa \(2020\).](#)

zewnętrzną zainteresowaną stroną, którą należy wziąć pod uwagę, jest Komisja Europejska. Warto też pamiętać o zainteresowanych stronach w innych państwach członkowskich, które mogą zaczerpnąć wiedzę z ewaluacji. Uwzględnienie potrzeb podmiotów innych niż bezpośrednio zainteresowane strony stanowi istotny wkład, jaki decydenci i zarządzający programami mogą wnieść we wzajemne uczenie się.

Ostatnią kwestią strategiczną przy wyborze obszarów wymagających uwagi podczas opracowywania ewaluacji wpływu opartych na metodach kontrfaktycznych jest rozważenie tych interwencji, które mogą pomóc wykazać korzyści z tego typu ewaluacji wpływu i pełnić rolę wzorca.

Opanowanie metod kontrfaktycznej ewaluacji wpływu

2.1.2. Wybór interwencji, które są odpowiednie do podejścia kontrfaktycznego

Po rozważeniu szerszych zagadnień strategicznych, które mogą uzasadniać wybór konkretnych interwencji na potrzeby ewaluacji wpływu opartej na metodach kontrfaktycznych, w tej części przeanalizowano konkretne cechy interwencji, które mogą sprawić, że będą one odpowiednie do podejścia kontrfaktycznego. Takich cech jest wiele i są one zróżnicowane. Może się zdarzyć, że w pewnych okolicznościach niektóre cechy interwencji będą zachęcały do przeprowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych, a w innych – będą utrudniały jej realizację. Dlatego nie da się opracować wyczerpującej listy czynników, które należy wziąć pod uwagę. Warto jednak wspomnieć o tych cechach interwencji, które z dużym prawdopodobieństwem doprowadzą do udanej ewaluacji wpływu opartej na metodach kontrfaktycznych.

Ramka 4. Najczęstsze rodzaje interwencji i grup docelowych wybieranych na potrzeby ewaluacji wpływu opartych na metodach kontrfaktycznych dotyczących EFS

Większość ewaluacji wpływu opartych na metodach kontrfaktycznych dotyczących interwencji w ramach EFS koncentruje się na interwencjach związanych z aktywną polityką rynku pracy skierowanych ogólnie do osób bezrobotnych i osób bezrobotnych należących do konkretnych podgrup w niekorzystnej sytuacji. Duża liczba analiz wpływu interwencji skierowanych do osób młodych wynika z przepisów dotyczących Inicjatywy na rzecz zatrudnienia ludzi młodych, które wymagają, aby takie ewaluacje były przeprowadzane w określonych odstępach czasu.

Około połowa ewaluacji wpływu opartych na metodach kontrfaktycznych zidentyfikowanych w ramach projektu Punktu Informacyjnego ds. Oceny²³ od 2015 r. odnosi się do celu tematycznego 8 – „promowanie trwałego i wysokiej jakości zatrudnienia oraz wsparcie mobilności pracowników”. Podobnie w większości ewaluacji wpływu opartych na metodach kontrfaktycznych badających interwencje finansowane w ramach celu tematycznego 9 uwaga jest zwrócona na skuteczność interwencji w zakresie integracji na rynku pracy osób bezrobotnych znajdujących się w trudnej sytuacji (np. niemiecka interwencja skierowana do osób długotrwale bezrobotnych finansowana w ramach PO Bund EFS na lata 2014–2020).

Najczęściej analizowanymi formami wsparcia udzielanego osobom bezrobotnym są szkolenia, staże lub inne formy zdobywania doświadczenia zawodowego oraz dotowane miejsca pracy. Wynika to wyraźnie z przykładów analizowanych w niniejszym przewodniku (zob. przykłady włoskiej ewaluacji w Marche, ewaluacje dotyczące szkoleń dla migrantów w Niemczech, szkolenia zawodowe dla młodzieży niekształcącej się, niepracującej ani nie szkolącej się na Łotwie). Analiza usług doradztwa i kojarzenia ofert pracy z poszukującymi zatrudnienia jest mniej powszechna (jak w przypadku omówionego w niniejszym przewodniku szwedzkiego przykładu, w którym działanie pilotażowe na rzecz intensyfikacji wsparcia dla bezrobotnych skierowane do publicznej służby zatrudnienia jest poddawane ewaluacji z zastosowaniem podejścia randomizowanego). Według danych Punktu Informacyjnego ds. Oceny interwencji w ramach

²³ Zob. https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/

EFS wspierające samozatrudnienie lub tworzenie przedsiębiorstw nie są często poddawane ewaluacji²⁴.

W niektórych przypadkach w ramach ewaluacji wpływu opartych na metodach kontrfaktycznych analizuje się różne rodzaje interwencji w sposób zbiorczy, co wiąże się z ryzykiem wymieszania różnych logik interwencji i zmniejszenia wiarygodności porównania między grupą objętą interwencją a grupą kontrolną.

Warta odnotowania jest próba ewaluacji efektów interwencji dla osób w trudnej sytuacji, nie tylko pod względem rezultatów w zakresie zatrudnienia, ale również „miękkich wyników”: na tym aspekcie skoncentrowano się przy ewaluacji wpływu opartej na metodach kontrfaktycznych przeprowadzonej w Niemczech: oceniono wówczas wpływ programów tworzenia miejsc pracy dla długotrwale bezrobotnych na postrzegane wskaźniki stanu zdrowia, zadowolenie z życia, poczucie przynależności i status społeczny.

Rzadziej oceniane przy zastosowaniu podejścia kontrfaktycznego są interwencje w ramach EFS w obszarze edukacji. Przeprowadza się mniej ewaluacji wpływu opartych na metodach kontrfaktycznych w odniesieniu do celu tematycznego 10 niż do celu tematycznego 8 lub 9, przy czym często ewaluacje skoncentrowane są na interwencjach finansowanych w ramach PI 10.IV, dotyczących kształcenia zawodowego i jego efektów w zakresie integracji na rynku pracy. Są dwie przyczyny takiego stanu rzeczy: bardziej ograniczona dostępność do zbiorów danych oraz bardziej rygorystyczne przepisy dotyczące prywatności²⁵. Niemniej jednak podjęto pewne próby przeprowadzenia takich ewaluacji: w hiszpańskiej Asturii oceniono interwencję w szkołach średnich mającą na celu odwołanie od wczesnego kończenia nauki; w Polsce (w województwie podlaskim) przeprowadzono ewaluację wpływu opartej na metodach kontrfaktycznych, aby zbadać efekty projektu mającego na celu popularyzację kształcenia zawodowego wśród młodych uczniów; w Portugalii przedmiotem analizy były dotacje wspierające studentów uczelni wyższych.

Ewaluacje wpływu oparte na metodach kontrfaktycznych mogą być przeprowadzane na różnych poziomach polityki (tj. w odniesieniu do co najmniej jednej osi priorytetowej, co najmniej jednego celu szczegółowego lub co najmniej jednej operacji²⁶ w ramach danego programu) i mogą obejmować krajowe lub regionalne programy EFS+, koncentrować się na jednorodnych grupach docelowych (mężczyznach lub kobietach, młodzieży lub słabszych grupach społecznych, osobach długotrwale bezrobotnych itp.) lub rodzajach interwencji (np. szkoleniach, usługach na rzecz włączenia społecznego lub zniechęcaniu do wczesnego kończenia nauki (zob. poprzednia ramka)).

Przykłady z państw członkowskich wskazują, że do objęcia ewaluacją wpływu opartą na metodach kontrfaktycznych odpowiednie będą różnorodne instrumenty stosowane w ramach EFS+, w tym szkolenia, zachęty do zatrudnienia i usługi rynku pracy (np. doradztwo zawodowe, coaching). Z drugiej strony większym wyzwaniem przy przeprowadzaniu ewaluacji wpływu opartej na metodach kontrfaktycznych są interwencje w zakresie rotacji stanowisk i dzielenia etatu, zachęty do rozpoczęcia działalności gospodarczej lub wsparcie dla systemów i struktur, a także interwencje w dziedzinie edukacji i włączenia społecznego.

Należy się zastanowić, które interwencje są bardziej obiecujące z perspektywy ewaluacji wpływu opartej na metodach kontrfaktycznych, rozważając następujące pytania:

²⁴ Oto niektóre z przykładów: [Borik V. i in., 2015](#) - [Ires Piemonte, 2019](#) – [Openfield, 2019](#).

²⁵ Zob. np. [Ismeri Europa – Ecorys – Institute for Employment Studies, 2019](#).

²⁶ Zgodnie z art. 2 pkt 4 rozporządzenia Parlamentu Europejskiego i Rady (UE) 2021/1060: „operacja” oznacza: a) projekt, umowę, działanie lub grupę projektów wybrane w ramach danych programów; b) w kontekście instrumentów finansowych – wkład z programu do instrumentu finansowego oraz późniejsze wsparcie finansowe udzielane na rzecz ostatecznych odbiorców z tego instrumentu finansowego.

Czy interwencja jest odrębna, czy wyróżnia się na tle innych i czy jest względnie jednolita?

*Oddziaływanie
wyraźnie
odróżniające się*

Oddziaływanie danej interwencji musi odróżniać się od oddziaływania innych interwencji. Ponadto musi istnieć wyraźny kontrast między tym, co otrzymują uczestnicy interwencji, a korzyściami dla innych podobnych grup osób. Jeśli oddziaływanie jest na tyle rozproszone, że identyfikacja odrębnej grupy odbiorców nie jest możliwa, zastosowanie podejścia kontrfaktycznego jest również niemożliwe lub nie jest pożądane.

Jeżeli status danej jednostki (przedsiębiorstwa lub osoby fizycznej) związany z oddziaływaniem w ramach interwencji wpływa na potencjalny wynik innych jednostek (poprzez tak zwane szersze „efekty równowagi ogólnej”), wykorzystanie metod kontrfaktycznej ewaluacji wpływu staje się bardzo skomplikowane i trudne. W przypadku programów szkoleniowych może to mieć miejsce, gdy absolwenci programu utrudniają znalezienie pracy w krótkim okresie innym osobom, które nie uczestniczyły w programie. W przypadku gdy taka sytuacja stanowi istotny problem (na przykład w przypadku interwencji na dużą skalę), konieczne może być przeprowadzenie analizy makroekonomicznej, aby ocenić zakres efektu zastąpienia i wyparcia. W przypadku gdy istnieje prawdopodobieństwo wystąpienia takich efektów, IZ powinny zasięgnąć porady eksperta.

Interwencja sama w sobie powinna być stosunkowo jednolita. Oznacza to, że wszyscy uczestnicy interwencji powinni zasadniczo otrzymać taki sam pakiet środków lub być objęci takimi samymi środkami. Jeżeli zakres działań podejmowanych przez uczestników w ramach jednej interwencji jest zbyt zróżnicowany, może to mieć szereg implikacji dla ewaluacji wpływu opartej na metodach kontrfaktycznych. Po pierwsze, może okazać się, że w rzeczywistości nie ma sensu mówić o jednej spójnej interwencji, ale raczej o kilku interwencjach z odrębnymi procesami przyczynowymi; po drugie, trudno będzie zinterpretować wpływ przedstawiany jako średnie efekty netto w grupie różnych interwencji; po trzecie, analiza podgrup może być uzasadniona, ale jeśli w grupie objętej interwencją istnieje zbyt wiele podgrup, restrykcje co do liczebności próby mogą ograniczyć zdolność do przedstawienia użytecznych wyników.

*Jednolite
interwencje*

Czy interwencję porównuje się do braku interwencji? Czy występują inne istotne formy interwencji?

EFS współfinansuje krajowe i regionalne polityki rynku pracy i włączenia społecznego. Dlatego też w każdym programie ewaluacji dotyczącym ewaluacji wpływu opartej na metodach kontrfaktycznych należy dokładnie rozważyć, czy interwencja jest wyraźnie identyfikowalna i czy objęte nią osoby mają możliwość otrzymania usług z innych (krajowych lub regionalnych) programów i źródeł finansowania. Ważne jest, aby oddziaływanie poddane ewaluacji rzeczywiście powodowało zmianę możliwości lub zasobów dostępnych dla uczestników w porównaniu z tymi, które są dostępne dla osób z grupy kontrolnej, oraz, aby można było zmierzyć i monitorować wszelkie różnice.

Takie kwestie dotyczące „złożonego oddziaływania” są zwykle specyficzne dla danego kontekstu. Utrudniają one projektowanie i wdrażanie ewaluacji wpływu opartej na metodach kontrfaktycznych. Ich obecność uwydatnia znaczenie starannego planowania ewaluacji – opracowania programu ewaluacji – przed jej wdrożeniem.

*Złożone
oddziaływanie*

Czy grupa docelowa jest duża i dobrze zdefiniowana?

Ewaluacje wpływu oparte na metodach kontrfaktycznych wymagają dużej liczebności próby w porównaniu z niektórymi innymi formami ewaluacji. Kluczowe znaczenie ma zapewnienie grupy docelowej składającej się z odpowiedniej liczby osób oraz grup kontrolnych o wystarczającej liczebności. Poniżej omówiono tę kwestię bardziej szczegółowo.

Duża liczebność próby

Ważne jest, aby interwencja brana pod uwagę do ewaluacji wpływu opartej na metodach kontrfaktycznych była skierowana do dobrze zdefiniowanej grupy. Bez jasnego zrozumienia, kim są grupy docelowe, trudno jest określić znaczącą grupę kontrolną. W przypadku niektórych interwencji celowo dąży się do rekrutowania ich uczestników za pomocą nieformalnych mechanizmów, sprzyjających procesom, które nie są z góry określone ani zbyt nakazowe (np. docelowe grupy społeczne w niekorzystnej sytuacji, takie jak młodzież niekształcąca się, niepracująca ani nieszkoląca się lub grupy defaworyzowane, mogą być angażowane w ramach okazjonalnych i indywidualnych procedur); może to utrudnić precyzyjną identyfikację osób poddanych interwencji i powiązanej grupy kontrolnej.

Ustalenie tożsamości grupy docelowej

Czy istnieje wyraźny mechanizm przyczynowo-skutkowy?

Jak wspomniano wcześniej przy okazji omówienia głównych metod ewaluacji i „teorii zmian”, często przed przeprowadzeniem ewaluacji wpływu opartej na metodach kontrfaktycznych lub w połączeniu z nią warto przeprowadzić ewaluację opartą na teorii. Opracowanie teorii zmiany lub szczegółowej logiki interwencji może na wiele sposobów pomóc osobom projektującym ewaluację wpływu opartą na metodach kontrfaktycznych: przede wszystkim może pomóc ustalić, czy u podstaw interwencji leży spójny mechanizm przyczynowo-skutkowy. Interwencje pozbawione jasnego i przekonującego mechanizmu przyczynowo-skutkowego prawdopodobnie nie będą miały wystarczająco dużego wpływu, aby można je było określić statystycznie za pomocą ewaluacji wpływu opartej na metodach kontrfaktycznych.

Odrębny mechanizm polityki

Czy wyniki można określić ilościowo?

Istnieje potrzeba uzyskania wymiernych wyników (lub rezultatów). Takie dane i wskaźniki można pozyskać ze źródeł administracyjnych lub w ramach specjalnie ukierunkowanych badań.

Konieczność pomiaru wyników

W niektórych okolicznościach ilościowy pomiar zamierzonych rezultatów interwencji może wymagać szczególnych rozwiązań. Na przykład interwencja może dotyczyć zmiany postaw, przekonań lub opinii. W takim przypadku, aby zmierzyć te zmiany, należy przeprowadzić badania ankietowe. W niektórych przypadkach rezultaty interwencji mogą być mało precyzyjne lub słabo zdefiniowane. Również w tym przypadku opracowanie logiki interwencji może pomóc lepiej zrozumieć zamierzone efekty interwencji oraz sposób, w jaki ma ona spowodować zmiany odpowiednich rezultatów.

Czy interwencja została wprowadzona w taki sposób, aby można było wskazać znaczącą grupę kontrolną?

Aby zidentyfikować znaczącą grupę kontrolną, należy rozważyć następujące kwestie: w jaki sposób jednostki poddane interwencji (osoby fizyczne lub przedsiębiorstwa) są wybierane do udziału w interwencji lub dlaczego decydują się na wzięcie w niej udziału; czy to samo źródło danych – np. ten

Mechanizm wyboru uczestników interwencji

sam instrument badania – można zastosować do próby kontrolnej i do grupy objętej interwencją; oraz czy konieczne jest wybranie prób kontrolnych, które podlegają tym samym warunkom na rynku pracy co grupa objęta interwencją. Kilka przykładów przedstawiono w ramce poniżej.

Jeśli interwencja jest obowiązkowa i realizowana w całej populacji docelowej mniej więcej jednocześnie, może się okazać, że trudno będzie wskazać część populacji docelowej niepoddaną interwencji, która mogłaby pełnić rolę grupy kontrolnej.

Ramka 5. Definiowanie grup kontrolnych

W przykładach ewaluacji wpływu opartych na metodach kontrfaktycznych przeanalizowanych w niniejszym przewodniku wybór grup kontrolnych był podyktowany charakterystyką interwencji (na przykład kryteriami kwalifikowalności), a także dostępnością odpowiednich danych.

W porównaniu z doświadczeniami z ewaluacji wpływu opartych na metodach kontrfaktycznych z poprzedniego okresu programowania identyfikacja i wybór grupy kontrolnej częściej opierają się na danych administracyjnych, szczególnie pochodzących z rejestrów bezrobotnych. Najczęstszą strategią jest identyfikacja osób, które mogą należeć do grupy kontrolnej, o podobnych cechach, zarejestrowanych jako bezrobotne w publicznych służbach zatrudnienia w danym okresie, zgodnie z kryteriami kwalifikowalności. Dotyczy to niektórych ewaluacji przeprowadzonych we Włoszech (Marche, Trydent i Piemont), w Polsce (województwa lubelskie i podlaskie), jak również niemieckich ewaluacji szkoleń dla migrantów, programów tworzenia miejsc pracy oraz działań integracyjnych dla bezrobotnych w Badenii-Wirtembergii.

W innych ewaluacjach, które jednak nie zostały omówione w niniejszym przewodniku, osoby, które zgłosiły się do udziału w interwencji, ale nie zostały wybrane, zostały następnie wybrane do grupy kontrolnej. Było tak w przypadku włoskiej ewaluacji programu dotyczącego gwarancji dla młodzieży, niemieckiej ewaluacji projektu pilotażowego „Finansowanie miejsc pracy dla obywateli w latach 2007–2013” oraz ewaluacji programu PIPOL realizowanego we Friuli-Wenecji Julijskiej²⁷. Strategia ta nie jest jednak często stosowana, ponieważ w większości krajów i regionów systemy monitorowania informacji nie uwzględniają informacji o osobach, które się zgłosiły, ale nie wzięły udziału w programie.

Niemieckie organy przyjęły inne podejście do ewaluacji interwencji skierowanych do długotrwale bezrobotnych²⁸: w ewaluacji oceniono „zamiar poddania interwencji” (ang. *intention to treat*, ITT), a nie „średni efekt na jednostki poddane interwencji” (ang. *average treatment effects on treated*, ATT). W tym przypadku grupa poddana interwencji składała się z osób potencjalnie kwalifikujących się przez cały okres realizacji programu, niezależnie od tego, czy były one faktycznie poddane interwencji (od 2015 r.), a grupa kontrolna składała się z osób o tych samych kryteriach kwalifikowalności, ale które były długotrwale bezrobotne przed wdrożeniem programu (2010–2012).

Jedynie w przypadku Szwecji, która przyjęła podejście randomizowane, grupa kontrolna została określona losowo: w fazie projektowania, grupa osób młodych została poddana interwencji (zintensyfikowane wsparcie publicznych służb zatrudnienia) w sposób losowy, natomiast grupa kontrolna otrzymywała zwykłe wsparcie oferowane przez publiczne służby zatrudnienia.

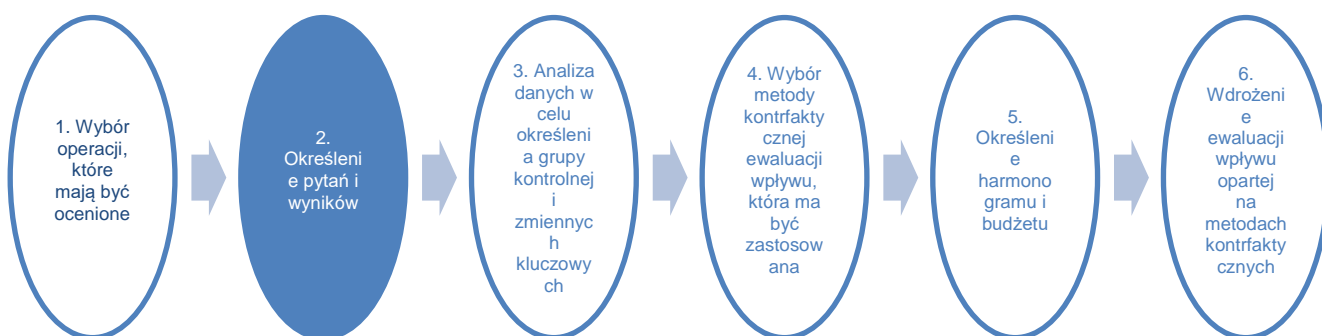
W nielicznych ewaluacjach koncentrujących się na działaniach na rzecz przedsiębiorstw w ramach EFS rozgraniczenie między grupą objętą interwencją a grupą kontrolną wyznaczono w podziale na wnioskodawców objętych finansowaniem i nieobjętych finansowaniem, podobnie jak w niektórych duńskich ewaluacjach, w których porównano wyniki przedsiębiorstw finansowanych z EFS z wynikami próby przedsiębiorstw o podobnych cechach, które nie otrzymały wsparcia²⁹.

²⁷ Zob. odpowiednio [Isfol, 2016](#) – [IAW Institut für Angewandte Wirtschaftsforschung - ISG Institut für Sozialforschung und Gesellschaftspolitik, GmbH, 2015](#) - [Ismeri Europa, 2018](#).

²⁸ Zob. [Boockmann B. i in., 2019](#).

²⁹ Na przykład: [Danmarks Statistik i in., 2017](#) oraz [Danmarks Statistik i in., 2018](#).

2.2. Pytania ewaluacyjne i zmienne wynikowe



2.2.1. Jakie są założenia i cele interwencji?

Przy określaniu programu ewaluacji należy przede wszystkim opisać założenia, cele i główne elementy interwencji.

W wielu przypadkach dokumenty, które określają założenia i cele interwencji, są już dostępne. W przypadku ewaluacji wpływu opartej na metodach kontrfaktycznych ważne jest jednak, aby określić rezultaty i zmiany, które interwencja ma przynieść, czyli jej oczekiwany wpływ.

Często warto sformułować teorię zmiany dotyczącą danej interwencji, określając środki, dzięki którym poszczególne wkłady interwencji i działania mają prowadzić do wygenerowania produktów, wyników (lub rezultatów), a tym samym do oczekiwanego wpływu.

2.2.2. Jaki jest cel ewaluacji?

Przy opracowywaniu programu ewaluacji wpływu opartej na metodach kontrfaktycznych należy starannie rozważyć cel ewaluacji. Aby mieć pewność, że ewaluacja dostarczy wymaganych dowodów, należy w przede wszystkim zrozumieć, dlaczego jest ona niezbędna. W kontekście ewaluacji interwencji finansowanych z EFS należy rozważyć szereg kwestii:

- Jaki jest cel i charakter ewaluacji w kontekście wymogów regulacyjnych i wytycznych KE?
- Kim są główne zainteresowane strony ewaluacji?
- Do czego zostaną wykorzystane wyniki ewaluacji?
- Na jakie konkretne pytania ewaluacja będzie musiała odpowiedzieć?

Jakie są cele i charakter ewaluacji?

W pierwszej kolejności należy określić, jaki jest **powód** przeprowadzenia ewaluacji. Zgodnie z rozporządzeniem (UE) 2021/1060 „[w] celu poprawy jakości opracowywania programów i ich wdrażania państwo członkowskie lub instytucja zarządzająca przeprowadzają ewaluację programów pod kątem co najmniej jednego z następujących kryteriów: skuteczności, efektywności, trafności, spójności i unijnej wartości dodanej. Ewaluacje mogą również obejmować inne odpowiednie kryteria, takie jak poziom włączenia społecznego, niedyskryminację i widoczność, i mogą obejmować więcej niż jeden program”³⁰. Jak wykazano powyżej, ustalenia poczynione w wyniku ewaluacji wpływu opartej na metodach kontrfaktycznych dotyczą na ogół

Projektowanie CIE z wykorzystaniem logiki interwencji

Cel i charakter ewaluacji

³⁰ Art. 44 ust. 1 rozporządzenia (UE) 2021/1060 w sprawie wspólnych przepisów (RWP).

skuteczności (w jakim stopniu osiągnięto oczekiwane rezultaty) oraz efektywności (wskaźnika efektywności kosztowej interwencji).

W szerszym ujęciu KE zachęca państwa członkowskie do stosowania w kształtowaniu polityki podejścia zorientowanego na rezultaty oraz do przeprowadzania ewaluacji, które spełniają wewnętrzne zapotrzebowanie państw członkowskich pod względem ich zakresu zastosowania, koncepcji i ram czasowych. W związku z tym w programach EFS+ ewaluacje wpływu oparte na metodach kontrfaktycznych mogą być również realizowane w celu uzyskania odpowiedzi na konkretne pytania ewaluacyjne lub zgodnie z krajową polityką w zakresie ewaluacji.

Po drugie, **należy określić charakter ewaluacji**³¹:

- Ewaluacje **charakteru wpływu** pozwalają zbadać efekty programu lub grupy programów w odniesieniu do priorytetów unijnych i krajowych (może to być makroekonomiczny wpływ EFS, ukierunkowanie na konkretne polityki i tematy lub priorytety horyzontalne, takie jak dzieciństwo i równość szans).
- Ewaluacje dotyczące **procesu (lub wdrożenia)** sprzyjają wdrożeniu programu, ponieważ polegają na analizie postępów i metod wdrażania i na ich podstawie można opracować zalecenia dotyczące możliwości ulepszenia programu.

Zasadniczo podejście kontrfaktyczne można stosować do ewaluacji wpływu, natomiast ewaluacja procesu wymaga innych metod (zob. także Wykres 1 powyżej).

Rozporządzenie w sprawie wspólnych przepisów (RWP) nie wymaga przeprowadzenia konkretnej liczby ewaluacji wpływu w okresie programowania 2021–2027, w przeciwieństwie do okresu 2014–2020. Zawarto w nim natomiast wymóg opracowania strategii ewaluacyjnej umożliwiającej ocenę tego, w jaki sposób wsparcie z funduszy europejskich przyczyniło się do osiągnięcia celów programu w odniesieniu do wszystkich głównych profili strategicznych³². Ponadto w rozporządzeniu w sprawie wspólnych przepisów nie określono, na których priorytetach lub interwencjach powinna skupiać się ewaluacja, ale pozostawiono tę decyzję w gestii IZ, które mogą określić ją w planie ewaluacji. Oznacza to, że strategia ewaluacji każdego indywidualnego programu musi obejmować połączenie ewaluacji wpływu i ewaluacji procesu, a także wskazywać, na jakich priorytetach i interwencjach należy skoncentrować najważniejsze działania. Ostateczna i ogólna ewaluacja wpływu musi zostać opracowana do czerwca 2029 r.³³, ale nie wprowadzono żadnych innych ograniczeń dotyczących terminów przeprowadzenia pozostałych ewaluacji, które należy określić w planie ewaluacji.

Kto jest głównym odbiorcą?

Należy określić odbiorcę ewaluacji. W zależności od charakteru ewaluacji może ona obejmować decydentów, IZ i zarządzających programami, inne IZ lub organy wdrażające w państwie członkowskim bądź władze krajowe lub regionalne prowadzące podobne programy. W przypadku gdy dane

*Identyfikacja
zainteresowanych
stron*

³¹ [Zob. Komisja Europejska, 2007.](#)

³² Art. 44 ust. 1 rozporządzenia (UE) 2021/1060 w sprawie wspólnych przepisów.

³³ Art. 44 ust. 2 rozporządzenia (UE) 2021/1060 w sprawie wspólnych przepisów.

dostarczane są przez instytucje niebędące zarządzającymi programem, podmioty te („właściciele danych”) również należy uznać za zainteresowane strony. Ważne jest włączenie do grupy sterującej odpowiedzialnej za ewaluację wszystkich głównych zainteresowanych stron w celu ustanowienia współodpowiedzialności za proces projektowania i przeprowadzania ewaluacji, a także należy włączyć do niej kilku ekspertów w dziedzinie ewaluacji pochodzących z instytucji akademickich lub publicznych, tak aby zapewnić IZ doradztwo techniczne.

Do czego zostaną wykorzystane wyniki ewaluacji?

Po zidentyfikowaniu odbiorców ewaluacji można określić sposób wykorzystania jej wyników. W praktyce można to osiągnąć przez zaangażowanie grupy sterującej w opracowanie pytań ewaluacyjnych oraz w ramach omawiania specyfikacji istotnych warunków zamówienia.

Ewaluacja wpływu oparta na metodach kontrfaktycznych często umożliwia podjęcie dwóch kluczowych decyzji, mianowicie dotyczących:

- tego, czy istniejąca interwencja powinna być kontynuowana; oraz
- tego, czy nowy rodzaj interwencji powinien być wdrożony na szerszą skalę.

W pierwszym przypadku ewaluacja wpływu oparta na metodach kontrfaktycznych może posłużyć do oceny skuteczności istniejącego lub trwającego programu, w przypadku gdy budżet jest pod presją, a istnieją potencjalne alternatywne sposoby wykorzystania zaangażowanych zasobów. W tej sytuacji mało prawdopodobne jest, że interwencja była już wcześniej oceniana z wykorzystaniem scenariuszy kontrfaktycznych.

W drugim przypadku mogą występować ograniczenia we wdrażaniu interwencji. Interwencja może być na przykład wdrażana w określonym regionie lub na określonym obszarze państwa członkowskiego bądź może być wdrażana tylko przez ograniczony okres. W takich kontekstach wyniki ewaluacji wpływu opartej na metodach kontrfaktycznych można wykorzystać do ustalenia, czy dana interwencja jest skuteczna i czy w związku z tym użyteczne może być jej wdrożenie w innym miejscu. Interwencje w takim przypadku określa się jako pilotażowe lub testowe i są one realizowane przed wdrożeniem na szerszą skalę.

Na jakie pytania należy odpowiedzieć?

Po ustaleniu celów interwencji i założeń oraz ostatecznego zastosowania ewaluacji, a także po wyraźnym określeniu grupy odbiorców powinno być możliwe określenie szczegółowych pytań, na które ewaluacja wpływu oparta na metodach kontrfaktycznych będzie musiała odpowiedzieć. W wielu okolicznościach szereg odbiorców i zainteresowanych stron będzie miało pytania o charakterze przyczynowym, na które ich zdaniem ewaluacja wpływu oparta na metodach kontrfaktycznych ma odpowiedzieć. Zadaniem IZ lub ewaluatora jest zebrać te pytania w drodze szeroko zakrojonych konsultacji, z uwzględnieniem różnych punktów widzenia i uwag od osób zaangażowanych na różnych etapach interwencji. Następnie IZ lub ewaluator muszą ustalić, które pytania mają priorytetowe znaczenie, i ukierunkować ewaluację na ważniejsze i bardziej odpowiednie kwestie.

Przy ustalaniu listy kluczowych pytań badawczych na potrzeby ewaluacji wpływu opartej na metodach kontrfaktycznych pod uwagę można wziąć niektóre z następujących kwestii:

Główne pytania badawcze

- Czy interwencja przyniosła zamierzone wyniki lub przyczyniła się do ich osiągnięcia w krótkiej, średniej i długiej perspektywie? Czy efekty krótkoterminowe znacznie różniły się od tych długoterminowych? Pytania odnoszące się do tych kwestii powinny być traktowane priorytetowo.
- Czy możliwe jest uzyskanie ilościowego pomiaru wyników? Ewaluacja wpływu oparta na metodach kontrfaktycznych musi opierać się na odpowiednim zbiorze danych (dane administracyjne lub bezpośrednie badania wysyłane do uczestników), niezależnie od preferowanego źródła.
- W jakim stopniu zmiany w sytuacji uczestników lub w kontekście społeczno-gospodarczym można przypisać interwencjom? Ten rodzaj pytania wymaga pomiaru efektu netto interwencji na potrzeby porównania z grupą kontrolną i wskaźnikami kontekstowymi.
- Czy efekty interwencji były takie same dla wszystkich członków grupy docelowej? Na przykład czy wpływ interwencji skierowanej do długotrwale bezrobotnych był taki sam dla mężczyzn i kobiet? Czy był taki sam dla osób poniżej 25. lub powyżej 50. roku życia? Zdolność do zbadania efektów w różnych podgrupach za pomocą ewaluacji wpływu opartej na metodach kontrfaktycznych będzie duża, jeśli zapewni się odpowiednio wysoką liczbę osób w grupie objętej interwencją i grupie kontrolnej.
- Czy interwencja była efektywna kosztowo (w porównaniu z alternatywnymi rozwiązaniami)? Jaki jest jej stosunek kosztów do korzyści? Aby można było przeprowadzić taką analizę efektywności, ewaluacja wpływu oparta na metodach kontrfaktycznych musi umożliwić dokładny pomiar efektów, jak również dostarczyć precyzyjnych informacji o bezpośrednich i pośrednich kosztach i korzyściach.
- Czy dostępna jest wystarczająca ilość informacji i wiedzy na temat wpływu podobnych interwencji? W jakim stopniu wiedza ta ma zastosowanie do badanej interwencji? W przypadku gdy dana polityka jest dobrze znana, a jej skutki zostały szeroko zbadane, użyteczne może być skoncentrowanie pytań ewaluacyjnych na konkretnych aspektach polityki, co pozwoli uniknąć powtarzania innych analiz. Szczegółowy przegląd dostępnej literatury może pomóc w podjęciu decyzji w tym względzie, a w każdym przypadku stanowi użyteczne wsparcie przy projektowaniu ewaluacji wpływu opartej na metodach kontrfaktycznych.

Ważne jest, aby przed zleceniem ewaluacji mieć jasny obraz zakresu pytań badawczych, które mają być przeanalizowane w ramach ewaluacji wpływu opartej na metodach kontrfaktycznych. Kluczowym elementem każdego programu ewaluacji jest omówienie zagadnień, które należy uwzględnić w ewaluacji.

Należy ustalić, które pytania mają priorytetowe znaczenie, tak aby nie popełnić częstego błędu i nie przeciążać ewaluacji zbyt dużą liczbą problemów, które mają być przeanalizowane. Ciężko jest sprawić, aby ewaluacja odpowiadała potrzebom różnych zainteresowanych stron o różnych interesach, a jednocześnie nie była zbyt rozbudowana. Przyjmując zbyt szeroki zakres pytań badawczych do zbadania w ramach ewaluacji, można stracić z oczu najważniejsze kwestie i w efekcie próbować rozpatrywać wiele różnych problemów w niewystarczający sposób. Często

Wskazanie pytań o priorytetowym znaczeniu

„mniej znaczy więcej”: ustalanie priorytetów jest najważniejszym etapem procesu planowania ewaluacji.

Aby ustalić, które pytania ewaluacyjne mają priorytetowe znaczenie, konieczne jest wyeliminowanie powtarzających się pytań i przypisanie każdemu z pytań wagi zgodnie z odpowiednimi zasadami. Mogą one obejmować: znaczenie i rzeczywiste zaangażowanie zainteresowanej strony, która sformułowała dane pytanie, odpowiednie dopasowanie i zgodność pytania z teorią zmiany dotyczącą programu, trafność pytania w odniesieniu do ogólnego celu ewaluacji, możliwość udzielenia odpowiedzi na pytanie na podstawie dostępnych danych, w określonym czasie i za pomocą dostępnych zasobów³⁴. Na tej podstawie uporządkowuje się pytania według ważności, co pozwala wyłonić te najbardziej istotne.

W niektórych przypadkach bardziej ogólne pytanie ewaluacyjne można uzupełnić kilkoma dodatkowymi pytaniami. Na przykład pytanie „jaki był efekt netto interwencji” można uzupełnić dalszymi pytaniami, jak choćby o wpływ na różne grupy uczestników, może nawet o wpływ w różnych okresach. Proces ten musi jednak zawsze prowadzić do opracowania takiego zestawu pytań, aby udzielenie na nie odpowiedzi było wykonalne.

Jakie kryteria ewaluacji mogą być powiązane z ewaluacją?

Wspomniano już o związku między kryteriami ewaluacji a pytaniami ewaluacyjnymi, ale warto to zagadnienie dokładnie omówić. Kryteria ewaluacji (efektywność, skuteczność, europejska wartość dodana, spójność itd.) są niezbędne do przypisania wartości wszelkim zestawionym dowodom i dokonania oceny polityki (efektywność itp.); natomiast pytania ewaluacyjne są niezbędne do tego, aby jasno określić wymagania zlecniodawcy i skupić się na głównych kwestiach politycznych. Kryteria ewaluacji i pytania ewaluacyjne są jednak ze sobą powiązane. Każde pytanie można zasadniczo odnieść do konkretnego kryterium i ta zależność jest ważna zarówno dla IZ, jak i ewaluatorów, ponieważ łączy projekt ewaluacji, oparty na pytaniami, z wymaganiami zawartymi w rozporządzeniu w sprawie wspólnych przepisów na lata 2021–2027, które są związane z kryteriami ewaluacji w inny sposób.

Kryteria ewaluacji i pytania ewaluacyjne

Poniżej przedstawiono kilka przykładów typowych pytań ewaluacyjnych pogrupowanych wraz z odpowiadającymi im kryteriami. Zostały one wybrane i opracowane na podstawie „Zestawu instrumentów służących lepszemu stanowieniu prawa”³⁵ (Narzędzie nr 47 *Kryteria ewaluacji i pytania ewaluacyjne*) Komisji Europejskiej w celu pokazania przykładów zgodnych z pytaniami, które wymagają przeprowadzenia analizy wpływu oraz, tam gdzie to możliwe, ewaluacji wpływu opartej na metodach kontrfaktycznych:

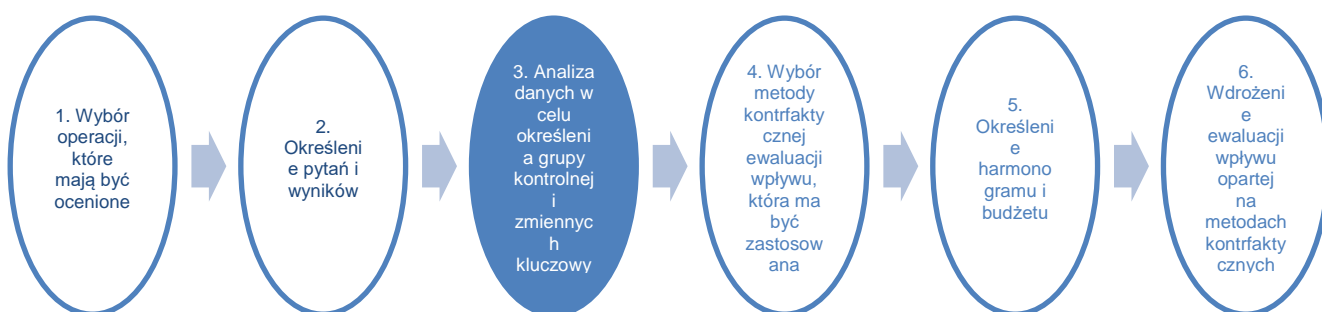
- *Typowe przykłady pytań dotyczących skuteczności*
 - Jakie były efekty ilościowe interwencji?
 - W jakim stopniu te zmiany/efekty można przypisać interwencji?
 - W jakim stopniu czynniki wpływające na zaobserwowane osiągnięcia można powiązać z interwencją UE?
- *Typowe przykłady pytań dotyczących efektywności*
 - W jakim stopniu interwencja była efektywna kosztowo?
 - W jakim stopniu koszty interwencji są uzasadnione, biorąc pod uwagę osiągnięte zmiany/efekty?

³⁴ Zob. na przykład [Centrum ds. Zapobiegania i Kontroli Chorób, 2013](#).

³⁵ Zob. [Komisja Europejska, 2017](#).

- Jeśli istnieją znaczące różnice w kosztach (lub korzyściach) między terytoriami, co jest ich przyczyną? Jaki jest związek między tymi różnicami a interwencją?
- *Typowe przykłady pytań dotyczących europejskiej wartości dodanej*
 - Jaka jest dodatkowa wartość wynikająca z interwencji w ramach EFS+ w porównaniu z wynikami podobnych interwencji krajowych lub regionalnych?
 - Jakie byłyby najbardziej prawdopodobne skutki wstrzymania lub wycofania istniejącej interwencji w ramach EFS+?

2.3. Dane do określenia grupy kontrolnej i pomiaru zmiennych wynikowych



2.3.1. Czy odpowiednie dane są lub mogą być dostępne?

Jak wynika z rozmów przeprowadzonych z instytucjami zarządzającymi i ekspertami w dziedzinie ewaluacji z całej UE, jednym z kluczowych wyzwań podczas wdrażania ewaluacji wpływu opartych na metodach kontrfakcyjnych jest dostęp do odpowiednich danych. Z kolei kluczową kwestią praktyczną jest dostępność takich danych. W tej sekcji przedstawiono uproszczoną kategoryzację wymaganych rodzajów danych, wraz z omówieniem źródeł, z których można pozyskać takie dane, a także rodzajów czynności gromadzenia danych pierwotnych, które mogą być wymagane. Poruszono również niezwykle ważną kwestię, jaką jest ochrona danych.

Warto zwrócić uwagę na jedną istotną kwestię dotyczącą odpowiedniego planowania. W przeszłości próby wdrożenia ewaluacji wpływu opartych na metodach kontrfakcyjnych były w pewnym stopniu udaremniane z powodu braku danych, ponieważ odpowiednie plany nie zostały wdrożone wystarczająco wcześnie. W przypadku istniejących interwencji ważne jest, aby zidentyfikować członków grupy poddanej interwencji i grupy niepoddanej interwencji oraz aby ustanowić mechanizmy gromadzenia danych pochodzących od tych grup, ponieważ to na nich będzie się koncentrować ewaluacja. W przypadku nowych interwencji należy podjąć wczesne działania w celu zapewnienia, aby właściwe rodzaje danych zostały zgromadzone w odpowiednim czasie.

Planowanie gromadzenia danych

Jakie rodzaje danych są wymagane?

Zasadniczo do przeprowadzenia ewaluacji wpływu opartej na metodach kontrfakcyjnych potrzebne są trzy rodzaje danych. W niektórych przypadkach jedno źródło danych może zawierać większą liczbę rodzajów danych. Są to: dane dotyczące grupy objętej interwencją i grupy kontrolnej, dane dotyczące wyników oraz dane kontekstowe.

Główne źródła danych

- **Dane dotyczące grupy objętej interwencją i grupy kontrolnej:** wymagane są źródła danych, które umożliwią ewaluatorom identyfikację poszczególnych jednostek grupy objętej interwencją i grupy kontrolnej (przedsiębiorstw, osób lub potencjalnie obszarów geograficznych).
- **Dane dotyczące wyników:** jak pokazano na wykresach 6 i 7 w rozdziale 3 niniejszego przewodnika, ewaluacje wpływu oparte na metodach kontrfaktycznych wymagają pomiaru wyników zarówno w odniesieniu do grupy objętej interwencją, jak i grupy kontrolnej. Najlepiej, gdyby dane dotyczące wyników obu grup były gromadzone przy użyciu tych samych metod gromadzenia danych oraz aby pomiarów dokonywano w tym czasie.
- **Dane kontekstowe:** dane te są niezbędne, ponieważ umożliwiają wybór pasujących do siebie grupy kontrolnej i grupy objętej interwencją, a także przeanalizowanie wszelkich pozostających różnic między nimi. Ważne jest, aby zebrać możliwie jak największą ilość danych na temat cech charakterystycznych jednostki i czynników, które mogą być związane zarówno z decyzją o udziale w interwencji, jak i z potencjalnymi rezultatami, szczególnie wskaźnikami rezultatu mierzonymi przed rozpoczęciem interwencji. Dane kontekstowe mogą również obejmować dane opisujące lokalne rynki pracy (np. lokalne stopy bezrobocia lub ocenę elastyczności rynku pracy) oraz te, które umożliwią analizę w podziale na podgrupy.

W tabeli 2 poniżej przedstawiono te trzy rodzaje danych i wskazano potencjalne źródła, z których można takie dane pozyskać. W pierwszej ramce poniżej przedstawiono przykłady danych wykorzystywanych na potrzeby ewaluacji wpływu opartych na metodach kontrfaktycznych w odniesieniu do EFS, natomiast w drugiej ramce zilustrowano przykłady zintegrowanego zbioru danych wykorzystywanego w analizie kwestii związanych z rynkiem pracy oraz w ewaluacjach polityki rynku pracy.

Tabela 2. Rodzaje i źródła danych

Rodzaje danych	Źródła
Dane dotyczące grupy objętej interwencją	<ul style="list-style-type: none"> - Dane dotyczące uczestnictwa w interwencji (zasadniczo w posiadaniu beneficjentów) - Dane z monitorowania EFS+ (charakterystyka interwencji, rozpoczęcie i zakończenie, dane odnośne, rejestr wniosków)
Dane dotyczące grupy kontrolnej	<ul style="list-style-type: none"> - Dane administracyjne, takie jak systemy ewidencji zabezpieczenia społecznego, rejestry dotyczące wykształcenia i rejestry osób pobierających zasiłek dla bezrobotnych - Dane dotyczące uczestnictwa (osoby, które kwalifikowały się do uczestnictwa, ale nie uczestniczyły z powodów innych niż kwalifikowalność³⁶) - Istniejące badania krajowe, takie jak BAEL
Dane dotyczące wyników (wymagane zarówno w odniesieniu do grupy objętej interwencją, jak i grupy kontrolnej)	<ul style="list-style-type: none"> - Dane administracyjne: na przykład dane dotyczące zabezpieczenia społecznego i bezrobocia mogą być również wykorzystywane do łączenia pomiarów wyników (dane dotyczące wypłaty świadczeń/zabezpieczenia społecznego), krajowe dane dotyczące zabezpieczenia społecznego i podatków (zarobki i status zatrudnienia) - Dane administracyjne od instytucji edukacyjnych (standardowe testy osiągniętych kompetencji, odsetek absolwentów, współczynniki skolaryzacji i dane dotyczące frekwencji) - Oficjalny spis przedsiębiorstw lub dane podatkowe (dotyczące produktywności lub obrotu przed szkoleniem wewnętrznym i po jego zakończeniu lub dotyczące liczby nowych pracowników) - Dane dotyczące zatrudnienia lub produkcji pochodzące z oficjalnych statystyk (w przypadku terytorialnych analiz konfraktycznych w celu zmierzenia poziomów zatrudnienia i PKB) - Odpowiednio dostosowane do potrzeb badania w grupie objętej interwencją i grupie kontrolnej
Dane kontekstowe lub zmiennie kontrolne (wymagane zarówno w odniesieniu do grupy objętej interwencją, jak i grupy kontrolnej)	<ul style="list-style-type: none"> - Systemy administracyjne – np. dane dotyczące świadczeń zawierające informacje na temat historii pobierania świadczeń przed interwencją Krajowe dane dotyczące ubezpieczeń i podatków, rejestry zarobków i zatrudnienia - Oficjalne statystyki dotyczące rynku pracy lub wykształcenia (np. badanie aktywności ekonomicznej ludności, podstawowe dane na poziomie regionalnym lub krajowym, które również dostarczają danych jednostkowych na poziomie indywidualnym do szczegółowego opracowania) - Badania wśród grupy kontrolnej i grupy objętej interwencją; w przypadku gdy zasady interwencji są jasno określone, grupy kontrolne można określić <i>ex ante</i> i zebrać dane podstawowe - Systemy monitorowania – w niektórych okolicznościach systemy monitorowania mogą być wykorzystywane do gromadzenia danych na potrzeby pomiarów zarówno od grupy objętej interwencją, jak i grupy kontrolnej, np. systemy składania wniosków, w których wnioskodawcy, którzy nie przeszli procesu pomyślnie, mogą być wykorzystani do utworzenia grupy kontrolnej.

³⁶ Są to uczestnicy „kwalifikujący się, ale nie przyjęci”; niedopuszczenie do udziału w projekcie wynika zazwyczaj z zakończenia finansowania lub z innych przyczyn zewnętrznych (np. choroba nauczyciela, paraliż komunikacyjny itp.). Uczestnicy ci mogą stanowić preferencyjną grupę kontrolną, ponieważ kwalifikują się tak samo i mają taką samą wolę uczestnictwa jak faktyczni uczestnicy; czasami jest ich jednak zbyt niewiele, aby można było z nich utworzyć grupę kontrolną, lub ich dane nie są rejestrowane.

Ramka 6. Przykłady danych wykorzystywanych do ewaluacji wpływu opartych na metodach kontryfaktycznych

Wśród przykładów szczegółowo przeanalizowanych w niniejszym przewodniku niemieckie ewaluacje są prawdopodobnie tymi, które opierają się na najbogatszych administracyjnych zbiorach danych. Zintegrowane Historie Zatrudnienia (IEB) łączą szereg różnych administracyjnych źródeł informacji zarządzanych przez Federalną Agencję Pracy i zawierają informacje na temat zatrudnienia danej osoby (z wyjątkiem samozatrudnienia) oraz jej okresów bezrobocia, cech społeczno-demograficznych, płatności transferowych (zasiłków dla bezrobotnych) oraz uczestnictwa w aktywnej polityce rynku pracy. IEB mają długą historię, a opracowywanie na ich potrzeby zbiorów danych i zarządzanie nimi wymaga ciągłych nakładów pracy i inwestycji. Więcej szczegółów na temat IEB przedstawiono w ramce 7³⁷. W niemieckiej ewaluacji programów tworzenia miejsc pracy dane te zostały uzupełnione badaniem przeprowadzonych w trzech falach w celu zebrania informacji na temat „miękkich wyników” (postrzeganych przez zainteresowane osoby) na różnych etapach interwencji.

We włoskich ewaluacjach szczegółowo omówionych w niniejszym przewodniku użyto podobnych danych. Archiwa osób bezrobotnych zarejestrowanych w publicznych służbach zatrudnienia zostały wykorzystane w celu identyfikacji potencjalnych grup kontrolnych, natomiast dane administracyjne dotyczące zmian w zatrudnieniu (Comunicazioni obbligatorie, COB), zawierające rejestr wszystkich umów o pracę zawartych przez firmy i pracodawców publicznych (z wyłączeniem samozatrudnienia), zostały wykorzystane do pomiaru zmiennych wynikowych. Jakość informacji jest zróżnicowana, ponieważ COB są zarządzane na poziomie regionalnym. W ciągu ostatnich kilku lat wprowadzono jednak ogólnokrajowe standardy, dzięki czemu jakość się poprawiła. W jednej z ewaluacji (dla prowincji Trydentu) ewaluator mógł także połączyć te dane z danymi dotyczącymi deklaracji podatkowych dostarczonymi przez INPS (krajowy instytut zabezpieczenia społecznego), dzięki czemu można było zmierzyć wpływ w odniesieniu do zarobków danej osoby.

W Polsce w ewaluacjach wykorzystywano dane administracyjne pochodzące z rejestrów publicznych służb zatrudnienia dotyczących osób bezrobotnych, aby zidentyfikować grupy kontrolne. Dane te nie zawierają jednak informacji na temat statusu ani historii zatrudnienia poszczególnych osób, a ewaluacje musiały opierać się na danych zastępczych (anulowanie statusu bezrobotnego w publicznych służbach zatrudnienia) lub wymagały gromadzenia informacji za pomocą badań na próbie osób, zarówno poddanych, jak i niepoddanych interwencji.

Na Łotwie ewaluator miał możliwość połączenia dwóch administracyjnych zbiorów danych. Dane z łotewskiej Państwowej Agencji Pracy (SEA) dostarczyły informacje zarówno o uczestnikach, jak i osobach nieuczestniczących w interwencji, zarejestrowanych jako osoby bezrobotne w określonym czasie, a dane z krajowego urzędu skarbowego (SRS) dostarczyły informacje na temat statusu zatrudnienia w różnych okresach, jak również na temat dochodów poszczególnych osób. Pozwoliło to ewaluatorowi ocenić efekty zarówno pod względem prawdopodobieństwa otrzymania zatrudnienia w różnych okresach, jak również pod względem dochodów.

W przykładzie szwedzkim – jedynym wśród opisywanych przypadków przykładem ewaluacji z wykorzystaniem randomizacji – wykorzystano zarówno dane administracyjne do pomiaru zmiennych wynikowych, jak i badanie przeprowadzone wśród urzędów i pośredników pracy w celu zmierzenia intensywności i rodzajów udzielonego wsparcia.

W ewaluacjach w dziedzinie edukacji dane administracyjne wykorzystano do zidentyfikowania grupy kontrolnej i oceny zmiennych wynikowych – z wyjątkiem Polski, gdzie zmienne wykorzystane jako wyniki trzeba było zmierzyć na podstawie badań.

Ogólnie rzecz biorąc, powyższe przykłady uwypuklają znaczenie odpowiednich danych administracyjnych dla ewaluacji wpływu opartych na metodach kontryfaktycznych – optymalnie byłoby wykorzystywać te dane zarówno do identyfikacji grup kontrolnych, jak i pomiaru zmiennych wynikowych. Z uwagi na powyższe IZ planujące ewaluacje wpływu oparte na metodach kontryfaktycznych powinny uprzednio zapewnić, aby dane administracyjne potrzebne do przeprowadzenia ewaluacji były dostępne, i podjąć odpowiednie działania w

³⁷ Podobne doświadczenia wynikają z irlandzkiej bazy danych „Jobseekers Longitudinal Dataset” (JLD) (Zbiór danych wzdłużnych dotyczących osób poszukujących pracy), przedstawiającej informacje dotyczące płatności na pomoc społeczną i ubezpieczenie społeczne, programów rynku pracy oraz przebiegu zatrudnienia w jednej wspólnej bazie danych. Baza danych JLD została wykorzystana między innymi do ewaluacji programu aktywizacji JobBridge finansowanego w okresie programowania 2007–2013. Zob. [Indecon, 2016](#).

celu rozwiązania potencjalnych problemów związanych z dostępnością, integracją danych lub innymi problematycznymi aspektami.

Ramka 7. Przykłady zintegrowanych baz danych na potrzeby ewaluacji wpływu opartych na metodach kontryfakcyjnych

Zbiór danych wzdłużnych dotyczących osób poszukujących pracy (JLD) w Irlandii

Zbiór danych wzdłużnych dotyczących osób poszukujących pracy (JLD) jest administracyjną bazą danych zarządzaną przez Departament Ochrony Socjalnej (DSP).

Stanowi ambitną próbę dostosowania danych administracyjnych do celów badawczych. Prace nad JLD rozpoczęły się około 10 lat temu, po tym jak DSP zlecił University College w Dublinie przeprowadzenie wstępnych badań dotyczących zarządzania rejestrem danych Live Register³⁸ oraz – bardziej ogólnie – danymi dotyczącymi rynku pracy. Badanie to pozwoliło wskazać szereg sugestii dotyczących usprawnienia gromadzenia danych i ujawniło pewne kwestie problematyczne (np. powielanie danych w różnych systemach danych, brakujące informacje itp.).

JLD łączy kilka źródeł informacji: dane dotyczące płatności i dane administracyjne pochodzące z DSP – np. dotyczące płatności na pomoc społeczną i ubezpieczenie społeczne dla ludności w wieku produkcyjnym uwzględnionej w Live Register oraz dane związane z programami aktywnej polityki rynku pracy zarządzanymi przez DSP; dane dotyczące programów rynku pracy zarządzanych przez SOLAS, krajowy organ organizujący programy kształcenia i szkolenia; dane zgromadzone przez organy podatkowe (komisarzy ds. dochodów). W odniesieniu do interwencji w ramach EFS JLD obejmuje częściowo interwencje finansowane z funduszu, ale koncentruje się bardziej na tych finansowanych przez DSP i SOLAS³⁹.

Dane z wyżej wymienionych źródeł są uporządkowane jako seria epizodów, przy czym jeden epizod rozpoczyna się, gdy dana osoba zaczyna okres bezrobocia, a kończy się, gdy osoba ta otrzymuje zatrudnienie lub rozpoczyna inną formę działalności bądź program szkoleniowy; następnie, gdy status jednostki ponownie się zmienia, rozpoczyna się nowy epizod. Początek okresu bezrobocia zbiega się z rozpoczęciem przez daną osobę ubiegania się o świadczenie i zasiłek dla osób poszukujących pracy⁴⁰. W JLD zaletą tej struktury składającej się z epizodów jest to, że występujące bezpośrednio po sobie okresy pobierania zasiłku i świadczeń dla osób poszukujących pracy mogą być połączone i występować jako jeden epizod bezrobocia. Warto zauważyć, że epizody, jak wyjaśniono powyżej, mogą się pokrywać, a badacz ma trudności z ustaleniem, który z epizodów ma większe znaczenie.

Od 2004 r. JLD monitoruje wnioski o pomoc społeczną, historie udziału w programach aktywizacji, szkoleniach i historii zatrudnienia. Baza zawiera około 13 milionów indywidualnych epizodów związanych z opieką społeczną i pracą dotyczących około 2 milionów osób. Każdy epizod ma datę rozpoczęcia i zakończenia oraz kod operacyjny⁴¹, który pozwala na identyfikację sytuacji danej osoby w każdym okresie. JLD obejmuje duży zestaw „zmiennych”: płeć, wiek, stan cywilny, obywatelstwo, poziom wykształcenia, poprzedni zawód, historię zatrudnienia i bezrobocia (czas trwania i liczba epizodów) oraz charakterystykę miejsc pracy (na przykład sektor), historię szkoleń dla bezrobotnych (rodzaj, czas trwania i liczba epizodów), rodzaj świadczeń (świadczenie/zasiłek dla osób poszukujących pracy), liczba dzieci, liczba osób pozostających na utrzymaniu, rodzaj płatności rodzinnych (tj. zasiłek dla dorosłych i dzieci na utrzymaniu, tylko dla dorosłych itp.), zarobki i podatki oraz położenie geograficzne⁴².

³⁸ Live Register zawiera informacje na temat osób rejestrujących się w lokalnych biurach DSP w celu otrzymania świadczeń dla osób poszukujących pracy lub zasiłku dla osób poszukujących pracy bądź innych uprawnień określonych ustawowo.

³⁹ Programy w obszarach, takich jak wymiar sprawiedliwości (Youth Diversion Project, Young Persons Probation Projects oraz Integration and Employment of Migrants), edukacja (Third Level Access oraz Adult Literacy), wspólnota (Community Training Centres), a także obrona (Defence Forces Employment Support Scheme) nie są uwzględnione w JLD.

⁴⁰ Świadczenie dla osób poszukujących pracy to cotygodniowa wypłata z DSP dla osób, które nie mają pracy, są całkowicie bezrobotne lub pracują w niepełnym wymiarze czasu pracy, ponieważ ich godziny pracy zostały zredukowane przez pracodawcę. Ma ono zastosowanie do osób poszukujących pracy, które płaciły składki na ubezpieczenie społeczne (PRSI) w odpowiedniej wysokości, w odpowiednim roku podatkowym wpłaciły wystarczającą liczbę składek i łącznie wpłaciły minimum 104 składki. Zasiłek dla osób poszukujących pracy jest świadczeniem uzależnionym od wysokości dochodów dla osób, które są bezrobotne i nie kwalifikują się do otrzymywania świadczenia dla osób poszukujących pracy lub których prawo do tego świadczenia wygasło.

⁴¹ Wskazujący kombinację działań aktywizacyjnych/szkoleniowych, wniosków o pomoc społeczną i czasu trwania okresu zatrudnienia.

⁴² Warto zwrócić uwagę na fakt, że niektóre dane odnoszą się do różnych okresów, na przykład zarobki są rejestrowane w ujęciu

Ponadto indywidualne identyfikatory łączą JLD z innymi danymi administracyjnymi; na przykład na potrzeby ewaluacji programu JobsPlus JLD została sparowana z oddzielną bazą danych z monitorowania, zawierającą szczegółowe informacje na temat daty rozpoczęcia lub zakończenia udziału w programie JobsPlus, liczby dni w Live Register na początku programu oraz rodzaju interwencji.

Istnieją dwa główne kanały dostępu do danych JLD: badania na zlecenie, które mają miejsce, gdy DSP uruchamia zaproszenie do składania ofert (RFT); lub na wniosek badaczy dysponujących odpowiednimi uprawnieniami. W tym drugim przypadku wnioski o dostęp są rozpatrywane indywidualnie, a w przypadku ich akceptacji musi zostać podpisana możliwa do wyegzekwowania na drodze prawnej umowa o udostępnianiu danych między badaczami/instytucjami a DSP. Wszystkie dane poddaje się pseudonimizacji i przekazuje się tylko te pola danych, które są istotne dla danego tematu badań. Dostęp do danych dla badacza lub instytucji badawczej jest ograniczony, a dane muszą zostać usunięte po zakończeniu projektu badawczego lub ewaluacyjnego. Wraz z danymi JLD badacze otrzymują klucz kodowy, chociaż, jeśli chodzi o dokumentację, konieczna jest znaczna poprawa.

Baza JLD była przez lata wykorzystywana do różnych działań badawczych i ewaluacji: Back to Education Allowance (2015) – dodatek dla osób wracających do nauki, program aktywacyjny JobBridge (2016), Back to Work Enterprise Allowance (2017) – dodatek na rzecz przedsiębiorczości i powrotu do pracy, programy JobPath (2019) i JobPlus (2020) to przykłady projektów, które zostały wybrane do ewaluacji.

Struktura i zawartość JLD uległy poprawie w porównaniu z początkiem istnienia bazy. Baza wymaga jednak ciągłych działań w celu zapewnienia jej aktualizacji oraz opracowywania ulepszeń i wypełniania luk, na przykład przez dodawanie bardziej szczegółowych informacji na temat wykształcenia oraz włączenie większej ilości szczegółowych danych dotyczących programu (treść, ukończenie itp.) i poprawę jakości danych dotyczących zarobków. W ramach nadchodzącego projektu należy wdrożyć ulepszenia JLD w zakresie: regularnej aktualizacji (np. co miesiąc) wraz z automatyzacją i testowaniem systemu oraz dodawaniem lub zastępowaniem źródeł danych w celu zapewnienia pełnego uwzględnienia wsparcia w zakresie zatrudnienia i szkoleń, innych programów ochrony socjalnej oraz danych o zarobkach w czasie rzeczywistym⁴³.

Instytut Badań nad Zatrudnieniem (IAB) w Niemczech i jego Zintegrowane Historie Zatrudnienia (IEB)

W Niemczech poprawiono dostęp do danych do celów naukowych w wyniku reform rynku pracy z lat 2003–2005, w których położono nacisk na ewaluację polityki oraz na zalecenia niemieckiej komisji ds. poprawy infrastruktury informacyjnej łączącej sektor nauki i danych statystycznych, która rekomendowała utworzenie ośrodków przetwarzania danych badawczych przy każdym publicznym podmiocie dostarczającym dane jednostkowe. W odpowiedzi na zalecenia w 2004 r. Federalna Agencja Pracy utworzyła ośrodek przetwarzania danych badawczych w ramach Instytutu Badań nad Zatrudnieniem (Institut für Arbeitsmarkt- und Berufsforschung (IAB)). IAB jest odpowiedzialny za pozyskiwanie danych z procesów administracyjnych w celu przygotowania zbiorów danych przydatnych na potrzeby badań empirycznych.

IAB zarządza ponad 15 zbiorami danych, które są dostępne dla społeczności naukowej. Dane pochodzą z danych administracyjnych z systemu zabezpieczenia społecznego, wewnętrznych procesów Federalnej Agencji Pracy oraz z badań przeprowadzonych przez IAB⁴⁴.

W odniesieniu do systemów zabezpieczenia społecznego wszyscy pracodawcy są zobowiązani do zgłaszania szeregu elementów danych i cech charakterystyki swoich pracowników, a dane te dostarczają bogatego zestawu informacji na temat historii zatrudnienia. Dane administracyjne związane z wewnętrznymi procedurami Federalnej Agencji Pracy obejmują obowiązkowe ubezpieczenie od utraty pracy, zasiłki dla bezrobotnych i odpowiadające im okresy przysługiwania tego świadczenia, spotkania doradcze z osobami bezrobotnymi, oferty pośrednictwa pracy oraz aktywne instrumenty rynku pracy. IAB łączy te dane w jeden kompleksowy zbiór danych o nazwie „Zintegrowane Historie Zatrudnienia (IEB)”. Gromadzenie tych danych administracyjnych rozpoczęło się w 1975 r., choć nie wszystkie zmienne są

rocznym, podczas gdy płatności świadczeń socjalnych są aktualizowane co tydzień.

⁴³ Pragniemy podziękować Frankowi Humphreysowi, Ciaranowi Judge'owi, Saidhbhín Hardiman i Krzysztofowi Gigonowi z DSP za przekazanie informacji na temat JLD.

⁴⁴ W 2011 r. w IAB powstał Ośrodek Łączenia Zapisów – projekt realizowany we współpracy z Uniwersytetem Duisburg-Essen, finansowany przez Niemiecką Fundację Badawczą; zadaniem ośrodka jest uproszczenie łączenia zbiorów danych bez określonego identyfikatora.

dostępne w odniesieniu do całego okresu obserwacji. IEB mogą być zintegrowane z szeregiem danych pochodzących z badań realizowanych przez IAB, takich jak: „IAB Establishment Panel” – panel zakładów pracy prowadzony przez IAB, „German Job Vacancy Survey” – badanie IAB dotyczące wolnych miejsc pracy w Niemczech, „German Management and Organizational Practices (GMOP)” – badanie dotyczące niemieckich praktyk w zakresie zarządzania i organizacji; „Panel Study Labour Market and Social Security (PASS)” – badanie panelowe dotyczące rynku pracy i zabezpieczenia społecznego; „Working and Learning in a Changing World” – praca i nauka w zmieniającym się świecie; „Employee survey: bonus payments, wage increases, and fairness (BLoG)” – badanie pracowników dotyczące premii, wzrostu płac i sprawiedliwości; „Linked Employer–Employee Data from the IAB (LIAB)” – połączone dane IAB dotyczące pracodawców i pracowników; Panel „WeLL” – badanie pracowników na potrzeby projektu „Further Training as a Part of Lifelong Learning” („Dalsze szkolenie jako element uczenia się przez całe życie”).

IAB regularnie aktualizuje swoje produkty wytworzone z danych i oferuje różne próbki tych bogatych źródeł danych administracyjnych do celów badawczych. Do każdego produktu wytworzonego z danych IAB dołącza szczegółową dokumentację w języku niemieckim i angielskim. Podstawę prawną dostępu do danych określono w niemieckim kodeksie socjalnym (szereg wersji). Mówiąc konkretnie, przewidziano w nim cztery rodzaje dostępu do danych dla społeczności naukowej: pliki Campus (w pełni zanonimizowane i przydatne jedynie do celów dydaktycznych); pliki do użytku naukowego – zanonimizowane dane jednostkowe przekazywane instytucjom naukowym w Niemczech i państwach członkowskich UE do celów projektów badawczych w dziedzinie badań rynku pracy, ale nie do celów dydaktycznych czy komercyjnych; bezpieczeństwo danych musi być zagwarantowane przez instytucję naukową ubiegającą się o dostęp do tych danych; słabo zanonimizowane dane z bardziej szczegółowymi informacjami są dostępne tylko na miejscu. IAB zapewnia wydzielone miejsca pracy w bezpiecznym środowisku komputerowym w Norymberdze oraz w różnych miejscach w Niemczech, Stanach Zjednoczonych i Wielkiej Brytanii. Badacze mają bezpośredni dostęp do danych, ale mogą uzyskać wyniki swoich programów dopiero po ich weryfikacji pod kątem ujawnianych informacji przez pracowników IAB; dostęp zdalny oznacza, że naukowcy przygotowują swoje programy, używając sztucznych danych i przesyłają je do systemu Job Submission Application (JoSuA). W tym procesie naukowcy nigdy nie mają wglądu w oryginalne dane i otrzymują jedynie swoje wyniki. W przypadku każdego dostępu do danych dostępne są znormalizowane formularze wniosków. Po zatwierdzeniu wniosku instytucja badawcza zawiera umowę z IAB regulującą wykorzystanie danych na potrzeby konkretnego projektu w określonym czasie. W umowie określa się przepisy o ochronie danych, a w przypadku ich naruszenia stosuje się surowe sankcje. Niektóre zbiory danych zarządzane przez IAB są dostępne wyłącznie do użytku na miejscu (na przykład połączone zbiory danych).

Na podstawie danych opracowanych przez IAB przeprowadza się regularne ewaluacje aktywnej polityki rynku pracy przy użyciu najnowszych metod empirycznych. W niektórych przypadkach wyniki takich ewaluacji doprowadziły do zmiany polityki⁴⁵. IAB zapewnia naukowcom dostęp do swoich zbiorów danych nie tylko w Niemczech, ale również w innych krajach, a liczba ich użytkowników stale rośnie (na przykład w 2016 r. prawie jedna trzecia wszystkich umów o wykorzystanie danych zawierana była z ośrodkami innymi niż niemieckie).

Aby dowiedzieć się więcej na temat wykorzystania danych administracyjnych na potrzeby ewaluacji wpływu opartych na metodach kontryfaktycznych i związanych z tym kwestii praktycznych, zob. również europejski przewodnik przygotowany przez badaczy z JRC, [Komisja Europejska, 2020](#).

⁴⁵ Przykładem może być ewaluacja obowiązkowej umowy w sprawie integracji między osobą poszukującą pracy a prowadzącym sprawę. Wykorzystując randomizowane badanie terenowe i monitorując losy osób objętych badaniem na rynku pracy, IAB był w stanie wykazać, że w przypadku niektórych grup bezrobotnych ta obowiązkowa regulacja przynosi efekty odwrotne do zamierzonych i powinna zostać zastąpiona bardziej elastycznym wykorzystywaniem instrumentu (van den Berg i in. 2016).

Jakie problemy mogą się pojawić w obszarze ochrony danych?

Mogą wystąpić trudności w uzyskaniu danych umożliwiających identyfikację osób lub przedsiębiorstw, które uczestniczyły w interwencjach finansowanych z EFS+⁴⁶. Ewaluacje wpływu oparte na metodach konfaktycznych wymagają danych jednostkowych – danych, które zawierają obserwacje dotyczące poszczególnych jednostek zarówno w grupie objętej interwencją, jak i grupie kontrolnej.

W rozporządzeniu (UE) 2021/1057 w sprawie EFS+ (załącznik I) zawarto wymóg dostarczenia danych dotyczących uczestników w podziale według płci, statusu na rynku pracy, grupy wiekowej, poziomu wykształcenia oraz grup w trudnej sytuacji (migranci, mniejszości, osoby z niepełnosprawnościami lub inne osoby w niekorzystnej sytuacji). W rozporządzeniu w sprawie wspólnych przepisów i rozporządzeniu w sprawie EFS+ na lata 2021–2027 na IZ nałożono prawny obowiązek gromadzenia i przetwarzania danych osobowych w postaci indywidualnych rejestrów uczestników. Ponadto w art. 17 ust. 6 rozporządzenia (UE) 2021/1057 w sprawie EFS+ w odniesieniu do monitorowania i wskaźników określono: „W przypadku gdy dane są dostępne w rejestrach lub równoważnych źródłach, państwa członkowskie mogą umożliwić instytucjom zarządzającym i innym podmiotom, którym powierzono zadanie zbierania danych niezbędnych do monitorowania i ewaluacji wsparcia ogólnego z komponentu EFS+ objętego zarządzaniem dzielonym, pozyskanie tych danych z tych rejestrów lub równoważnych źródeł, zgodnie z art. 6 ust. 1 lit. c) i e) rozporządzenia (UE) 2016/679”.

Rozporządzenie w sprawie EFS+ w kontekście zarządzania danymi

Zasady te, określone w rozporządzeniu w sprawie wspólnych przepisów i rozporządzeniu w sprawie EFS+, ułatwiają dostęp do danych osobowych potrzebnych do monitorowania EFS+ oraz – w przypadku ewaluacji wpływu opartej na metodach konfaktycznych – do określenia grupy poddanej interwencji, a także zapewniają możliwość korzystania z tych danych. Dostęp do danych osobowych niezbędnych do utworzenia grup kontrolnych oraz przetwarzanie danych dotyczących grupy objętej interwencją i grupy kontrolnej muszą jednak być zgodne z rozporządzeniem (UE) 2016/679 (ogólne rozporządzenie o ochronie danych – RODO), które obejmuje ogólne przekazywanie i wykorzystywanie danych osobowych, w tym szczególnych kategorii danych osobowych⁴⁷ w UE. W poniższej ramce przedstawiono przegląd najważniejszych treści i obowiązków regulowanych przez RODO.

Ramka 8. Unijne ramy regulacyjne dotyczące przetwarzania danych osobowych

Odpowiednie przepisy dotyczące przetwarzania danych osobowych w Europie obejmują przede wszystkim rozporządzenie (UE) 2016/679 oraz wytyczne i środki przyjęte przez właściwe organy takie jak Europejska Rada Ochrony Danych (EROD). Przepisy te określają wiele warunków i ograniczeń dotyczących przetwarzania danych osobowych w celu ochrony praw i wolności osób, których dane dotyczą. Niezależnie od sytuacji należy znaleźć równowagę między wynikającą z przepisów i regulacji koniecznością zapewnienia ochrony i zabezpieczenia a potrzebą nieograniczania badań naukowych, czy wręcz przyczyniania się do ich rozwoju. Z tego powodu i w tych ramach prawnych Unia Europejska przewidziała,

⁴⁶ Zob. [Ismeri Europa – Ecorys – Institute for Employment Studies, 2019](#).

⁴⁷ W RODO nie używa się już terminu „dane wrażliwe” i obecnie mówi się o „szczególnych kategoriach danych osobowych”. Zgodnie z art. 9 zabronione jest przetwarzanie takich danych, tj. „danych osobowych ujawniających pochodzenie rasowe lub etniczne, poglądy polityczne, przekonania religijne lub światopoglądowe, przynależność do związków zawodowych oraz przetwarzanie danych genetycznych, danych biometrycznych w celu jednoznacznej zidentyfikacji osoby fizycznej lub danych dotyczących zdrowia, seksualności lub orientacji seksualnej tej osoby”.

że pod pewnymi warunkami możliwe są wyjątki umożliwiające prowadzenie działalności badawczej i rozpowszechnianie jej wyników, o ile zagwarantowane jest podstawowe prawo zainteresowanych osób, tj. prawo do prywatności⁴⁸.

OGÓLNE ROZPORZĄDZENIE O OCHRONIE DANYCH (UE) 2016/679

Ogólne rozporządzenie o ochronie danych (RODO) weszło w życie w dniu 24 maja 2016 r., a we wszystkich państwach członkowskich zaczęło w pełni obowiązywać w dniu 25 maja 2018 r. RODO „ma zastosowanie do przetwarzania danych osobowych w sposób całkowicie lub częściowo zautomatyzowany oraz do przetwarzania w sposób inny niż zautomatyzowany danych osobowych stanowiących część zbioru danych lub mających stanowić część zbioru danych”⁴⁹.

Jeśli chodzi o terytorialny zakres stosowania, rozporządzenie ma zastosowanie zarówno do przetwarzania danych osobowych przez osoby znajdujące się na terytorium Unii Europejskiej, jak i w przypadkach, gdy przetwarzanie odnosi się do osób, których dane dotyczą, znajdujących się na terytorium Unii Europejskiej, nawet jeśli administrator danych lub podmiot przetwarzający dane znajdują się poza UE.

Z merytorycznego punktu widzenia, a także w porównaniu z poprzednim prawodawstwem, RODO wzmacnia prawa osób, których dane dotyczą, i nakłada na administratorów danych szereg obowiązków zgodnie z logiką opartą na analizie ryzyka i na zasadzie rozliczalności. Ponadto w rozporządzeniu przewidziano szereg wymogów, które musi spełnić administrator danych, a które nie były zawarte w poprzednim prawodawstwie: chodzi między innymi o ocenę wpływu na prywatność (art. 35–36), minimalizację operacji przetwarzania zgodnie z kryteriami uwzględnienia ochrony prywatności już w fazie projektowania i domyślnej ochrony prywatności (art. 25), przyjęcie rejestru czynności przetwarzania (art. 30) i wyznaczenie inspektora ochrony danych (art. 37–39).

Ważne jest podkreślenie głównych zasad określonych w art. 5 RODO, których muszą przestrzegać podmioty przetwarzające dane osobowe. W szczególności chodzi o następujące zasady:

- a) Zgodność z prawem, rzetelność i przejrzystość: dane osobowe muszą być przetwarzane zgodnie z prawem, rzetelnie i w sposób przejrzysty.
- b) Ograniczenie celu: dane osobowe muszą być zbierane w konkretnych, wyraźnych i prawnie uzasadnionych celach i nieprzetwarzane dalej w sposób niezgodny z tymi celami. W przypadku przetwarzania do celów statystycznych lub do celów badań naukowych administrator danych przyjmuje niezbędne gwarancje i środki ochrony.
- c) Minimalizacja danych: należy gromadzić wyłącznie dane bezwzględnie niezbędne do osiągnięcia konkretnych celów.
- d) Prawidłowość: gromadzone dane muszą być prawidłowe i w razie potrzeby uaktualniane.
- e) Ograniczenie przechowywania: dane osobowe muszą być przechowywane w formie umożliwiającej identyfikację osoby, której dane dotyczą, przez okres nie dłuższy, niż jest to niezbędne do celów, w których dane te są przetwarzane. W tym celu dane osobowe można przechowywać przez okres dłuższy, o ile będą one przetwarzane wyłącznie do celów archiwalnych w interesie publicznym, do celów badań naukowych lub historycznych lub do celów statystycznych na mocy art. 89, z zastrzeżeniem, że takich przypadkach wdrożone zostaną odpowiednie środki techniczne i organizacyjne w celu ochrony praw i wolności osób, których dane dotyczą.
- f) Integralność i poufność: dane osobowe muszą być przetwarzane w sposób zapewniający ich odpowiednie bezpieczeństwo, w tym ochronę przed niedozwolonym lub niezgodnym z prawem przetwarzaniem oraz przypadkową utratą, zniszczeniem lub uszkodzeniem, za pomocą odpowiednich środków technicznych lub organizacyjnych.
- g) Rozliczalność: jest jedną z najważniejszych zasad określonych w tym rozporządzeniu. Stanowi ona, że administrator jest odpowiedzialny za przetwarzanie danych i musi być w stanie wykazać wdrożenie wszelkich wymaganych środków.

Powyższy zbiór zasad stanowi główną strukturę, na której opiera się RODO, oraz określa szereg obowiązków, które muszą spełnić administrator danych i podmiot przetwarzający.

⁴⁸ Wśród innych istotnych przepisów europejskich, które współtworzą ramy prawne dotyczące przetwarzania danych osobowych, można wymienić: rozporządzenie Komisji (UE) nr 557/2013 w sprawie europejskiej statystyki w zakresie dostępu do poufnych danych do celów naukowych; rozporządzenie (UE) 2018/1725 w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych przez instytucje, organy i jednostki organizacyjne Unii i swobodnego przepływu takich danych.

⁴⁹ Art. 2 rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679.

PRZETWARZANIE DANYCH OSOBOWYCH W KONTEKŚCIE CIE: WARUNKI, OGRANICZENIA I NAJWAŻNIEJSZE KWESTIE

Przeprowadzenie ewaluacji wpływu opartej na metodach konfaktycznych wiąże się z przetwarzaniem dużych ilości danych, w tym danych osobowych. W nawiązaniu do opisanych powyżej ram regulacyjnych ważne jest, aby zapoznać się z warunkami, ograniczeniami i najważniejszymi kwestiami, z jakimi spotykają się państwa członkowskie i IZ podczas przeprowadzania ewaluacji wpływu opartych na metodach konfaktycznych.

Podstawa prawna i cele przetwarzania

Do przetwarzania danych osobowych musi być spełniony co najmniej jeden z warunków zgodności przetwarzania z prawem wskazanych w art. 6 RODO⁵⁰. Jeśli natomiast dane, które mają być przetwarzane, są „szczególnymi kategoriami danych osobowych”, należy również odnieść się do art. 9 RODO.

W kontekście ewaluacji wpływu opartej na metodach konfaktycznych najbardziej odpowiednią podstawą prawną przetwarzania danych wydaje się **interes publiczny** zgodnie z art. 6 ust. 1 lit. e) RODO („wykonanie zadania realizowanego w interesie publicznym lub w ramach sprawowania władzy publicznej powierzonej administratorowi”). Taka podstawa prawna musi być ustanowiona w prawie unijnym lub krajowym, jak określono w art. 6 ust. 3 RODO. „Interes publiczny” wyraźnie odzwierciedla obowiązki instytucji zarządzających określone w art. 17 ust. 6 rozporządzenia (UE) 2021/1057 w sprawie EFS+ dotyczącym wykorzystywania danych, które są „dostępne w rejestrach lub równoważnych źródłach”. Ponadto art. 4 rozporządzenia (UE) 2021/1060 w sprawie wspólnych przepisów stanowi, że „[p]aństwa członkowskie i Komisja mogą przetwarzać dane osobowe tylko wtedy, gdy jest to konieczne do celów wykonywania ich odpowiednich obowiązków wynikających z niniejszego rozporządzenia, w szczególności do celów monitorowania, sprawozdawczości, komunikacji, publikacji, ewaluacji [...]”. Przepisy krajowe mogą również nadawać administratorom danych podobne uprawnienia do zarządzania danymi i przetwarzania ich w interesie publicznym.

Inne podstawy prawne w odniesieniu do ewaluacji wpływu opartej na metodach konfaktycznych mogą być przywołane w innych warunkach wymienionych w art. 6 ust. 1, a w szczególności są to:

- **zgoda** osoby, której dane dotyczą (art. 6 ust. 1 lit. a) RODO). Zgoda może być odpowiednią podstawą prawną, na przykład gdy dane na potrzeby ewaluacji wpływu opartej na metodach konfaktycznych są gromadzone za pomocą badania, a osoby, których dane dotyczą, mogą łatwo wyrazić zgodę na przetwarzanie (zob. art. 7 RODO dotyczący zgody). Ogólnie rzecz biorąc, kwestia zgody staje bardziej złożona, jeśli nie zostanie zaplanowana z dużym wyprzedzeniem; można ją uznać za „szczątkową” podstawę prawną ewaluacji wpływu opartej na metodach konfaktycznych, gdy inne podstawy prawne nie mają zastosowania;
- przetwarzanie jest niezbędne do **wypełnienia obowiązku prawnego** ciążącego na administratorze (art. 6 ust. 1 lit. c) RODO w odniesieniu do art. 17 rozporządzenia w sprawie EFS+). Ta podstawa prawna musi być ustanowiona przez prawo i może dotyczyć podmiotów prywatnych lub publicznych. Możliwa jest na przykład sytuacja, w której ze względu na szczególne przepisy prawne podmiot prywatny lub publiczny odpowiedzialny za zbiór danych jest zobowiązany do współpracy z instytucją zarządzającą przy ewaluacji wpływu opartej na metodach konfaktycznych.

Ponadto administrator danych **może wykorzystać zgromadzone dane do innych celów, jeśli są one zgodne z pierwotnymi celami**. W tym zakresie administrator danych będzie musiał ocenić warunki określone w art. 6 ust. 4 i art. 5 ust. 1 lit. b) RODO. Przepisy te są szczególnie istotne w przypadku przetwarzania danych do celów badań naukowych lub do celów statystycznych. Pozwalają one na wykorzystanie danych administracyjnych do celów innych niż pierwotne i nie wymagają uzyskania szczególnej zgody na nowe wykorzystanie, ale muszą być zgodne z zasadami ochrony określonymi w

⁵⁰ Warunki te są następujące: a) osoba, której dane dotyczą wyraziła zgodę na przetwarzanie swoich danych osobowych w jednym lub większej liczbie określonych celów; b) przetwarzanie jest niezbędne do wykonania umowy, której stroną jest osoba, której dane dotyczą, lub do podjęcia działań na żądanie osoby, której dane dotyczą, przed zawarciem takiej umowy; c) przetwarzanie jest niezbędne do wypełnienia obowiązku prawnego ciążącego na administratorze; d) przetwarzanie jest niezbędne do ochrony żywotnych interesów osoby, której dane dotyczą, lub innej osoby fizycznej; e) przetwarzanie jest niezbędne do wykonania zadania realizowanego w interesie publicznym lub w ramach sprawowania władzy publicznej powierzonej administratorowi; f) przetwarzanie jest niezbędne do celów wynikających z prawnie uzasadnionych interesów realizowanych przez administratora lub przez stronę trzecią, z wyjątkiem sytuacji, w których nadrzędny charakter wobec tych interesów mają interesy lub podstawowe prawa i wolności osoby, której dane dotyczą, wymagające ochrony danych osobowych, w szczególności gdy osoba, której dane dotyczą, jest dzieckiem.

art. 89 ust. 1, głównie pseudonimizacji (zob. poniżej). Przepisy te mogą być na przykład istotne w przypadku danych z rejestru bezrobotnych, które mają być wykorzystane do ewaluacji wpływu opartej na metodach konfaktycznych.

Ze względu na specyfikę ewaluacji wpływu opartych na metodach konfaktycznych administrator danych gromadzący dane nie musi być podmiotem prowadzącym badanie. W takim przypadku musi zaistnieć warunek, który pozwoli zapewnić, aby przekazanie danych było zgodne z prawem i aby odbiorca mógł zrealizować ewaluację wpływu opartą na metodach konfaktycznych.

Anonimizacja i pseudonimizacja: przetwarzanie do celów statystycznych

Art. 89 RODO stanowi, że w przypadku przetwarzania danych w interesie publicznym, w kontekście badań naukowych lub do celów statystycznych należy zapewnić odpowiednie zabezpieczenia dla praw i wolności osób, których dane dotyczą, oraz poszanowanie w szczególności zasady „minimalizacji”. Oznacza to korzystanie z technik pseudonimizacji⁵¹.

Jeżeli cele te można zrealizować w drodze dalszych operacji przetwarzania danych, które nie pozwalają albo przestały pozwalać na zidentyfikowanie osoby, której dane dotyczą, cele należy realizować za pomocą anonimizacji danych, a następnie przetwarzania ich w formie zagregowanej. Wyniki ewaluacji wpływu opartej na metodach konfaktycznych są z założenia agregowane i ryzyko to nie występuje, chyba że oryginalne zestawy danych są publikowane na potrzeby badań naukowych. W takim przypadku zbiory danych muszą zostać zanonimizowane (zob. przykład w ramce poniżej).

Przechowywanie i bezpieczne przetwarzanie danych

Jednym z głównych aspektów przetwarzania danych jest ich przechowywanie. Przepisy nie określają sposobu przechowywania danych, ale zgodnie ze wspomnianymi wyżej zasadami wymagane jest, by przechowywanie i przetwarzanie danych było zawsze związane z celem badań. Gdy cel przetwarzania danych zostanie osiągnięty, a ich zatrzymywanie nie będzie już konieczne, należy je zakończyć. Ta ogólna zasada musi być określona w oświadczeniu o ochronie prywatności, które jest przekazywane osobom, których dane dotyczą, gdy są one rejestrowane w administracyjnych zbiorach danych lub, biorąc pod uwagę prawdopodobny brak możliwości osobistego poinformowania wszystkich osób, których dane dotyczą, można znaleźć alternatywne sposoby przekazania tych informacji (np. przez opublikowanie na stronie internetowej instytucji zarządzającej strony informacyjnej zawierającej politykę prywatności dotyczącą działalności badawczej).

Cel uważa się za osiągnięty w szczególności po przeprowadzeniu ewaluacji wpływu opartej na metodach konfaktycznych, udostępnieniu i zagregowaniu wyników do celów statystycznych. Na tym etapie dane powinny zostać zanonimizowane (na wypadek, gdyby miały zostać wykorzystane ponownie na późniejszym etapie i do innych celów) albo usunięte. Jeśli pozwalają na to przepisy ustawowe lub wykonawcze, przechowywanie może być nieograniczone w czasie, ale tylko na podstawie wyraźnego odniesienia w przepisach.

Kolejnym warunkiem wymaganym przez RODO jest to, że przetwarzanie danych musi być otoczone odpowiednimi i właściwymi środkami bezpieczeństwa. Art. 32 RODO stanowi, że każdy, kto dokonuje przetwarzania (administrator lub podmiot przetwarzający), wdraża odpowiednie środki techniczne i organizacyjne, aby zapewnić stopień bezpieczeństwa odpowiadający temu ryzyku. Oznacza to, że nie przewidziano żadnych standardowych środków, ale muszą być one określone osobno dla każdego przypadku, ze szczególnym odniesieniem do ryzyka stwarzanego przez poszczególne operacje przetwarzania.

⁵¹ „Pseudonimizacja” oznacza przetworzenie danych osobowych w taki sposób, by nie można ich było już przypisać konkretnej osobie bez użycia dodatkowych informacji. Anonimizacja odnosi się do przetwarzania danych osobowych w sposób, który uniemożliwia identyfikację osób na ich podstawie. Przegląd technik pseudonimizacji można znaleźć na stronie internetowej <https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices>.

ZOBOWIĄZANIA PRAWNE

Informacje dla osób, których dane dotyczą

Pierwszym obowiązkiem administratora danych jest informowanie osób, których dane dotyczą. W przypadku zbierania danych od osoby, której dane dotyczą, obowiązek ten określono w art. 13 RODO, a w innych przypadkach odniesieniem jest art. 14. Należy poinformować osobę, której dane dotyczą, o celach i metodach przetwarzania, podstawie prawnej, okresach przechowywania oraz prawach, z których osoba, której dane dotyczą, może skorzystać. W przypadku ewaluacji wpływu opartej na metodach konfaktycznych informacje te są podawane w momencie gromadzenia danych, które będą później wykorzystywane do badań statystycznych i ewaluacyjnych, takich jak monitorowanie EFS lub zbiory danych z rejestrów bezrobotnych. Gdy wskazana powyżej komunikacja jest niemożliwa z praktycznego punktu widzenia, a w szczególności w przypadku osób figurujących w rejestrach wykorzystywanych na potrzeby ewaluacji wpływu opartej na metodach konfaktycznych, komunikację można prowadzić za pomocą strony internetowej lub innych ogólnych narzędzi informacyjnych.

Zarządzanie relacjami między poszczególnymi podmiotami zaangażowanymi w przetwarzanie danych

Zasadniczo w działalność badawczą zaangażowana jest większa liczba podmiotów i w takich przypadkach stosunki między nimi muszą być regulowane szczególnymi ustaleniami znanymi jako „umowy o przetwarzaniu danych”⁵². Należy je określić indywidualnie dla każdego przypadku, odzwierciedlając wkład różnych zainteresowanych stron. Przykłady możliwych relacji to:

- relacja *administrator danych – administrator danych* występuje, gdy podmioty współpracują przy realizacji projektu, choć w różnych warunkach i przy różnych zadaniach, przy czym każdy z nich zachowuje swój odrębny cel przetwarzania;
- relacja *administrator danych – podmiot przetwarzający* występuje wtedy, gdy jeden podmiot (administrator danych) określa środki i cele przetwarzania danych i wykorzystuje inny podmiot (podmiot przetwarzający) do wykonywania określonych czynności przetwarzania. W tym przypadku odniesienia i obowiązki można znaleźć w art. 28 RODO;
- relacja między *współadministratorami* jest określona w art. 26 RODO, który stanowi, że: „Jeżeli co najmniej dwóch administratorów wspólnie ustala cele i sposoby przetwarzania, są oni współadministratorami”. W takich przypadkach strony muszą w przejrzysty sposób określić odpowiednie zakresy swojej odpowiedzialności dotyczącej przestrzegania przepisów, w szczególności w odniesieniu do praw osób, których dane dotyczą.

Ocena skutków dla ochrony danych

Art. 35 RODO stanowi, że w przypadku gdy przetwarzanie danych osobowych może stanowić potencjalnie wysokie ryzyko naruszenia praw i wolności osób, których dane dotyczą, przed rozpoczęciem przetwarzania administrator danych musi dokonać oceny skutków⁵³. Ocena skutków w ramach ewaluacji wpływu opartych na metodach konfaktycznych może być istotna jedynie w przypadku przetwarzania szczególnych kategorii osób, których dane dotyczą, na dużą skalę. W takich przypadkach administrator danych w administracji może przeprowadzić ocenę skutków zgodnie z zasadami i narzędziami udostępnionymi przez krajowy organ ochrony danych.

PLANY NA PRZYSZŁOŚĆ: AKT O ZARZĄDZANIU DANYMI

Przedstawione powyżej ramy odzwierciedlają aktualny stan wiedzy i przepisy obowiązujące w UE. RODO wprowadzono specjalnie po to, by umożliwić przełom w ochronie danych osobowych i stworzyć wspólny system w całej UE zgodnie z aktualnym rozwojem technologicznym i z myślą o współczesnym społeczeństwie opartym na danych. W przygotowaniu są kolejne reformy, które mogą mieć znaczący wpływ na gromadzenie danych osobowych w nadchodzących latach.

Akt w sprawie zarządzania danymi obecnie jest wciąż na etapie wniosku w Komisji Europejskiej⁵⁴.

⁵² Jest to wymagane np. w przypadku, gdy administrator danych i osoby faktycznie prowadzące badania są różnymi podmiotami. W takim przypadku należy zawrzeć umowę, która ureguluje relacje między stronami w zakresie dostępu do danych i prowadzenia badań.

⁵³ Więcej informacji można znaleźć w wytycznych opublikowanych przez EROD <https://ec.europa.eu/newsroom/article29/items/611236>.

⁵⁴ Wniosek dotyczący aktu w sprawie zarządzania danymi został przyjęty przez Komisję pod koniec 2020 r. Informacje na temat stanu prac nad aktem w sprawie zarządzania danymi można znaleźć tutaj: <https://eur-lex.europa.eu/legal-content/PL/HIS/?uri=CELEX%3A52020PC0767>.

Wyraźnym celem tego aktu jest promowanie dostępności danych użytkowych poprzez zwiększenie zaufania do pośredników w zakresie danych i wzmocnienie mechanizmów udostępniania danych w całej UE. Przedmiotem tej nowej normy będą prawdopodobnie dane osobowe, których wykorzystanie do celów statystycznych i badawczych może być jeszcze większe dzięki wprowadzeniu nowego podmiotu: pośrednika w udostępnianiu danych osobowych. Taki środek mógłby przyczynić się do stworzenia bardziej sprzyjającego środowiska dla ewaluacji wpływu opartych na metodach konfaktycznych, umożliwiając łatwiejszy dostęp do dużej ilości informacji na dużą skalę.

Oczywiste jest, że przepisy RODO nie wykluczają wdrożenia ewaluacji wpływu opartej na metodach konfaktycznych. W związku z tym, aby zapewnić bezpieczeństwo ewaluacji wpływu opartej na metodach konfaktycznych, we wszystkich państwach członkowskich należy podjąć pewne podstawowe działania proceduralne i operacyjne. Na przykład w normalnym przypadku, gdy dane z monitorowania (dotyczące osób poddanych interwencji) są połączone z danymi z rejestru publicznego (dotyczących grupy kontrolnej), a zewnętrzny ewaluator przeprowadza ewaluację wpływu opartą na metodach konfaktycznych, należy podjąć następujące najważniejsze działania:

Najważniejsze działania służące zapewnieniu lepszej zgodności z RODO

1. IZ zawiera porozumienie z organami administracyjnymi odpowiedzialnymi za dane (np. rejestr bezrobotnych, rejestr podatkowy itp.) niezbędne do identyfikacji grupy kontrolnej oraz analizy grupy objętej interwencją i grupy kontrolnej. Jeśli nie uzyskano wyraźnej zgody, IZ weryfikuje, czy wykorzystanie tych danych jest zgodne z art. 6 ust. 1 lit. e), lub art. 6 ust. 4 i art. 5 ust. 1 lit. b) RODO.
2. IZ zawiera umowy z innymi podmiotami (właścicielami danych i ewaluatorem) w celu uregulowania przepływu informacji i wzajemnych obowiązków zgodnie z przepisami RODO. Umowa o świadczenie usług między ewaluatorem a IZ musi zawierać specjalną klauzulę dotyczącą ochrony danych; w przypadku innych organów administracji ochrona danych przy wdrażaniu ewaluacji wpływu opartych na metodach konfaktycznych może być regulowana protokołem ustaleń lub szczegółowymi procedurami krajowymi.
3. Na podstawie wskazanej powyżej umowy IZ otrzymuje dane w formie pseudonimizowanej od właściciela danych i przekazuje ewaluatorowi dane na potrzeby przetwarzania informacji o grupie poddanej interwencji i grupie kontrolnej zgodnie z RODO⁵⁵.
4. Przechowywanie danych osobowych grupy objętej interwencją i grupy kontrolnej odbywa się wyłącznie zgodnie z przepisami dotyczącymi przechowywania danych na czas trwania i do celów badania oraz zgodnie z podstawowymi przepisami dotyczącymi bezpieczeństwa określonymi w RODO. Przepisy te muszą być przestrzegane przez wszystkie podmioty zaangażowane w ewaluację wpływu opartą na metodach konfaktycznych.

Praktyki krajowe są zasadniczo zgodne z RODO, ale różnią się w poszczególnych państwach członkowskich, a zgodnie z informacjami przekazywanymi przez ewaluatorów krajowe przepisy o ochronie danych nadal stanowią poważne przeszkody w wykorzystaniu danych jednostkowych. Może to wynikać z czasu potrzebnego na dostosowanie

⁵⁵ Jeśli grupa objęta interwencją i grupa kontrolna są analizowane za pomocą badania, w takim badaniu należy zawrzeć zgodę na wykorzystanie danych osób badanych do celów badawczych.

krajowych przepisów i zwyczajów do nowszego od nich RODO lub z różnych interpretacji RODO w poszczególnych administracjach krajowych, nadmierne rygorystycznego wdrażania, zaostrzania RODO w niektórych krajach lub innych błędnych interpretacji. W niektórych krajach do pokonania tych przeszkód w duchu RODO i rozporządzenia w sprawie EFS+ konieczna byłaby konkretna inicjatywa ze strony IZ lub organów krajowych. W poniższej ramce opisano, w jaki sposób włoski region Wenecja Euganejska poradził sobie z niektórymi z najczęściej występujących trudności. Dobrą praktyką jest skontaktowanie się z krajowym organem ochrony danych w celu omówienia proponowanych ustaleń przed ich sfinalizowaniem.

Ramka 9. Ochrona i wymiana danych

Dobry przykład dostępu do zanonimizowanych danych osobowych w stosunkowo krótkim czasie można znaleźć w ewaluacji kontryfakcyjnej opisanej w dokumencie „Projekt pilotażowy i studium wykonalności dotyczące trwałości i skuteczności rezultatów dla uczestników Europejskiego Funduszu Społecznego z wykorzystaniem ewaluacji wpływu opartych na metodach kontryfakcyjnych” („Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations”)⁵⁶ przeprowadzonego dla DG EMPL w 2019 r.

Ewaluacja – jeden z czterech przypadków opisanych w badaniu – koncentrowała się na interwencjach w ramach EFS (PI 9.i) przeprowadzonych w regionie Wenecji Euganejskiej w latach 2015–2016 skierowanych do osób długotrwale bezrobotnych. Dane z monitorowania dotyczące uczestników i rodzajów środków zostały połączone z danymi dotyczącymi okresów zatrudnienia przed przeprowadzeniem interwencji w ramach EFS i po ich przeprowadzeniu (Comunicazioni obbligatorie). Proces ten trwał około dwóch miesięcy (od końca lipca do końca września) od pierwszego spotkania w sprawie wymogów dotyczących danych z udziałem instytucji zarządzającej i instytucji przechowującej dane, regionalnym instytutem Veneto Lavoro. Instytut Veneto Lavoro był odpowiedzialny za anonimizację danych, a wszystkie dostarczone zbiory danych można było łatwo powiązać ze sobą⁵⁷ za pomocą unikalnego wspólnego identyfikatora zapewnionego przez Veneto Lavoro.

Chociaż zaangażowane organizacje nie zlecały bezpośrednio ewaluacji, pozytywne doświadczenia wynikają z dwóch głównych czynników: ścisłej (choć nieformalnej) współpracy między regionalnymi biurami instytucji zarządzającej a regionalnym instytutem Veneto Lavoro, która była niezbędna do tego, by rozwiązać wszelkie problemy w zakresie ochrony danych osobowych, oraz istnienia bazy danych (Mercurio) zarządzanej przez Veneto Lavoro i dostępnej jako plik do użytku publicznego do celów badawczych⁵⁸.

Mercurio jest bazą danych statystycznych zawierającą wszystkie dane zarejestrowane przez urzędy pracy w Wenecji Euganejskiej dotyczące okresów zatrudnienia i bezrobocia wśród tamtejszej populacji. Ponadto obejmuje ona szczegółowe informacje na temat wszystkich zarejestrowanych pracowników i przedsiębiorstw. W momencie przeprowadzania analizy baza Mercurio zawierała informacje na temat ponad 4 mln pracowników, 17,6 mln zdarzeń związanych z zatrudnieniem i 4 mln zdarzeń związanych z bezrobociem⁵⁹.

Zbiór danych jest okresowo „aktualizowany i czyszczony”, dlatego dopasowanie go do danych z monitorowania instytucji zarządzającej jest stosunkowo proste. Jedno z potencjalnych ograniczeń wynika z faktu, że plik do użytku publicznego nie jest aktualizowany wystarczająco często (na przykład ostatnia dostępna obecnie wersja została zaktualizowana w grudniu 2020 r.), ponieważ zapewnienie dobrej jakości danych wymaga ogromnego wysiłku⁶⁰. Z doświadczeń wynika, że przy opracowywaniu planów ewaluacji IZ

⁵⁶ Ismeri Europa – Ecorys – Institute for Employment Studies, 2019.

⁵⁷ Zapewniono dodatkowy zbiór danych zawierający daty rozpoczęcia i zakończenia każdego oświadczenia o gotowości do niezwłocznego podjęcia pracy podpisanego przez wszystkie osoby poddane i niepoddane interwencji.

⁵⁸ Opis bazy Mercurio i zasady dostępu do niej można znaleźć na stronie internetowej <https://www.venetolavoro.it/public-use-file>

⁵⁹ Z danych administracyjnych wyodrębniono następujące zmienne będące przedmiotem zainteresowania: daty rozpoczęcia i zakończenia każdego okresu zatrudnienia i bezrobocia; rodzaj umowy w przypadku każdego okresu zatrudnienia (umowa na czas nieokreślony, umowa o pracę na czas określony, umowa o staż); przewidywany czas trwania umów o pracę na czas określony; informacje społeczno-demograficzne (płeć, data urodzenia, wykształcenie, obywatelstwo).

⁶⁰ Oznacza to, że konieczne jest opieranie się na bardziej aktualnych informacjach niż te zawarte w najnowszej dostępnej wersji

powinna z wyprzedzeniem dokonać ustaleń dotyczących: możliwych form współpracy z podmiotami zewnętrznymi zarządzającymi danymi niezbędnymi do opracowania scenariusza kontryfaktycznego; usunięcia ewentualnych przeszkód prawnych; wyjaśnienia sposobów dostępu do danych administracyjnych, zarządzania nimi i dostarczania ich w formie zanonimizowanej.

2.3.2. W jaki sposób należy określić grupę „poddaną interwencji”?

W celu przeprowadzenia ewaluacji wpływu opartej na metodach kontryfaktycznych należy jasno określić, co oznacza poddanie interwencji lub uczestnictwo w interwencji. Ponadto po dokładnym ustaleniu, kiedy daną osobę lub dane przedsiębiorstwo uważa się za poddane interwencji, ważne jest, aby możliwa była ich identyfikacja. W tej sekcji przedstawiono główne kwestie związane z definicją grupy objętej interwencją i grupy kontrolnej, przy czym bardziej szczegółowe wyjaśnienia dotyczące ich charakterystyki metodologicznej można znaleźć w rozdziale 3.

Zdefiniowanie uczestnictwa może na pozór wydawać się proste. Istnieje jednak szereg kwestii, które nie są na pierwszy rzut oka oczywiste, ale mają kluczowe znaczenie i wymagają gruntownego przemyślenia. Na przykład: czy uczestników szkolenia, którzy rezygnują z udziału w interwencji, uznaje się za poddanych interwencji? W ilu sesjach kursu szkoleniowego muszą wziąć udział uczestnicy szkolenia, aby można było ich uznać za uczestników? Należy również wziąć pod uwagę efekt wyprzedzający. W oczekiwaniu na przeprowadzenie interwencji niektóre osoby ubiegające się o świadczenia z zabezpieczenia społecznego mogą zrezygnować z korzystania z pomocy socjalnej, aby uniknąć działań aktywizujących. Czy takie osoby uważa się za poddane interwencji, mimo że np. nigdy nie stawiły się na umówione spotkania w urzędzie publicznych służb zatrudnienia?

Definicja grupy objętej interwencją

Przy definiowaniu „grupy objętej interwencją” istnieje rozróżnienie między „zamiarem poddania oddziaływaniu” a „poddaniem oddziaływaniu osób poddanych interwencji”. Z punktu widzenia polityki zazwyczaj kluczowe pytanie brzmi: czy przedmiotem zainteresowania są efekty otrzymania możliwości udziału w interwencji, czy raczej efekty faktycznego udziału? W pierwszym przypadku osoby, którym zaproponowano udział w interwencji, mogą, ale nie muszą w niej uczestniczyć. W drugim przypadku, gdy przedmiotem zainteresowania jest wpływ programu na osoby należące do grupy poddanej interwencji, grupa poddana interwencji obejmuje tylko uczestników⁶¹.

Zamiar poddania oddziaływaniu lub poddanie oddziaływaniu grupy poddanej interwencji

Początkowo decydenci często zakładają, że przedmiotem ich zainteresowania jest określenie efektów netto interwencji w odniesieniu do osób, które w niej uczestniczą. Po dłuższym zastanowieniu kwestie te mogą być jednak mniej oczywiste. Jeżeli możliwe jest zidentyfikowanie osób, którym zaproponowano udział w programie, z punktu widzenia polityki bardziej pomocne może być określenie ich jako grupy „poddanej interwencji”. Dotyczy to w szczególności okoliczności, w których udział w interwencji nie jest obowiązkowy. Decydenci nie mogą zmusić osób, które otrzymały możliwość udziału w interwencji, do uczestnictwa w niej, zatem właściwe pytanie brzmi:

Oferta lub faktyczna interwencja

Mercurio, a ponadto należy podjąć konkretne dalsze działania w celu włączenia i wyczyszczenia najbardziej aktualnych informacji na temat okresów zatrudnienia.

⁶¹ Jeżeli udział w interwencji jest obowiązkowy, nie ma zasadniczo różnicy między tymi dwoma statusami – każda osoba, której zaproponowano wzięcie udziału w interwencji, musi w nim uczestniczyć. W większości przypadków jednak interwencje nie są obowiązkowe (tak też przyjęto w niniejszym przewodniku).

jaki jest wpływ na późniejsze znalezienie zatrudnienia i zarobki osób, którym zaproponowano udział w programie szkoleniowym?

Aby oszacować wpływ możliwości uczestnictwa w interwencji na szereg rezultatów, osoby, którym zaoferowano taką możliwość, muszą być możliwe do zidentyfikowania. W wielu przypadkach może być to trudne.

Gdzie można znaleźć odpowiednie dane?

Po ustaleniu, kogo uważa się za osobę poddaną interwencji i co stanowi interwencję, należy rozważyć, w jaki sposób osoby poddane interwencji będą identyfikowane na potrzeby ewaluacji. Każdorazowo oznacza to znalezienie źródła danych, na podstawie którego można wykazać lub objąć próbą wszystkie jednostki poddane interwencji – niezależnie od tego, czy są to osoby fizyczne czy przedsiębiorstwa. Dane te pochodzą zazwyczaj z systemów monitorowania EFS oraz, jeśli są dostępne, z pozostałych źródeł danych ustanowionych na potrzeby danej interwencji.

Wyszukiwanie źródeł danych dotyczących osób poddanych interwencji

Ze względu na wymogi dotyczące monitorowania i sprawozdawczości w odniesieniu do EFS (i EFS+) organizacje będące beneficjentami muszą rejestrować liczbę i niektóre cechy osobowe osób będących beneficjentami usług świadczonych w ramach interwencji. Na potrzeby ewaluacji wpływu opartej na metodach konfraktycznych interwencje należy usprawnić, tak aby dostarczały danych jednostkowych na temat osób, które w nich uczestniczyły. Ewaluatorzy będą wymagać danych dotyczących każdej jednostki poddanej interwencji (przedsiębiorstwa lub osoby fizycznej) na temat jej głównych cech (płeć, wiek, poziom wykształcenia itp.). Dane te można poddać anonimizacji/pseudonimizacji w celu ochrony prywatności, jednak w przypadku gdy do przeprowadzenia ewaluacji wpływu opartej na metodach konfraktycznych potrzebne jest badanie, konieczne będzie ustalenie, czy warunki wymagane w RODO pozwalają na przekazanie tożsamości tych jednostek (nazwisk, adresów, numerów telefonów itp.), tak aby można było objąć je próbą. Aby ułatwić łączenie wpisów z różnych źródeł danych, wymagane są również niepowtarzalne identyfikatory poszczególnych jednostek.

2.3.3. Czynniki, które należy uwzględnić przy identyfikacji grupy kontrolnej

Aby oszacować scenariusz konfraktyczny, należy zwykle określić grupę kontrolną. Wybór grupy kontrolnej jest zazwyczaj uwarunkowany tym, czy interwencja jest obowiązkowa dla uczestników, a także tym, czy interwencję przeprowadza się powszechnie w danej jurysdykcji, czy też jest ona ograniczona do określonego obszaru lub czasu. Wybór odpowiedniej grupy kontrolnej ma trzy aspekty: 1) analityczny; 2) związany z polityką; oraz 3) praktyczny.

Trzy aspekty, które należy wziąć pod uwagę przy określaniu grupy kontrolnej

Określenie grupy kontrolnej z perspektywy analitycznej

Celem ewaluacji wpływu opartej na metodach konfraktycznych jest uzyskanie wartości szacunkowych nieobarczonych błędem systematycznym dotyczących wpływu interwencji na szereg rezultatów. Aby to osiągnąć, należy oszacować rezultaty konfraktyczne. Rezultaty konfraktyczne szacuje się na podstawie grupy kontrolnej (zob. sekcja 1.1). Jak przedstawiono na wykresach 6 i 7, wpływ oblicza się poprzez odjęcie wartości szacunkowej

rezultatu kontrfaktycznego od rezultatu odnotowanego w grupie „poddanej interwencji”. Zakres, w jakim wpływ jest obciążony błędem systematycznym, zależy od stopnia, w jakim rezultat kontrfaktyczny obliczony w odniesieniu do grupy kontrolnej oznacza wynik, który uzyskano by w odniesieniu do grupy objętej interwencją, gdyby nie została ona poddana interwencji, przy założeniu, że wszystkie inne zmienne pozostają takie same.

Aby znaleźć odpowiednik grupy kontrolnej w przypadku braku randomizacji, musi ona być równoważna grupie objętej interwencją we wszystkich ważnych aspektach, zarówno w tych możliwych, jak i tych niemożliwych do zaobserwowania. Dotyczy to wszystkich podejść quasi-eksperymentalnych, w związku z czym jest warunkiem koniecznym dla wszystkich ewaluacji wpływu opartych na metodach kontrfaktycznych, w których nie stosuje się podejścia opartego na randomizacji.

Ponieważ prawie wszystkie interwencje w ramach EFS są: a) dobrowolne (grupa docelowa nie ma obowiązku udziału w interwencji) lub b) ograniczone w inny sposób – jeżeli są to interwencje pilotażowe lub instrumenty ograniczone do konkretnego obszaru lub jurysdykcji, ewaluatorzy będą mieli do czynienia z pulą jednostek, którą można by wykorzystać jako grupę kontrolną. Konieczne będzie pewne uporządkowanie tej potencjalnej puli jednostek w celu dopracowania ostatecznego wyboru osób do grupy kontrolnej, tak aby były one dobrze dopasowane względem uczestników (grupy poddanej interwencji). W wielu przypadkach dostępne są cztery opcje:⁶²

*Warianty
dotyczące wyboru
grup kontrolnych*

- **Lokalizacja** – grupy kontrolne składająca się z osób podobnych do tych uczestniczących w interwencji, ale zlokalizowanych w obszarach państwa członkowskiego, w których interwencja jest niedostępna (jeżeli takie obszary istnieją). W przypadku, gdy takie grupy kontrolne i odpowiednie dane są dostępne, preferowanym podejściem jest często model różnic w różnicach. Populacje w różnych lokalizacjach mogą być bardzo podobne do siebie i takie grupy nie otrzymają możliwości udziału w interwencji lub odmówią przyjęcia zaproszenia, co pozwoli uniknąć tego ważnego źródła potencjalnego błędu systematycznego. Populacje w różnych lokalizacjach będą jednak podlegać innym warunkom panującym na rynku pracy. Metoda różnic w różnicach w odniesieniu do takich zmian sprawdza się dość dobrze, ponieważ zazwyczaj różnice w warunkach panujących na lokalnych rynkach pracy są w miarę niezmiennie w czasie. Rządziej zaleca się jednak dobór prób kontrolnych z różnych lokalnych rynków pracy w przypadkach, gdy do oszacowania wpływu stosuje się parowanie. Wykazano, że błąd systematyczny związany z wyborem prób kontrolnych z różnych rynków pracy może być większy niż błąd selekcji⁶³.
- **Czas** – grupy kontrolne składające się z osób podobnych do uczestników, ale poddanych obserwacji w różnym czasie, przed interwencją albo po niej. Taki dobór grupy kontrolnej konieczny jest często w przypadku, gdy interwencja jest powszechna i obowiązkowa – innymi słowy, gdy wszyscy członkowie grupy docelowej są zobowiązani do wzięcia w niej udziału, a program jest realizowany na terenie całej jurysdykcji. Grupy kontrolne utworzone w ten sposób mają istotną wadę, a mianowicie ich rezultaty będą mierzone w różnym czasie niż rezultaty grupy objętej interwencją, a

⁶² Niniejsza sekcja opiera się na [Card, D., Ibarraran, P. i Villa, J.M., 2011.](#)

⁶³ Zob. [Heckman, J.J., Ichimura, H., Smith, J. i Todd, P., 1998.](#)

zatem będą podatne na wahania cykliczne, zmiany składu i zmieniające się tendencje makroekonomiczne, które mogą osłabić zdolność do określenia rezultatu kontrfaktycznego nieobarczonego błędem systematycznym. Takie grupy kontrolne należy wziąć pod uwagę jedynie w przypadku ograniczonej zmienności wyników w czasie oraz w przypadku, gdy aktualna grupa kontrolna jest niedostępna.

- **Kwalifikowalność** – w tym przypadku grupy kontrolne wybiera się z grup znajdujących się w tym samym miejscu i czasie, ale spośród kandydatów, którzy z tego czy innego powodu nie kwalifikowali się do udziału. Takich grup kontrolnych często poszukuje się, gdy interwencja jest powszechna, wskaźniki uczestnictwa są wysokie lub udział jest obowiązkowy oraz gdy istnieją jasne zasady kwalifikowalności, tak że np. osoby „niekwalifikujące się” stanowią potencjalne źródło grup kontrolnych. Celem jest znalezienie takich grup, które są podobne do grup poddanych interwencji, ale które z dobrze znanych i ustalonych powodów (które można określić ilościowo w danych) nie kwalifikowały się do poddania interwencji. Dostęp do interwencji w ramach EFS+ jest często oparty na wyraźnych zasadach kwalifikowalności, które można łatwo zmierzyć i którymi nie można manipulować (określony wiek uczestników, czas trwania statusu bezrobotnego itp.); w związku z tym przy wyborze grup kontrolnych można wykorzystać te progi jako czynniki rozróżniające, na podstawie których dokonuje się podziału na grupę poddana interwencji i grupę kontrolną.
- **Wybór/świadomość** – zasadniczo zarówno grupa objęta interwencją, jak i grupa kontrolna (a nie tylko grupa objęta interwencją) podlegają procesom wyboru, w których wybór podyktowany jest czynnikami potencjalnie niemożliwymi do zaobserwowania⁶⁴. Członków grupy kontrolnej można wybrać spośród osób, które kwalifikowały się do uczestnictwa w interwencji, ale nie wzięły w niej udziału. Zaletą tego rozwiązania jest to, że osoby te pochodzą zazwyczaj z tego samego rynku pracy co osoby poddane interwencji. Takie grupy kontrolne powinny być dokładnie przemyślane, zwłaszcza w przypadku, gdy projekt ewaluacji wpływu opartej na metodach kontrfaktycznych opiera się na parowaniu i gdy możliwe jest pozyskanie obszernych danych w celu podjęcia decyzji o wyborze. W innych okolicznościach, na przykład gdy wdrażana jest „różnica w różnicach”, grupy kontrolne określone na podstawie aspektu wyboru/świadomości będą mniej atrakcyjne.

Należy zwrócić uwagę na jeszcze jedną kwestię. W przypadku, gdy pomiary rezultatów uzyskane przed interwencją są dostępne dla grupy objętej interwencją i grupy kontrolnej, w ramach pomiaru rezultatów w odniesieniu do obu grup należy zbadać tendencje przed interwencją. Sprawdzenie założenia o „wspólnych tendencjach” rozwiązuje problem tymczasowych spadków wskaźników zatrudnienia i zarobków sprzed interwencji, które wystąpią w przypadku niektórych osób kwalifikujących się do aktywnej polityki rynku pracy (w przeciwnym razie nie kwalifikowałyby się one do wsparcia – tzw. problem spadku Ashenfeltera). Ewaluator poszukuje podobnych wykazywanych w czasie tendencji w pomiarach rezultatów zarówno w odniesieniu do grupy objętej interwencją, jak i grupy kontrolnej, tak aby nie pomylić poprawy sytuacji po utracie pracy krótkoterminowej lub zarobków z

*Analiza tendencji
przed interwencją*

⁶⁴ To właśnie [Card, D., Ibarrran, P. i Villa, J.M., 2011](#) określają mianem „błędu selekcji dwustronnej”.

długoterminowym względnym wzrostem, który ewaluacja wpływu oparta na metodach kontryfaktycznych próbuje wykryć.

Wybór odpowiednich grup kontrolnych jest zadaniem złożonym pod względem technicznym i metodologicznym. Zaleca się, aby podczas opracowywania programów ewaluacji urzędnicy zapoznali się z głównymi koncepcjami i odpowiednio wcześniej podjęli działania w celu wyłonienia potencjalnych grup kontrolnych. Ważne jest, aby zlecniodawcy ewaluacji zaangażowali ekspertów na wczesnym etapie projektowania w celu zapewnienia wsparcia i doradztwa.

Jakie są istotne kwestie związane z polityką?

Wybór odpowiedniej grupy kontrolnej nie jest jedynie procesem technicznym lub analitycznym. Chociaż aspekty analityczne związane z wyłonieniem odpowiednich grup kontrolnych mają zasadnicze znaczenie, ważne jest również, aby grupa kontrolna stanowiła odpowiednią alternatywę względem interwencji, z perspektywy kształtowania polityki.

Określenie alternatywy względem interwencji

Ewaluacje wpływu oparte na metodach kontryfaktycznych mogą przybierać różne formy: mogą na przykład polegać na porównaniu rezultatów w odniesieniu do grupy objętej interwencją lub kilku grup objętych interwencją z rezultatami grupy kontrolnej, która nie została poddana interwencji; lub mogą polegać na porównaniu jednej interwencji z drugą, bez uwzględniania grupy kontrolnej, która nie została poddana interwencji. Wybór grupy kontrolnej będzie uzależniony od tego, który rodzaj porównania jest najbardziej istotny z punktu widzenia polityki i czy wskazanie grupy kontrolnej „niepoddanej interwencji” jest w ogóle możliwe. W ramce 10 poniżej przedstawiono przykład porównania jednej interwencji z drugą, bez uwzględniania grupy kontrolnej „niepoddanej interwencji” – celem jest ocena, czy należy kontynuować jedną interwencję zamiast drugiej. Należy również zauważyć, że porównywanie jednego programu z innym, bez korzyści płynących z uwzględnienia grupy kontrolnej niepoddanej interwencji, może powodować niejasności (kwestię tę omówiono szczegółowo w ramce 11).

Porównanie interwencji z brakiem interwencji lub z rozwiązaniem alternatywnym

Ramka 10. Zagadnienia polityczne związane z programem szkoleniowym

Za przykład może posłużyć sytuacja, w której decydent zamierza wprowadzić nową interwencję w zakresie szkolenia, która ma być finansowana w ramach EFS – nazwijmy ją interwencją A. Załóżmy, że w państwie członkowskim istnieje już program szkoleniowy (B) skierowany do tych samych osób, ale finansowany z funduszy krajowych. W takim przypadku można mieć do czynienia z następującym problemem politycznym: czy poziom zatrudnienia i zarobków uczestników interwencji A jest wyższy niż poziom zatrudnienia i zarobków uczestników interwencji B po zakończeniu udziału w interwencji? A co za tym idzie, czy interwencja A oferuje lepszy stosunek jakości do ceny? Jeżeli zarobki są wyższe w przypadku uczestników interwencji A, wówczas oczywistym rozwiązaniem politycznym jest zaprzestanie stosowania interwencji B na rzecz interwencji A, jeżeli jej realizacja okaże się również efektywna kosztowo.

Ramka 11. Interpretacja efektów netto

Badanie może nie wykazać różnicy w zarobkach między uczestnikami interwencji A a uczestnikami interwencji B. Rozwiązanie polityczne dotyczące tej informacji może być niejasne, jeśli na przykład interwencja B była bardzo skuteczna w porównaniu z brakiem interwencji. Oznaczałoby to, że obie interwencje są bardzo skuteczne. W niektórych przypadkach może się jednak okazać, że nie istnieją dowody na skuteczność interwencji B w porównaniu z brakiem interwencji. Alternatywnie obie interwencje mogą być nieskuteczne, choć jedna z nich może wydawać się stosunkowo bardziej skuteczna od drugiej.

W przypadku gdy pewne grupy w populacji mogą zostać objęte większą liczbą interwencji, nadal więcej informacji może dostarczyć znalezienie odpowiedniej grupy jednostek niepoddanych interwencji, które posłużą jako punkt odniesienia.

Należy zauważyć, że różnicy w różnicach nie można wykorzystać do porównania wielu interwencji w przypadku braku grupy kontrolnej niepoddanej interwencji.

Jakie kwestie praktyczne są wymagane przy wyborze grupy kontrolnej?

Oprócz kwestii analitycznych i politycznych należy wziąć pod uwagę praktyczne aspekty wyboru grup kontrolnych. Wybór lub dobór próby (osób fizycznych lub przedsiębiorstw) mających należeć do grupy kontrolnej wymaga znalezienia odpowiedniego operatu losowania. Ponadto operaty losowania powinny obejmować poszczególne jednostki, które spełniają wymogi analityczne i polityczne. Najlepsze rozwiązanie będzie różniło się w zależności od konkretnej ewaluacji i specyficznego kontekstu analizowanej interwencji.

W wielu przypadkach przy określaniu odpowiednich grup kontrolnych często wykorzystuje się dwa źródła danych. W przypadku każdego z nich wymagane jest, aby tożsamość grupy objętej interwencją była znana.

Do wyszukiwania grup kontrolnych można wykorzystać różnego rodzaju ewidencje ludności. Na przykład jeżeli interwencja związana z aktywną polityką rynku pracy jest skierowana do osób w wieku 18–24 lat korzystających z zasiłku dla bezrobotnych, do zidentyfikowania populacji docelowej można wykorzystać ich rejestry świadczeń. Ponadto, jeżeli grupa poddana interwencji jest znana i można ją dopasować do danych dotyczących świadczeń, możliwe jest znalezienie osób w wieku 18–24 lat niepoddanych interwencji, a zatem takich, które mogą należeć do grupy kontrolnej. Kolejny przykład: przyjmijmy, że interwencja jest skierowana do małych i średnich przedsiębiorstw. Do zdefiniowania populacji docelowej można by wykorzystać krajowe rejestry przedsiębiorstw (o ile są dostępne), a dzięki dostępnym informacjom na temat przedsiębiorstw poddanych interwencji można wskazać potencjalne grupy kontrolne.

Ewidencja ludności i ewidencja podatkowa przedsiębiorstw

Dane dotyczące wnioskodawców mogą być wykorzystywane w przypadku, gdy interwencja nie jest powszechna, np. gdy nie wszystkie osoby ubiegające się o udział w programie szkoleniowym zostają przyjęte (grupa kontrolna określona na podstawie aspektu wyboru/świadomości). Podobnie nie wszystkie przedsiębiorstwa, które ubiegają się o finansowanie, zostaną wybrane, a przedsiębiorstwa, które nie zostały przyjęte do szkolenia lub finansowania, mogą w niektórych przypadkach zostać wykorzystane jako grupy kontrolne (zob. poprzednia część tej sekcji dotycząca zachowania niezbędnej ostrożności przy wyborze grup kontrolnych w takich okolicznościach).

Dane dotyczące wnioskodawców

2.3.4. Jakiego rodzaju kwestie dotyczące danych należy przedstawić w programie ewaluacji?

Zarządzanie źródłami danych

Jakie rodzaje danych są wymagane i w jaki sposób będą one gromadzone?

Odnotowano, że ewaluacje wpływu oparte na metodach kontrfaktycznych wymagają zazwyczaj dostępu do znacznych ilości danych jednostkowych (w niektórych przypadkach można wykorzystać dane grupowe, np. dane

regionalne). Dane te należy gromadzić, zestawiać i dokumentować; dane z różnych źródeł należy łączyć ze sobą na podstawie wspólnych obszarów identyfikacji; należy je przechowywać i przekazywać w bezpieczny sposób między podmiotami zarządzającymi ewaluacjami wpływu opartymi na metodach kontryfaktycznych i wykonującymi je; na podstawie tych źródeł danych należy utworzyć zestawy danych analitycznych, aby ułatwić oszacowanie wpływu.

Przy opracowywaniu programu ewaluacji należy rozważyć następujące kwestie dotyczące danych:

Podstawowa lista kontrolna dotycząca zarządzania danymi

Dostępność

- Jakie źródła można wykorzystać do uzyskania różnych rodzajów danych grupowych/jednostkowych?
- Czy dostępne są niezbędne dane indywidualne? Czy dotyczy to również „szczególnych kategorii” danych (w stosownych przypadkach)?
- Jeżeli zachodzi potrzeba przeprowadzenia badania, czy grupa docelowa i grupa kontrolna zostały zidentyfikowane w sposób umożliwiający monitorowanie ich za pomocą badań ankietowych – czy dane kontaktowe są dostępne i aktualne?

Spójność

- Czy istnieje jedno źródło danych, czy też konieczne jest połączenie źródeł danych (np. statystyki dotyczące bezrobocia, świadczeń społecznych, zabezpieczenia społecznego, dane dotyczące przedsiębiorstwa/zakładu itp.)
- Czy źródła są ze sobą spójne? Czy można w nich zidentyfikować w sposób spójny poszczególne osoby ze wszystkich źródeł?

Osiągalność

- Czy istnieje możliwość dostępu do źródeł danych krajowych dotyczących poszczególnych karier zawodowych, zarobków lub świadczeń społecznych w celu porównania uczestników EFS z potencjalną grupą kontrolną?

Umowy

- Czy obowiązują szczegółowe umowy w sprawie dostępności danych pomiędzy IZ a innymi właścicielami danych? Jakie bariery prawne lub organizacyjne należy omówić?
- Kto będzie odpowiadał za negocjowanie dostępu i uzyskanie zgody na wykorzystanie danych?

Przechowywanie

- Gdzie będą przechowywane dane? Jakie systemy i infrastruktura informatyczna będą wymagane?
- Jakie kroki zostaną podjęte w celu zapewnienia bezpiecznego przechowywania danych i zastrzeżenia dostępu dla osób, które potrzebują danych do celów ewaluacji?
- W jaki sposób dane są anonimizowane? Czy możliwe jest śledzenie osób w czasie i na podstawie powiązanych źródeł danych?

Zgodność z przepisami ochrony danych osobowych

- Czy dotychczasowe rozwiązania są zgodne z podstawowymi przepisami RODO i rozporządzenia w sprawie EFS+? Czy ewentualne przeszkody w

dostępności danych mają uzasadnienie w postaci stosowania różnych podstaw prawnych zgodnie z art. 6 lit. a), c) i e) RODO i art. 17 rozporządzenia w sprawie EFS+ oraz ewentualnego wykorzystania danych zanonimizowanych?

W jaki sposób będą przetwarzane dane?

W wielu przypadkach do ewaluacji wpływu opartej na metodach konfraktycznych będą potrzebne dane jednostkowe, tj. dane, które obejmują obserwacje dotyczące poszczególnych jednostek (zwykle poszczególnych osób fizycznych lub przedsiębiorstw) zarówno w grupie objętej interwencją jak i grupie kontrolnej (czasem można wykorzystać dane grupowe, np. dane regionalne lub dane na poziomie urzędu publicznych służb zatrudnienia). Rozróżniono trzy główne rodzaje wymaganych danych: a) dokumentacja dotycząca grupy objętej interwencją i grupie kontrolnej, b) dokumentacja dotycząca wyników oraz c) dane kontekstowe (dane wykorzystywane do sprawdzenia ważnych potencjalnych różnic pomiędzy grupą objętą interwencją i grupą kontrolną). Dane te mogą pochodzić z tego samego źródła lub z różnych źródeł. Źródła należy ustrukturyzować w celu utworzenia zbiorów danych analitycznych (lub próbek analitycznych), które są wykorzystywane do oszacowania wpływu. Ustrukturyzowanie to będzie w wielu przypadkach obejmować łączenie danych poszczególnych osób fizycznych lub przedsiębiorstw pochodzących z różnych źródeł. Takie łączenie wymaga zastosowania identyfikatorów na poziomie indywidualnym (np. indywidualnych numerów identyfikacyjnych zabezpieczenia społecznego), które umożliwiają dostosowanie danych dotyczących osoby (np. w danych podatkowych) do danych dotyczących uczestnictwa, albo wystarczającej ilości danych, aby powiązać dane z różnych źródeł (np. nazwisko i data urodzenia muszą być dostępne w różnych źródłach). Należy rozważyć, które źródła danych będą wykorzystywane na potrzeby planowanej ewaluacji wpływu opartej na metodach konfraktycznych, a także czy możliwe będzie łączenie danych z różnych źródeł.

Łączenie danych jednostkowych z różnych źródeł

2.3.5. Jakie są kluczowe ograniczenia w analizie danych i rezultatów?

Jak omówiono powyżej, wyniki ewaluacji wpływu opartych na metodach konfraktycznych określa się zwykle porównując rezultaty uzyskane w grupie objętej interwencją z rezultatami uzyskanymi w grupie kontrolnej, przy czym to podstawowe porównanie stanowi w istocie część podejścia do ewaluacji wpływu opartej na metodach konfraktycznych. Różnicę między grupą objętą interwencją a grupą kontrolną określa się jako wpływ lub efekt netto interwencji. Dokładny sposób szacowania wpływu będzie jednak zależał od przyjętego projektu badań.

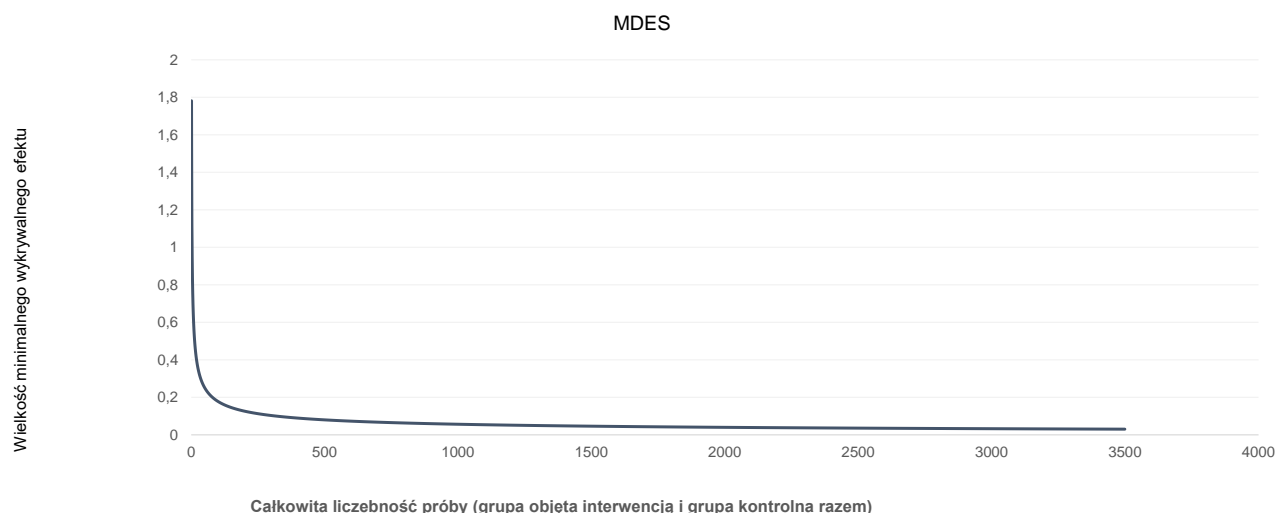
Przy planowaniu ewaluacji wpływu opartej na metodach konfraktycznych należy rozważyć, czy skala interwencji jest wystarczająco duża i czy ma ona szansę wywrzeć wpływ, który można wykryć metodami statystycznymi. W rzeczywistości metody konfraktycznej ewaluacji wpływu opierają się na zasadzie, że grupa docelowa i kontrolna stanowią próbę całości populacji potencjalnie poddanej interwencji i kontrolnej, przy czym liczba osób należących do obu grup musi być statystycznie istotna, aby można było uzyskać znaczące wyniki.

Przy rozpatrywaniu, czy liczebność próby jest wystarczająco duża do przeprowadzenia analizy, przydatnym pojęciem jest „minimalny wykrywalny efekt”.⁶⁵ Prościej rzecz ujmując, minimalny wykrywalny efekt to najmniejszy rzeczywisty wpływ, jaki można wykryć w przypadku próby o danej liczebności przy standardowych poziomach ufności i mocy statystycznej⁶⁶.

Ocena liczebności próby i wielkości efektu

Na wykresie 4 poniżej przedstawiono, zmiany minimalnego wykrywalnego efektu w zależności od całkowitej liczebności próby (całkowita liczba prób w grupie objętej interwencją i grupie kontrolnej) na poziomie ufności wynoszącym 95% i na poziomie mocy statystycznej wynoszącym 80%. Przesuwając się od lewej do prawej, minimalny wykrywalny efekt spada gwałtownie w miarę jak liczebność próby zbliża się do 500 (250 jednostek należących do grupy objętej interwencją i 250 jednostek należących do grupy kontrolnej). Innymi słowy, wraz ze wzrostem całkowitej liczebności próby projekt ewaluacji wpływu opartej na metodach kontrfaktycznych staje się bardziej precyzyjny i można wykorzystać go do wykrycia mniejszego wpływu.

Wykres 4. Minimalne wykrywalne efekty w próbach o różnej liczebności



Uwaga: 95% statystycznie istotne i 80% mocy statystycznej. Na wykresie przyjęto projekt randomizowany

W planowaniu ewaluacji wpływu opartej na metodach kontrfaktycznych przydatne jest oszacowanie prawdopodobnej liczebności prób w celu określenia grupy poddanej interwencji i grupy kontrolnej; oszacowanie to opiera się na prognozach dotyczących liczby jednostek, które zostaną poddane interwencji, projektowaniu prawdopodobnej metody, która zostanie wykorzystana do ewaluacji wpływu opartej na metodach kontrfaktycznych, a także liczebności odpowiednich grup kontrolnych utworzonych na podstawie dostępnych danych administracyjnych lub odpowiedniego badania. Przyjmując pewne założenia, informacje te można następnie wykorzystać do

Prawdopodobna liczebność próby

⁶⁵ Bloom H. S., 1995, udziela praktycznych wskazówek dotyczących obliczania minimalnych wykrywalnych efektów dla projektów eksperymentalnych. W przypadku podejść quasi-eksperymentalnych takie obliczenia będą wymagały korekty. Zasadniczo podejścia quasi-eksperymentalne wymagają większej liczebności prób w porównaniu z tymi, które są niezbędne do projektu eksperymentalnego.

⁶⁶ Poziom ufności jest statystyczną miarą wiarygodności procedury szacowania, natomiast poziom mocy statystycznej oznacza prawdopodobieństwo, że test wykryje rzeczywisty efekt. Te dwa parametry zależą również od liczebności próby i na odwrót, można je wykorzystać do obliczenia liczebności próby w odniesieniu do odpowiednich poziomów istotności. Na ogół poziom ufności wynosi 95%, a moc statystyczna 80%.

sprawdzenia, czy wynikające z nich minimalne wykrywalne efekty są wystarczająco znaczące oraz czy wdrożenie ewaluacji wpływu opartej na metodach kontrfaktycznych jest rozsądnym wyborem. Kluczowym aspektem tego zadania jest to, czy dana interwencja może przynieść efekty o skali odpowiadającej szacowanym minimalnym wykrywalnym efektom; w tej kwestii pomocna może być analiza istniejącej literatury i innych podobnych ewaluacji.

Należy również zwrócić uwagę, że często celem ewaluacji wpływu opartych na metodach kontrfaktycznych jest zbadanie efektów w różnych subpopulacjach uczestniczących w interwencji (mężczyźni/kobiety, młodzież/osoby starsze, osoby gorzej/lepiej wykształcone itp.) Analizy te z konieczności wymagają ograniczenia liczebności grup poddanych interwencji i grup kontrolnych dla każdej subpopulacji (mężczyźni, kobiety, młodzież itp.), a w konsekwencji obniża się poziom istotności. Gdy badanie skutków dla subpopulacji stanowi ważny składnik ewaluacji, w każdej subpopulacji potrzebna jest wystarczająco duża liczba osób należących do grupy poddanej interwencji i grupy kontrolnej.

Określenie liczebności próby wymaga umiejętności technicznych, którymi instytucja zarządzająca nie zawsze dysponuje. W takim przypadku ewaluator musi obliczyć liczebność próby i uzgodnić z instytucją zarządzającą wykonalność i zakres ewaluacji wpływu opartej na metodach kontrfaktycznych. Nie jest możliwe określenie progu liczebności próby, powyżej którego przeprowadzenie ewaluacji wpływu opartej na metodach kontrfaktycznych jest zawsze wykonalne, ponieważ próg ten może różnić się w zależności od przyjętych metod, przewidywanych analiz dotyczących subpopulacji, przyjętych poziomów istotności i innych elementów. W wielu ewaluacjach liczba osób wynosząca 2 000 w odniesieniu do łącznej liczby osób w grupie poddanej interwencji i grupie kontrolnej okazała się jednak wystarczająca do przeprowadzenia dość szczegółowej analizy (w tym głównych podgrup, takich jak mężczyźni/kobiety itp.). Wskazanie to nie może być stosowane jako „próg” naukowy, lecz jedynie jako pomoc w przeprowadzeniu analizy potencjalnego wdrożenia ewaluacji wpływu opartej na metodach kontrfaktycznych.

Liczebność próby ma wpływ na wiarygodność rezultatów ewaluacji wpływu opartej na metodach kontrfaktycznych. Niektóre przykłady niepewności dotyczącej interpretacji wyników i ich związek z liczebnością próby wyjaśniono w ramce poniżej.

Ramka 12. Niepewność dotycząca interpretacji wyników

Wśród przykładów ewaluacji interwencji finansowanych w ramach EFS, ewaluacja szeregu działań na rzecz długotrwale bezrobotnych w regionie Marche we Włoszech nie przyniosła statystycznie istotnych rezultatów w przypadku jednego z czterech analizowanych działań – „talonów szkoleniowych”. Wynika to prawdopodobnie z faktu, że liczebność analizowanej próby osób poddanych interwencji była niewielka w porównaniu z dużą liczbą rzeczywistych świadczeniobiorców. Podobny przypadek miał miejsce podczas ewaluacji działań promujących kształcenie zawodowe w Polsce w województwie podlaskim. Pomijając kwestię poziomu istotności, przykład ten pokazuje, że podczas analizy rezultatów poleganie na zbyt małej liczebności prób grup objętych interwencją może prowadzić do problemów związanych z uogólnianiem ustaleń.

Ewaluacja szkoleń językowych dla imigrantów w Niemczech wykazuje pozytywne wyniki, w przeciwieństwie do niektórych wcześniejszych badań oceniających wpływ programów szkoleń językowych na imigrantów. Zdaniem autora potencjalnym wyjaśnieniem tej różnicy jest fakt, że analizowany niemiecki program składał

się również z modułu doświadczenia zawodowego. Hipoteza ta nie mogła być przetestowana przez autora na podstawie dostępnych danych ze względu na brak możliwości rozróżnienia komponentów interwencji dotyczących szkolenia językowego i doświadczenia zawodowego. Może to stanowić potencjalny przedmiot zainteresowania w przyszłych badaniach. Przykład ten pokazuje, że w celu wyjaśnienia mechanizmów, które decydują o rezultatach interwencji, scenariusze kontrfaktyczne należy powtarzać z wykorzystaniem lepszych i bardziej obszernych danych lub uzupełniać je innymi metodami ewaluacji.

2.3.6. Lista kontrolna służąca weryfikacji przygotowania i wykonalności ewaluacji wpływu opartej na metodach kontrfaktycznych

W tym miejscu warto podsumować na podstawowej liście kontrolnej główne czynniki, które instytucja zarządzająca powinna rozważyć i zweryfikować podczas przygotowywania ewaluacji wpływu opartej na metodach kontrfaktycznych:

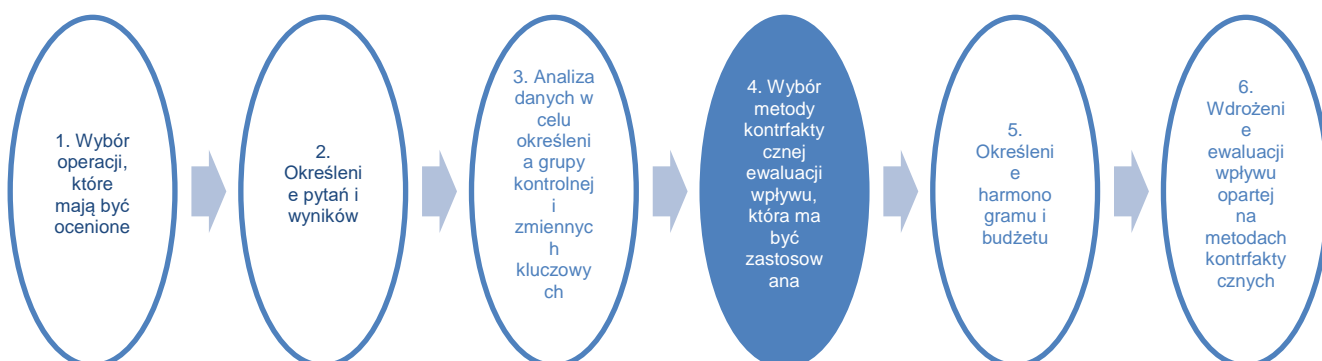
Wykaz głównych czynników, które należy wziąć pod uwagę

- Czy interwencja wybrana na potrzeby ewaluacji jest odpowiednia w przypadku ewaluacji wpływu opartej na metodach kontrfaktycznych?
 - o Czy interwencja jest odrębna, czy wyróżnia się na tle innych i czy jest względnie jednolita?
 - o Czy „teoria zmiany” w odniesieniu do interwencji sugeruje przekonujący mechanizm przyczynowo-skutkowy dotyczący wyników, które należy zbadać w ramach ewaluacji wpływu opartej na metodach kontrfaktycznych?
 - o Czy grupę kontrolną można łatwo określić zgodnie z zasadami interwencji?
- Czy wpływ na uczestników można ująć ilościowo?
 - o Czy mierzone rezultaty interwencji są zrozumiałe i zgodne z teorią zmiany interwencji?
 - o Czy od zakończenia interwencji upłynęło wystarczająco dużo czasu, aby można było określić jej wpływ na uczestników?
- Czy pytania ewaluacyjne wymagają pomiaru wpływu (efektów netto) interwencji?
- Czy odpowiednie dane (dane dotyczące grupy objętej interwencją i grupy kontrolnej, dane dotyczące wyników, dane kontekstowe) są dostępne lub czy można je udostępnić?
 - o Czy dostępne dane są zgodne z RODO? Jeśli nie, czy można wprowadzić niezbędne porozumienia międzyadministracyjne i rozwiązania techniczne, aby dane były zgodne z przepisami?
 - o Czy na podstawie dostępnych danych można określić odpowiednią grupę kontrolną?
 - o Czy liczebność próby jest wystarczająco duża, aby osiągnąć wymagany poziom istotności w odniesieniu do wyników ewaluacji wpływu opartej na metodach kontrfaktycznych?

Udzielenie odpowiedzi twierdzących na wszystkie powyższe pytania oznacza, że zaprojektowanie i wdrożenie ewaluacji wpływu opartej na metodach kontrfaktycznych jest możliwe. Udzielenie odpowiedzi przeczących na część z tych pytań nie wyklucza wdrożenia ewaluacji wpływu opartej na metodach

kontrfaktycznych, ale wymaga dodatkowej kontroli warunków wykonalności oraz poprawy dostępności danych lub innych kluczowych czynników.

2.4. Metoda kontrfaktycznej ewaluacji wpływu, która ma być zastosowana



Wybór metody, która zostanie wykorzystana do ewaluacji wpływu opartej na metodach kontrfaktycznych, jest decydującym krokiem w przygotowaniu ewaluacji, ponieważ wybrana metoda wpływa zarówno na jakość ustaleń, jak i na jakość ogólną. Zasadniczo żadna metoda nie jest lepsza od innej, ale jedna może być bardziej odpowiednia do wykorzystania dostępnych danych, podczas gdy ta inna może być bardziej właściwa do zbadania określonego rodzaju interwencji; wskazanie najbardziej odpowiedniej metody określonej ewaluacji stanowi decydujący krok w procesie projektowania.

Niektóre podstawowe warunki dotyczące określenia najbardziej odpowiedniej metody

Aspekty techniczne i nieodłączna charakterystyka każdej metody zostały szczegółowo omówione w **rozdziale 3** – w tym miejscu należałoby pokrótce zastanowić się nad praktycznymi konsekwencjami wyboru.

Po pierwsze, przed rozpoczęciem interwencji należy ustalić podejście oparte na randomizacji, w ramach którego poszczególne osoby są losowo przydzielane do grupy objętej interwencją lub kontrolnej. Zakłada to zarówno ukończenie projektu ewaluacji, jak i organizację umożliwiającą wykonanie tego zadania. W interwencjach w ramach EFS podejście to było rzadko stosowane, ale, jak się okaże, jest ono możliwe w pewnych warunkach i bardzo skuteczne z metodycznego punktu widzenia.

Po drugie, wybór metody kontrfaktycznej ewaluacji wpływu może wykraczać poza możliwości instytucji zarządzającej, ponieważ wymaga zaawansowanych umiejętności technicznych i szczególnego doświadczenia. W takim przypadku ewaluator wybierze najbardziej odpowiednią metodę, ale oczywiście dopiero wtedy, gdy zostanie wyznaczony. Może to przynieść dodatkową korzyść polegającą na działaniu jako kryterium oceny zdolności osób oceniających w procedurze wyboru ewaluatora. W każdym razie instytucja zarządzająca musi mieć poczucie pewności przeprowadzając ewaluację wpływu opartą na metodach kontrfaktycznych po udzieleniu odpowiedzi twierdzących na wszystkie lub większość pytań z poprzedniej listy kontrolnej oraz nabyć – wewnątrz lub zewnątrz – umiejętności niezbędne do zatwierdzenia wyboru ewaluatora.

Po trzecie, metody oparte na randomizacji mogą różnić się od metod quasi-eksperymentalnych pod względem czasu trwania i kosztów, a w badaniu randomizowanym z grupą kontrolną ewaluator powinien być zaangażowany od początku interwencji. Metody quasi-eksperymentalne nie różnią się znacząco od siebie; ewaluator może przystąpić do pracy po rozpoczęciu

interwencji (ale w czasie umożliwiającym zorganizowanie i przygotowanie niezbędnych danych), a jakość dostępnych danych może różnić się w zależności od czasu i kosztów.

Ponadto rodzaj interwencji i dostępność danych umożliwiają czasami korzystanie z większej liczby metod quasi-eksperymentalnych niż jedna. W tym przypadku przydatne może być zastosowanie wielu metod, ponieważ potwierdzenie tych samych efektów przy użyciu różnych metod zwiększa wiarygodność tych ustaleń.

2.5. Harmonogram i budżet



2.5.1. Jakie zasoby są dostępne?

Kluczową kwestią, którą należy rozważyć przy opracowywaniu programu ewaluacji wpływu opartej na metodach kontrfakcyjnych, jest dostępność zasobów. Może to być szeroko zakrojony zestaw zagadnień, które podzielono na trzy działy: a) zasoby ekspertów; b) czas; oraz c) zasoby finansowe.

Jacy eksperci zewnętrzni i pracownicy wewnętrzni są niezbędni do celów ewaluacji wpływu opartej na metodach kontrfakcyjnych?

W większości przypadków ewaluacja wpływu zostanie zlecona dostawcy zewnętrznemu. Zleceniem muszą jednak zarządzać pracownicy w ramach instytucji zarządzającej, którzy dysponującą wiedzą na temat metod kontrfakcyjnej ewaluacji wpływu. Taka wiedza jest wymagana dla zapewnienia jakości i skutecznej współpracy z ekspertami zewnętrznymi. W ramach instytucji zarządzającej mogą być również wymagane inne formy wiedzy specjalistycznej, takie jak umiejętności statystyczne oraz wiedza specjalistyczna w zakresie gromadzenia danych i zarządzania nimi. Ważne jest, aby wcześniej rozważyć, czy instytucja zarządzająca ma odpowiednio wykwalifikowany i przeszkolony personel oraz czy personel ten posiada zdolność do wspierania ewaluacji.

Personel wewnętrzny

Zlecenie wykonania skutecznej ewaluacji wpływu opartej na metodach kontrfakcyjnych wymaga wykonawców, którzy posiadają niezbędne umiejętności i doświadczenie w przeprowadzeniu takich ewaluacji. Ponadto odpowiedni wykonawcy będą musieli rozumieć kontekst polityczny i administracyjny w państwach członkowskich, znać potencjalne źródła danych oraz biegle posługiwać się odpowiednimi językami. Warto zastanowić się, co jest konieczne do rozwinięcia bazy dostawców na potrzeby ewaluacji wpływu opartej na metodach

Personel zewnętrzny

kontrafaktycznych w danym państwie członkowskim (więcej informacji na ten temat znajduje się w rozdziale 4).

Skuteczne ewaluacje wpływu oparte na metodach kontrafaktycznych wymagają współpracy ze strony osób zarządzających ocenianym programem lub interwencją. Na przykład konieczny będzie dostęp do ewidencji prowadzonej przez kierowników interwencji. Ewidencja ta stanowi źródło informacji o osobach lub przedsiębiorstwach, które uczestniczyły w interwencji.

Personel zarządzający programami/interwencjami

Kierownicy programu/interwencji mogą udzielić porad i wskazówek dotyczących tego typu danych. Oprócz działań, które musieliby wykonywać w przypadku braku ewaluacji wpływu, do ich zadań może również należeć prowadzenie ewidencji.

Aby rozwiązać problem gromadzenia danych z różnych źródeł, osoby planujące ewaluację wpływu opartą na metodach kontrafaktycznych będą musiały nawiązać współpracę z pracownikami zajmującymi się oficjalnymi źródłami danych (np. ewidencją bezrobotnych, danymi dotyczącymi zabezpieczenia społecznego, urzędami statystycznymi itp.) Umożliwi to planowanie w zakresie przekazywania danych z dużym wyprzedzeniem.

Wiedza fachowa w dziedzinie statystyki

Jakie czynniki są istotne przy ustalaniu harmonogramu ewaluacji wpływu opartej na metodach kontrafaktycznych?

Przeprowadzenie ewaluacji wpływu opartej na metodach kontrafaktycznych wymaga zaangażowania wielu różnych zasobów ludzkich; ponadto takie ewaluacje przeprowadza się w okresach o znacznej długości trwania. Program ewaluacji powinien obejmować ogólny harmonogram z najważniejszymi celami pośrednimi projektu, które będą dotyczyły zarówno osób zaangażowanych w samą interwencję, jak i osób związanych z ewaluacją. Konieczne będzie ustalenie ogólnego harmonogramu z uwzględnieniem zarówno ewaluacji, jak i działań związanych z przeprowadzeniem interwencji, a także kluczowych celów pośrednich polityki.

CIE są czasochłonne

Opracowanie znaczącego i realistycznego harmonogramu ewaluacji wpływu opartej na metodach kontrafaktycznych jest trudnym zadaniem. Z jednej strony, instytucja zarządzająca (lub instytucja pośrednicząca) planująca ewaluację musi wziąć pod uwagę kluczowe terminy, w których trzeba będzie podjąć decyzje zależne od ustaleń poczynionych w trakcie ewaluacji. Z drugiej strony, nieuniknione będą ograniczenia, które będą miały wpływ na terminowość sprawozdań. Na niektóre rezultaty trzeba będzie czekać latami, a harmonogramy gromadzenia danych, analizy i sprawozdawczości będą musiały to w jak największym stopniu odzwierciedlać (zob. sekcja 2.5.2 poniżej).

Planowanie opóźnienia czasowego, aby umożliwić wystąpienie skutków

Ewaluację należy przeprowadzić odpowiednio wcześniej w okresie programowania, aby umożliwić wprowadzenie zmian oraz wykorzystanie zdobytych doświadczeń i wniosków w pozostałym czasie. W niektórych okolicznościach takie same lub podobne interwencje można wesprzeć w kolejnych okresach programowania. Rezultaty ewaluacji wpływu opartej na metodach kontrafaktycznych koncentrujące się na interwencjach z poprzednich okresów

Skupienie się na konkretnych momentach

programowania mogą być niezwykle pomocne w informowaniu o wdrażaniu i projektowaniu w kolejnych okresach programowania.

Należy również rozważyć, w jaki sposób termin przeprowadzenia ewaluacji kontrfaktycznej może być powiązany z terminem przeprowadzenia innych elementów ewaluacji. Prawdopodobnie przed przeprowadzeniem ewaluacji wpływu opartej na metodach kontrfaktycznych należałoby przeprowadzić ewaluację opartą na teorii. W przypadku innowacyjnych interwencji (np. interwencji w ramach EFS na lata 2014–2020, które wspierały dzieci w Republice Czeskiej, lub usług wspierających włączenie społeczne dla osób otrzymujących nowe wsparcie dochodu we Włoszech, lub poszczególnych środków wspierających pracowników i studentów w pracy zdalnej podczas obowiązywania środków izolacji związanych z pandemią COVID-19 w wielu państwach członkowskich), może zaistnieć potrzeba wcześniejszego zgłoszenia kluczowych elementów ewaluacji procesu, aby umożliwić ulepszenie projektu ewaluacji wpływu opartej na metodach kontrfaktycznych, albo w późniejszym terminie, aby ułatwić szczegółowe zbadanie przyczyn i skutków. W przypadku przeprowadzania ewaluacji wpływu opartej na metodach kontrfaktycznych dla zaawansowanej, trwającej interwencji, należałoby przeprowadzić ewaluację procesu równoległe z ewaluacją wpływu.

Kolejność różnych rodzajów ewaluacji

Dostępność danych będzie również miała wpływ na harmonogram. Aktualizacja źródeł danych może być czasochłonna, jak ma to często miejsce w przypadku danych podatkowych. Eliminowanie prawnych i instytucjonalnych barier utrudniających pozyskiwanie wymaganych danych może być również czasochłonne i kosztowne. Ponadto wykorzystywanie danych z różnych źródeł, zapewnienie ich spójności, sprawdzenie ich jakości i przekształcenie ich w formę, którą można stosować do szacowania wpływu, wymaga znacznego czasu i wysiłku.

Gromadzenie danych może być czasochłonne

Jak można oszacować koszty?

Ważne jest, aby ustalić orientacyjny budżet, jaki instytucja zarządzająca jest w stanie i chce przeznaczyć na ewaluację wpływu opartą na metodach kontrfaktycznych. Budżet ten składa się z dwóch elementów: kosztów wewnętrznych i zewnętrznych. Pierwsze z nich odnoszą się do wysiłku związanego z zasobami wewnętrznymi w celu przeprowadzenia ewaluacji; w ramach administracji koszty te zazwyczaj nie są wyszczególnione w kategoriach finansowych, ale należy je uwzględnić w celu zapewnienia odpowiedniej kontroli i działań następczych w związku z ewaluacją. Koszty zlecenia ekspertom zewnętrznym przeprowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych należy starannie oszacować, aby zapewnić wysoką jakość ewaluacji przy rozsądnych wydatkach. W tym przypadku nacisk kładzie się na koszty zewnętrzne.

Należy dokonać rozróżnienia między ewaluacjami interwencji rutynowych, w przypadku których wydatki są zazwyczaj niższe, a działaniami innowacyjnymi lub pilotażowymi, w przypadku których gromadzenie stosunkowo dużej ilości danych, wykorzystanie nowych źródeł danych oraz zaangażowanie nowych zainteresowanych stron może wiązać się z wyższymi wydatkami. Nie jest to jednak regułą i należy to rozpatrywać w kontekście pytań ewaluacyjnych, których

Potencjalna metoda oszacowania kosztów

złożoność i liczba może wymagać mniejszych lub większych nakładów finansowych.

Wybór podejścia do ewaluacji również ma znaczenie. Jak wspomniano powyżej, podejście oparte na randomizacji wymaga obecności ewaluatora w celu gromadzenia i weryfikowania informacji w całym okresie trwania interwencji. Z drugiej strony, wysiłek włożony w metody quasi-eksperymentalne zależy w dużej mierze od liczby źródeł danych, ich jakości i dostępności.

Wytyczne wydane przez Komisję⁶⁷ zawierają wskazówki dotyczące kwoty potrzebnej do przeprowadzenia ewaluacji programu i określono w nich, że programy o dużej skali i programy rutynowe powinny przeznaczać na ocenę nie więcej niż 1% budżetu programu, natomiast inicjatywy innowacyjne lub pilotażowe mogą przeznaczać do 10%. Wytyczne te nie odnoszą się jednak wyraźnie do zapotrzebowania na zasoby w zakresie ewaluacji wpływu opartej na metodach kontrfaktycznych i można założyć, że te wartości procentowe stanowią pułap kosztów ewaluacji obejmujący jedną lub większą liczbę ewaluacji wpływu opartych na metodach kontrfaktycznych.

Jak można sobie wyobrazić, podanie orientacyjnego kosztu ewaluacji wpływu opartej na metodach kontrfaktycznych nie jest możliwe dla tak ogromnej liczby różnych interwencji i warunków operacyjnych w ramach programu EFS+ w 27 państwach członkowskich. W tym przypadku bardziej przydatne jest zaproponowanie metody szacowania rozsądnego budżetu dla ewaluacji wpływu opartej na metodach kontrfaktycznych w różnych kontekstach. W poniższej tabeli przedstawiono proste podejście: w wierszach wyszczególniono główne zadania do wykonania, a w kolumnach koszty podstawowe.

Czas trwania i złożoność każdego działania określa konieczny nakład pracy pod względem dni roboczych zespołu ewaluacyjnego, a co za tym idzie, jego koszt podstawowy. Koszty gromadzenia danych w formie badań ilościowych dotyczących uczestników i członków grupy kontrolnej są znaczne. W przypadku gdy ewaluacja wpływu oparta na metodach kontrfaktycznych opiera się na istniejących źródłach danych administracyjnych, które są dość łatwo dostępne, całkowite koszty będą niższe. Dane administracyjne często wymagają jednak niezbędnych działań „przygotowawczych” (korekta brakujących lub błędnych danych, dostosowanie formatu bazy danych, weryfikacja przepisów administracyjnych regulujących dane itp.).

⁶⁷ Zob. [Komisja Europejska, 2013.](#)

Tabela 3. Struktura kosztów podstawowych CIE

Główne działania	Zespół ewaluacyjny			Inne koszty (sprzęt, materiały, dostawy, podróże itp.)	Uwagi
	Kierownik programu	Starsi eksperci	Młodszy eksperci		
Planowanie i koordynacja					Proces planowania obejmuje wykonalność ewaluacji wpływu opartej na metodach kontrfaktycznych, jej organizację oraz ostateczną wersję metodyki w porozumieniu z instytucją zarządzającą. Na tym etapie należy poświęcić wystarczająco dużo czasu na analizę potencjalnych luk w danych. Koordynacja obejmuje cały czas trwania ewaluacji wpływu opartej na metodach kontrfaktycznych i uwzględnia organizację pracy oraz interakcje z instytucją zarządzającą i innymi odpowiednimi zainteresowanymi stronami w celu zakończenia prac nad pytaniami ewaluacyjnymi.
Przegląd literatury					Przegląd literatury uzupełnia zarys teorii zmiany, identyfikację zmiennych, które zostaną wykorzystane oraz zrozumienie kontekstu. Umożliwia również dopracowanie pytań ewaluacyjnych oraz wykorzystanie dotychczasowych wyników i sformułowanie hipotez.
Gromadzenie i przygotowanie danych					Starania wymagane w celu podjęcia tego działania mają bardzo zróżnicowany charakter w zależności od metody lub metod gromadzenia danych. Gromadzenie danych pierwotnych w oparciu o badania może być czasochłonne i kosztowne. Inne metody mogą wiązać się z inwestycjami w technologię (oprogramowanie lub sprzęt) lub zawieraniem umów z właścicielami danych. Koszty można zredukować poprzez włączenie właścicieli danych w ich przygotowanie.
Analiza danych					Analiza danych wymaga wysokich kwalifikacji i doświadczenia. Czas potrzebny do wykonania tego działania jest uzależniony od metody, ilości analiz i jakości danych.
Przygotowanie sprawozdania lub sprawozdań					Wysiłek związany z tym działaniem różni się w zależności od liczby i rodzaju sprawozdań oraz innych narzędzi komunikacji. Koszty mogą obejmować druk i projekt graficzny, a także czas przygotowania.
Spotkania uzupełniające					Spotkania uzupełniające oraz inne działania mające na celu przekazywanie informacji stanowią ważny krok w upowszechnianiu ustaleń. Koszty różne mogą obejmować wynajem powierzchni i wyżywienie.
Ogółem					

W przypadku braku innych warunków zamówienia koszt personelu można obliczyć odnosząc się do kosztów EFS+ związanych z usługami szkoleniowymi lub służbami zatrudnienia przyjętymi w każdym państwie w odniesieniu do starszych i młodszych instruktorów lub ekspertów. Na ogół wynagrodzenia te są porównywalne z wynagrodzeniami młodszego i starszego badacza i nawet jeśli nie muszą być równoważne, stanowią przydatny parametr, który należy wziąć pod uwagę. Prosta analiza rynku i kilka wywiadów z naukowcami może pomóc w określeniu przewidywanego kosztu ewaluacji wpływu opartej na metodach kontrfaktycznych.

2.5.2. Kiedy należy dokonać ewaluacji interwencji?

Kluczowe znaczenie ma ustalenie, kiedy w trakcie trwania interwencji należy przeprowadzić ewaluację wpływu, jak również określenie krytycznych kwestii, tj. kiedy rezultaty należy zmierzyć, a wpływ oszacować.

Kiedy należy przeprowadzić ewaluację nowych i trwających interwencji?

Dyskusja na temat tego, kiedy w trakcie trwania interwencji należy przeprowadzić ewaluację wpływu opartą na metodach kontrfaktycznych, będzie uzależniona od tego, czy interwencja jest nowa, czy jest to zaawansowany, trwający program. W przypadku nowej interwencji potrzeba więcej czasu, aby osiągnęła zaawansowany poziom i stan stabilności. Przeprowadzenie ewaluacji wpływu opartej na metodach kontrfaktycznych przed tym momentem będzie przedwczesne i może być źródłem wprowadzających w błąd dowodów. W przypadku nowych interwencji wstępna ewaluacja procesu przeprowadzona przed dokonaniem ewaluacji wpływu opartej na metodach kontrfaktycznych jest często przydatna do celów zidentyfikowania początkowych problemów i wskazania rozwiązań.

Różne terminy w przypadku nowych i trwających interwencji

Przy określaniu optymalnego terminu, w jakim należy przeprowadzić ewaluację wpływu opartą na metodach kontrfaktycznych dla nowej interwencji, należy wziąć pod uwagę szereg innych czynników, w tym: kroki niezbędne w celu zapewnienia dostępności odpowiednich źródeł danych, powołanie wewnętrznego zespołu ds. projektu składającego się z odpowiednio przeszkolonego personelu oraz wyznaczenie wykonawcy zewnętrznego. Ponadto kluczowym ograniczeniem będą potrzeby związane z procesem decyzyjnym, na który ostatecznie ukierunkowana jest ewaluacja.

Kwestie związane z zasobami

Termin, w jakim należy przeprowadzić ewaluację wpływu trwającej interwencji, będzie wynikał głównie z wymogów praktycznych i politycznych. Wtedy interwencja powinna być już utrwalona i powinna osiągnąć zaawansowany poziom pozwalający na zastosowanie ewaluacji wpływu opartej na metodach kontrfaktycznych. Kolejną kwestią, którą należy rozważyć, są inne reformy przeprowadzane równoległe z ocenianą interwencją. Efekty tych reform mogą wpłynąć na wpływ rozważanej interwencji. Decydenci będą musieli przeanalizować, czy przeprowadzanie innych reform w obszarze polityki ma znaczenie dla decyzji politycznych, których uzasadnienie będą stanowiły wyniki przedmiotowej ewaluacji wpływu opartej na metodach kontrfaktycznych.

Ewaluacje dotyczące EFS koncentrują się zazwyczaj na jednym okresie programowania. Szczególnie w przypadku stabilnych interwencji, które stanowiły już część programu EFS w poprzednim okresie, przydatne mogłoby okazać się połączenie oceny retrospektywnej poprzedniego okresu z trwającą ewaluacją bieżącego okresu w celu objęcia ewaluacją dłuższego okresu funkcjonowania interwencji.

Wiele okresów programowania

Kiedy należy mierzyć wyniki i obliczać wpływ?

Drugą kluczową kwestią związaną z terminem, w jakim należy przeprowadzić ewaluację jest to, kiedy należy zmierzyć i oszacować wpływ lub, mówiąc dokładniej, przewidywany czas wystąpienia skutków w następstwie interwencji.

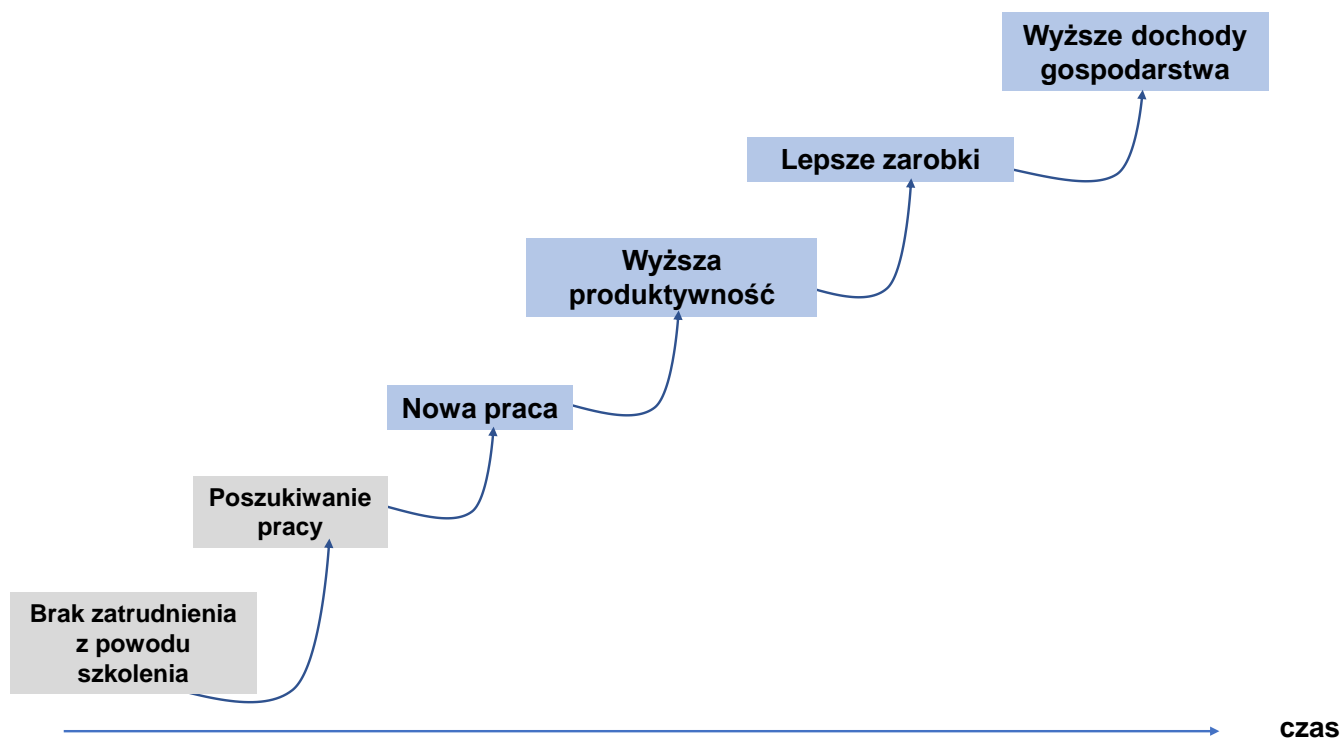
Analizując interwencję w zakresie szkolenia skierowaną do osób bezrobotnych, należy zadać pytanie, kiedy można oczekiwać wyższych wskaźników zatrudnienia. Dobrze znaną cechą programów szkoleniowych

jest to, że w krótkim okresie mają one tendencję do zmniejszania zatrudnienia uczestników. Wynika to z tak zwanego efektu „lock-in”. Interwencje w zakresie szkoleń mają tendencję do odwracania uwagi bezrobotnych uczestników od poszukiwania pracy ze względu na ich udział w kursach. Z kolei osoby w grupie kontrolnej są zaangażowane w znalezienie pracy. Jeżeli wpływ zostanie obliczony zbyt wcześnie, może być on negatywny lub niedoszacowany. Planując ewaluację wpływu opartą na metodach kontrfaktycznych ważne jest, aby realistycznie podejść do kwestii czasu wystąpienia efektów oraz do tego, kiedy mogą być one wymierne. Uproszczony model dotyczący późniejszych skutków przedstawiono na wykresie 5 poniżej.

Rozważając, kiedy najlepiej zmierzyć wyniki i oszacować wpływ, trzeba będzie uwzględnić wymogi decydentów dotyczące informacji w określonych terminach. W przypadku interwencji, których celem jest poprawa długoterminowej zdolności do zatrudnienia, z perspektywy analitycznej rozsądnym posunięciem może się okazać obserwowanie uczestników przez okres dwóch lat lub dłużej po objęciu ich interwencją, aby sprawdzić, czy ich zarobki i wskaźniki zatrudnienia są wyższe niż w przypadku równoważnej grupy osób niepoddanych interwencji. Z kolei zarządzający programami często potrzebują natychmiastowych ustaleń, w którym to przypadku należy osiągnąć kompromis między tym, co jest rozsądnym okresem działań następczych z punktu widzenia interwencji, a potrzebą decydentów, aby dowody były dostępne na czas.

Między rozsądnym a wykonalnym

Wykres 5. Uproszczony harmonogram dotyczący rezultatów programu szkoleniowego



Jeżeli pomiary wyników uzyskuje się ze źródeł administracyjnych (np. ewidencji ubezpieczeń społecznych zawierającej szczegółowe informacje na temat zatrudnienia i zarobków), praktycznym rozwiązaniem będzie ciągłe monitorowanie wyników przez dłuższy czas i szacowanie wpływu (nawet raz w miesiącu). Ryzyko polega na tym, że charakter ustaleń może z czasem ulec zmianie. Jeżeli do pomiaru wyników wymagane jest gromadzenie danych

Skupienie się na konkretnych momentach w czasie

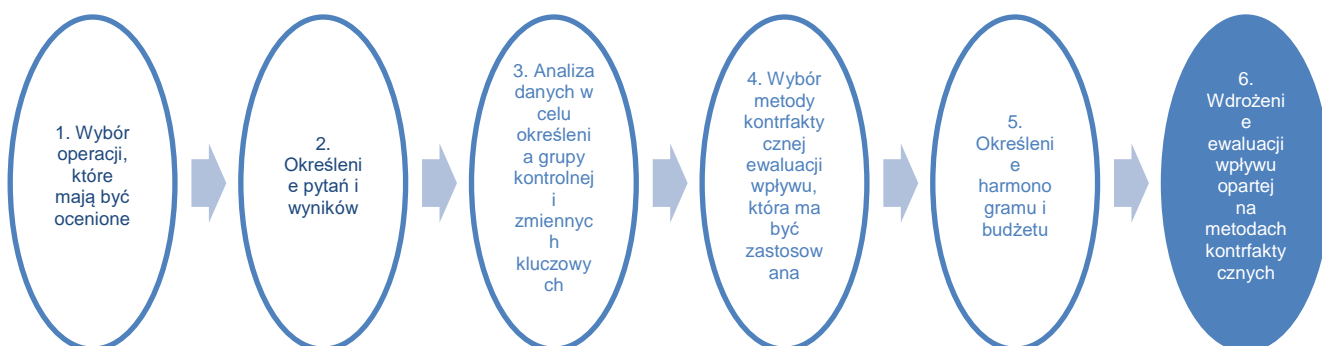
pierwotnych w formie badań reprezentacyjnych, szacowanie skutków w regularnych odstępach czasu stałoby się bardzo kosztowne, chyba że można w sposób wiarygodny zgromadzić dane retrospektywne dotyczące wyników. Nie należy jednak bagatelizować kosztów wyodrębniania danych z wielu systemów administracyjnych i tworzenia jednego analitycznego zbioru danych.

Jak omówiono w sekcji 1.5, sformułowanie teorii zmiany (lub logiki interwencji) może pomóc w ustaleniu, kiedy należy oszacować wpływ.

Rozwiązaniem alternatywnym dla osób planujących ewaluację wpływu opartą na metodach kontrfaktycznych w przypadku braku teorii zmiany (co jest również przydatnym rozwiązaniem dla osób, które mogą oprzeć się na jasnej teorii zmiany) jest przeprowadzenie krótkiego przeglądu wcześniejszych badań oceniających interwencje podobne do przedmiotowej interwencji. Szczegółowa analiza wyników poprzednich badań może dać dobre wyobrażenie o właściwym pomiarze rezultatów i obliczaniu wpływu.

*...lub przegląd
innych
najnowszych
badań*

2.6. Wdrożenie ewaluacji wpływu opartej na metodach kontrfaktycznych



Po określeniu harmonogramu i budżetu planowanie ewaluacji wpływu opartej na metodach kontrfaktycznych jest zakończone i można przejść do jej realizacji. Celem etapu końcowego niniejszego przewodnika jest przedstawienie kluczowych działań związanych z przeprowadzeniem ewaluacji oraz sposobu, w jaki instytucje zarządzające mogą wyeliminować potencjalne przeszkody, które napotykają. W szczególności w niniejszej sekcji omówiono wnioski wyciągnięte z działalności Punktu Informacyjnego ds. Oceny, za pośrednictwem którego KE i DG ds. Zatrudnienia zapewniają wsparcie dla instytucji zarządzających we wszystkich państwach członkowskich. Realizację ewaluacji omówiono ogólnie w „EVALSED: The resource for the evaluation of Socio-Economic Development”⁶⁸; w przewodniku omówiono cztery główne działania, szczególnie z perspektywy ewaluacji wpływu opartej na metodach kontrfaktycznych:

1. wybór ewaluatora,
2. nadzór nad wdrożeniem ewaluacji wpływu opartej na metodach kontrfaktycznych,
3. sprawozdawczość,
4. korzystanie z wyników.

⁶⁸ Dostępne pod adresem: https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf

2.6.1. Wybór ewaluatora

Pierwsza decyzja przy wyborze ewaluatora wiąże się z pytaniem, czy skorzystać z usług kandydata wewnętrznego czy zewnętrznego. W tym ostatnim przypadku wybór ewaluatora odbywa się w drodze postępowania o udzielenie zamówienia.

*Niezależność
ewaluatora*

W art. 44 ust. 3 rozporządzenia w sprawie wspólnych przepisów uznano istnienie kwestii niezależności i stwierdzono, że: „Przeprowadzenie ewaluacji powierza się funkcjonalnie niezależnym ekspertom wewnętrznym lub zewnętrznym”. W związku z tym pierwszym kryterium wyboru ewaluatora jest jego rzeczywista niezależność od instytucji zarządzającej i procesów decyzyjnych w ramach programu EFS+. Drugie kryterium dotyczy kwestii, czy umiejętności i zasoby niezbędne do przeprowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych są wewnątrznie dostępne. Jest to prawdopodobnie najczęstsza przeszkoda w przeprowadzaniu ewaluacji wewnętrznych, ponieważ wiedza techniczna niezbędna do przeprowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych jest bardzo specyficzna i rzadko dostępna w większości organów administracji.

Po dokonaniu wyboru między ewaluatorem wewnętrznym a zewnętrznym administracja musi sporządzić wykaz specyfikacji technicznych ewaluacji; w przypadku wyboru ewaluatora zewnętrznego specyfikacje te będą częścią postępowania o udzielenie zamówienia. W specyfikacji istotnych warunków zamówienia ewaluacji wpływu opartej na metodach kontrfaktycznych należy rozważyć w szczególności następujące aspekty:

*Główne elementy
specyfikacji
istotnych
warunków
zamówienia*

- **cel ewaluacji** oraz pytania badawcze muszą być jasno sformułowane i muszą konsekwentnie prowadzić do ewaluacji wpływu i, w miarę możliwości, do ewaluacji wpływu opartej na metodach kontrfaktycznych;
- należy określić **dostępne dane**; jest to jedna z głównych kwestii wymagająca rozwiązania przy projektowaniu i wdrażaniu ewaluacji wpływu opartej na metodach kontrfaktycznych. Dzięki temu ewaluatorzy uświadamiają sobie istniejące problemy związane z dostępnością danych, co umożliwia im sporządzenie realistycznych i dokładnych ofert;
- **wniosek o przeprowadzenie ewaluacji wpływu opartej na metodach kontrfaktycznych** można wyraźnie określić w specyfikacji istotnych warunków zamówienia, ale wskazanie konkretnych metod, które powinny zostać zastosowane, należy pozostawić w gestii ewaluatora. W ten sposób może on zaproponować oryginalne rozwiązania i zademonstrować swoje zdolności.
- **Proponowany zespół** musi łączyć wiedzę specjalistyczną z kilku dziedzin, w szczególności z zakresu: ewaluacji i ewaluacji wpływu opartej na metodach kontrfaktycznych (wiedza w zakresie projektowania i wdrażania ewaluacji), ekonometrii (stosowanie właściwych metod) oraz dziedziny polityki ocenianej interwencji (rekonstrukcja teorii zmiany i pełna interpretacja ustaleń).
- **Kryteria udzielenia zamówienia** muszą przedkładać jakość nad koszt oferty. Jakość ewaluacji w dużej mierze zależy od zdolności ewaluatora do jej zaprojektowania i przeprowadzenia. Ramy prawne dotyczące zamówień publicznych często nie są idealnie dopasowane do potrzeb ofert dotyczących ewaluacji, ale zleceniodawca powinien w miarę

możliwości dostosować istniejące przepisy, aby wynagradzać jakość i potencjał techniczny.

- W ramach **procesu wyboru** w skład komisji selekcyjnej powinni wchodzić eksperci z dziedziny ewaluacji i ewaluacji wpływu opartej na metodach kontrfaktycznych, aby w pełni została doceniona jakość wniosków i dokonana właściwa ocena różnych rozwiązań w zakresie metodyki i organizacji.

W niektórych państwach członkowskich doświadczenie w zakresie ewaluacji wpływu opartej na metodach kontrfaktycznych jest nadal znikome. Może to stanowić problem w procesie wyboru, ponieważ rynek na tę usługę jest zbyt ograniczony. Prawdopodobnie bardzo niewiele przedsiębiorstw lub ekspertów będzie w stanie spełnić warunki techniczne określone w specyfikacji istotnych warunków zamówienia i przedstawić dobrej jakości ofertę. W takich przypadkach otwarte seminaria instytucji zarządzającej z udziałem pracowników akademickich i przedsiębiorstw, których przedmiotem zainteresowania są ewaluacje, mogą pomóc w przygotowaniu ewaluacji wpływu opartej na metodach kontrfaktycznych, a następnie w sformułowaniu zaproszenia do składania wniosków. Seminaria umożliwiłyby zebranie propozycji od ekspertów krajowych i zapewniłyby im czas na przygotowanie się do wzięcia udziału w przetargu; w celu zachowania przejrzystości i uniknięcia konfliktu interesów seminaria powinny mieć charakter publiczny, a najważniejsze informacje należy udostępniać również tym, którzy nie mogli w nich uczestniczyć.

2.6.2. Nadzór nad ewaluacją wpływu opartą na metodach kontrfaktycznych

Po wyborze ewaluatora i rozpoczęciu ewaluacji wpływu opartej na metodach kontrfaktycznych IZ musi sprawować nadzór nad procesem ewaluacji. Obejmuje to głównie kontrolę i zatwierdzanie wyników pracy ewaluatora, organizowanie spotkań dotyczących postępów z ewaluatorem i zainteresowanymi stronami oraz, w razie potrzeby, zawieranie umów z właścicielami danych.

Instytucja zarządzająca musi być w stanie wypełniać te zadania. Oznacza to również dysponowanie zasobami do celów pełnienia nadzoru nad całym procesem oraz umiejętnościami na potrzeby sprawdzania rozwiązań technicznych ewaluacji wpływu opartej na metodach kontrfaktycznych. Na ogół instytucja zarządzająca przydziela specjalny personel do przeprowadzania ewaluacji, a w przypadku gdy nie ma on wystarczającego doświadczenia związanego z ewaluacją wpływu opartą na metodach kontrfaktycznych, wsparcie powinni zapewnić niezależni eksperci zewnętrzni. Ekspertów tych można zatrudnić na podstawie umów krótkoterminowych lub, jeśli są pracownikami instytucji publicznych, ich usługi będą bezpłatne. Wymagane zaangażowanie będzie prawdopodobnie wynosić od 6 do 12 dni roboczych, co powinno wystarczyć na dokonanie przeglądu sprawozdań i udział w niektórych spotkaniach z instytucją zarządzającą, ewaluatorem i głównymi zainteresowanymi stronami.

*Umiejętności
wewnętrzne i
zewnętrzne w
procesie nadzoru*

Ustanowienie grupy sterującej jest częstym sposobem nadzorowania procesów ewaluacji programu. W skład tej grupy wchodzi zazwyczaj instytucja zarządzająca, urzędnicy reprezentujący inne działy wdrażające, główne zainteresowane strony (partnerzy społeczni, organizacje pozarządowe, pozostałe administracje) oraz niektórzy eksperci ze środowisk

akademickich lub instytucji publicznych (np. krajowego instytutu statystycznego). Grupa sterująca uczestniczy w określaniu pytań ewaluacyjnych, omawia ustalenia poczynione w trakcie ewaluacji, rozpowszechnia je i wykorzystuje w procesach kształtowania polityki. Bardziej elastyczna i wyspecjalizowana podgrupa mogłaby również nadzorować wyniki ewaluacji we współpracy z instytucją zarządzającą.

2.6.3. Sprawozdawczość

Sprawozdawczość jest kluczowym działaniem w ramach ewaluacji i stanowi główny sposób przekazywania rezultatów. Sprawozdania oceniające muszą być jasne, zwięzłe, zrozumiałe dla osób niebędących technikami i muszą zawierać przejrzyste oceny i zalecenia dotyczące polityki. Należy w nich również wykazać wiarygodność ewaluacji określając w tym celu metodykę i wykorzystane dane, jak również wskazując ewentualne ograniczenia dotyczące analizy.

Zasadniczo w przypadku każdej ewaluacji wpływu opartej na metodach konfraktycznych sporządza się trzy sprawozdania:

- **Sprawozdanie wstępne** zwykle przekazuje się tuż po podpisaniu umowy i przedstawia się w nim ostateczną wersję metodyki zgodnie z ustaleniami dokonanymi z instytucją zarządzającą oraz wstępnym przeglądem dostępnych danych.
- **Sprawozdanie okresowe** przekazuje się na pośrednim etapie ewaluacji i w ramach ewaluacji wpływu opartej na metodach konfraktycznych może być poświęcone przedstawieniu zgromadzonych danych, składu grupy objętej interwencją i grupy kontrolnej oraz doboru próby.
- **Sprawozdanie końcowe** zawiera analizy i ustalenia dotyczące ewaluacji wpływu opartej na metodach konfraktycznych. Zawiera ono szczegóły dotyczące metodyki i danych oraz uwypukla się w nim wszelkie implikacje polityczne wynikające ze zgromadzonych dowodów. Struktura sprawozdania końcowego dotyczącego ewaluacji wpływu opartej na metodach konfraktycznych musi zawierać szereg kluczowych sekcji: pytania ewaluacyjne; teorię zmiany (lub logikę interwencji) oraz wskazanie wyników, które należy ocenić; przyjętą metodykę; wykorzystane dane oraz charakterystykę grupy objętej interwencją i grupy kontrolnej; szacowane efekty; odpowiedzi na pytania ewaluacyjne oraz związane z nimi implikacje polityczne.

Odrębne sprawozdania dotyczące monitorowania postępów i korzystanie z wyników

W stosownych przypadkach zleceniodawca dokonuje przeglądu wszystkich sprawozdań przy wsparciu ekspertów zewnętrznych i omawia uwagi z ewaluatorem, który przed formalnym przyjęciem sprawozdania wprowadza wymagane wyjaśnienia i poprawki.

W sprawozdaniu dotyczącym ewaluacji wpływu opartej na metodach konfraktycznych należy unikać nadmiernej ilości szczegółów technicznych, a ustalenia muszą być zrozumiałe i czytelne dla zainteresowanych stron i osób niebędących ekspertami. W specjalnych załącznikach można zawrzeć bardziej techniczne etapy analizy, zapewniając prostą i płynną narrację w głównym sprawozdaniu. Szczegóły metodyczne mają jednak zasadnicze znaczenie na potrzeby wykazania wiarygodności ewaluacji wpływu opartej na metodach konfraktycznych i muszą również być przejrzyste. W sprawozdaniu należy uwzględnić identyfikację zmiennych wynikowych, jakość danych, skład grup objętej interwencją i grupy kontrolnej, liczebność próby, metodę szacowania wpływu oraz zastosowane testy statystyczne. Co więcej, zgodnie

z praktyką akademicką, wykorzystane dane zanonimizowane oraz metody przetwarzania danych mogłyby zostać udostępnione badaczom, którzy zamierzają powtórzyć i zweryfikować analizy.

W sprawozdaniu końcowym dotyczącym wszystkich ewaluacji wpływu opartych na metodach kontrfaktycznych należy przedstawić minimalny zbiór znormalizowanych informacji. Cel minimalnego zbioru informacji jest dwójaki: 1) ułatwienie porównania ustaleń między różnymi programami operacyjnymi i państwami; 2) zgromadzenie informacji, które będą dostępne na potrzeby przyszłych metaanaliz.

Minimalny zestaw informacji, które należy zamieścić w sprawozdaniu końcowym dotyczącym CIE

Wzór stosowany do gromadzenia minimalnego zbioru znormalizowanych informacji na temat każdej ewaluacji dotyczącej EFS jest zgodny z zaleceniami zawartymi w „Projekcie pilotażowym i studium wykonalności [...]”. Z operacyjnego punktu widzenia informacje te mogłyby być podane w załączniku do sprawozdania końcowego i mogłyby również zostać przesłane do KE. W

Tabela 4 poniżej przedstawiono możliwy wzór gromadzenia minimalnego zestawu potrzebnych informacji.

Tabela 4. Podstawowe informacje, które należy zamieścić w dokumentacji przedstawiającej CIE

Rodzaj informacji	Szczegółowe informacje, które należy gromadzić i zgłaszać
1. Informacje dotyczące interwencji	<ul style="list-style-type: none"> - Nazwa ocenianej interwencji - Charakterystyka interwencji (szkolenia, służby zatrudnienia, usługi w zakresie włączenia społecznego itp.) - Podejście dotyczące programu EFS+ (osie priorytetowe, cel szczegółowy, działanie) - Instytucja zarządzająca lub inne organy wdrażające - Wydatki - Czas trwania/intensywność oddziaływania w ramach interwencji - Grupa uczestników według wieku, płci i statusu kwalifikowalności (beneficjenci ubezpieczenia od utraty pracy, długotrwale bezrobotni, młodzież niekształcąca się, niepracująca ani niekształcąca się) - Zakres terytorialny interwencji
2. Pomiar skuteczności interwencji	<ul style="list-style-type: none"> - Wskaźnik dotyczący tego, czy ewaluacja wpływu oparta na metodach kontrfaktycznych szacuje pozytywny i statystycznie istotny wpływ interwencji na odpowiedni wynik lub wyniki - Rzeczywista wielkość tego wpływu - Dwie powyższe miary (+/-) i (wielkość) w odniesieniu do zestawu odpowiednich i porównywalnych wyników: np. wskaźnik zatrudnienia, zarobki itp. - Wskaźniki dotyczące „stosunku wartości do ceny” (jeżeli oszacowano)
3. Dane i metodyka	<ul style="list-style-type: none"> - Źródło danych (badanie, dane administracyjne) - Okres obserwacji - Ekonometryczna metoda kontrfaktyczna ewaluacji wpływu - Horyzont czasowy (ogólnie: krótki ≤ 12 miesięcy od zakończenia interwencji, średni > 12 i ≤ 24 miesiące, efekty długoterminowe > 24 miesiące) - Liczebność próby grupy objętej interwencją i grupy kontrolnej - Dostępność/uwzględnienie danych przed programem

2.6.4. Korzystanie z wyników

Wykorzystanie wyników jest ważne, ponieważ jeżeli nie zostaną one skutecznie rozpowszechnione i nie dotrą do zamierzonych odbiorców, wpływ ewaluacji będzie niewielki.

Rozpowszechnianie ustaleń dotyczących CIE

Rozpowszechnianie ustaleń i wyników ewaluacji zazwyczaj obejmuje:

*ma kluczowe
znaczenie*

- co najmniej jedno pisemne sprawozdanie oceniające, w tym abstrakt i streszczenie;
- co najmniej jedną prezentację ustną ustaleń, uzupełnioną slajdami lub podobnymi narzędziami;
- część techniczną (lub załącznik) sprawozdania zawierającą szczegółowy opis zastosowanej metody, przyjęte kluczowe założenia oraz przyjęte podejście do analiz statystycznych.

Wszystkie sprawozdania oceniające należy opublikować. To postanowienie zawarto w rozporządzeniu w sprawie wspólnych przepisów na okres programowania 2021–2027⁶⁹. Ponadto ustalenia dotyczące ewaluacji należy przedstawić i przedyskutować na forum komitetu monitorującego program; jest to obowiązkowe, ale samo w sobie nie jest wystarczające do promowania pogłębionej debaty, ponieważ na posiedzeniach komitetów monitorujących przeważają na ogół priorytety operacyjne. Przykład krajowej praktyki w zakresie przedstawiania i rozpowszechniania ustaleń dotyczących ewaluacji znajduje się w ramce poniżej.

Nie ma jednej skutecznej polityki komunikacyjnej, która upowszechniłaby i rozpoczęła debatę na temat wyciągniętych wniosków; każdy kontekst wymaga własnej strategii. W związku z tym ważne jest, aby opracować skuteczną strategię, a w szczególności zapewnić, aby zainteresowane strony spoza instytucji zarządzającej zapoznały się z ustaleniami. Dopuszcza się możliwość podzielenia „odbiorców” ewaluacji wpływu opartej na metodach kontrfaktycznych na główne grupy i określenie podstawowych rodzajów komunikacji, na przykład:

*Różni odbiorcy i
wiele podejść do
komunikacji*

- decydenci wysokiego szczebla i przedstawiciele polityczni są przyzwyczajeni do czytania krótkich dokumentów (komunikatów politycznych, streszczeń) zawierających najważniejsze wyniki i zalecenia;
- urzędnicy administracji zaangażowani w realizację interwencji lub podobnych polityk – seminaria lub warsztaty mogą być skutecznym narzędziem służącym do prezentowania rezultatów ewaluacji wpływu opartej na metodach kontrfaktycznych, dostosowując poziom techniczny prezentacji do poziomu interakcji z uczestnikami;
- zainteresowane strony w zakresie polityki (partnerzy społeczni, organizacje pozarządowe, beneficjenci itp.) – podsumowanie sprawozdań i coroczne spotkania na temat ustaleń poczynionych w trakcie ewaluacji mogą być instrumentem łącznym angażującym te podmioty poprzez debatę. W szczególności, jeżeli podmioty te przyczyniły się do określenia pytań ewaluacyjnych, sprawozdanie z rezultatów ewaluacji powinno być powiązane z tymi pytaniami.
- eksperci i naukowcy – konferencje naukowe lub seminaria mogą być najlepiej służyć do przedstawienia rezultatów ewaluacji. Podmioty te zatwierdzają wyniki ewaluacji z naukowego punktu widzenia i w ten sposób wzmacniają ich wiarygodność.

⁶⁹ Art. 44 ust. 7 rozporządzenia w sprawie wspólnych przepisów 2021/1060.

Metaewaluacja⁷⁰ jest potężnym dodatkowym instrumentem służącym rozpowszechnianiu rezultatów ewaluacji wpływu opartej na metodach kontrfaktycznych. Umożliwia ona potwierdzanie i uogólnianie głównych ustaleń, których nie można sformułować w ramach samej ewaluacji wpływu opartej na metodach kontrfaktycznych. Aby przeprowadzić metaewaluację, konieczne jest dysponowanie dużą liczbą wiarygodnych i dobrej jakości ewaluacji wpływu opartych na metodach kontrfaktycznych dotyczących porównywalnych polityk; obecnie metaewaluacje są nadal rzadko przeprowadzane i nie koncentrują się wyłącznie na interwencjach w ramach EFS. Szerokie zastosowanie metaewaluacji wymagałoby bardziej powszechnego korzystania z ewaluacji wpływu opartych na metodach kontrfaktycznych i ich przewidywanego planowania, aby ułatwić ich późniejsze porównywanie. W okresie programowania 2021–2027 EFS+ metaewaluacje można wdrażać w państwach, w których prowadzi się znaczną liczbę programów, lub w ramach porównania między państwami.

Metaewaluacja na potrzeby większego rozpowszechniania i uogólniania

Ramka 13. Doświadczenia Polski związane z Konferencjami Ewaluacyjnymi

Międzynarodowa Konferencja Ewaluacyjna, organizowana przez Ministerstwo Funduszy i Polityki Regionalnej oraz Polską Agencję Rozwoju Przedsiębiorczości, stanowi jeden z kluczowych elementów w obszarze ewaluacji interwencji publicznej w Polsce. Od momentu jej zorganizowania stała się ona uznaną platformą dyskusji pomiędzy zainteresowanymi stronami odpowiedzialnymi za kształtowanie i wdrażanie polityki oraz ocenę skutków.

Od 2005 r. konferencję przez 10 lat organizowano co roku, a od 2015 r. co dwa lata, przy czym w 2021 r. miała miejsce jej 14. edycja. Obecnie konferencję organizuje się jako wydarzenie dwudniowe. Jeden dzień poświęcony jest ustaleniom wynikającym z badań i efektem interwencji; w tym dniu omawia się kwestie polityczne. Drugi dzień koncentruje się na kwestiach metodologicznych i ewaluacji jako procesie.

W trakcie 14. edycji konferencji metody kontrfaktyczne były wielokrotnie przedmiotem dyskusji panelowych lub wystąpień ekspertów, a w ciągu tych dwu dni omówiono je z różnych perspektyw. Na przykład w 2017 r. szacowane efekty zaprezentowano na sesjach dotyczących wpływu interwencji, natomiast dyskusje na temat zalet i wad technik szacowania odbyły się osobno.

Na uwagę zasługuje również fakt, że konferencja ma zróżnicowane grono odbiorców, ponieważ skierowana jest nie tylko do badaczy czy specjalistów z zakresu metodologii, ale również do osób zajmujących się projektowaniem i wdrażaniem interwencji. Cel wydarzenia jest dwójaki: budowanie potencjału systemu ewaluacji dzięki rozwijaniu umiejętności i wiedzy jego członków, a także tworzenie sieci kontaktów dzięki zapewnieniu zainteresowanym stronom przestrzeni do wymiany doświadczeń pomiędzy partnerami krajowymi, regionalnymi i zagranicznymi. W rzeczywistości każda edycja skupia około 300–400 przedstawicieli różnych sektorów – administracji publicznej, środowiska akademickiego, przedsiębiorstw świadczących usługi doradztwa i organizacji pozarządowych. Jest to znaczna liczba, biorąc pod uwagę fakt, że w Polsce podejście do polityki oparte na dowodach nie jest jeszcze w pełni zakorzenione.

Międzynarodowa Konferencja Ewaluacyjna stanowi część systemu ewaluacji dotyczącej administracji publicznej. Znaczne zwiększenie potencjału było możliwe dzięki szkoleniom prowadzonym przez Wspólne Centrum Badawcze, Bank Światowy lub dzięki finansowaniu w ramach pomocy technicznej w programach operacyjnych. Szkolenia z zakresu ewaluacji, w tym z zakresu podejścia kontrfaktycznego, dostępne, były nie tylko dla pracowników w zespołach ewaluacyjnych instytucji zarządzających, ale także dla instytucji takich jak lokalne czy regionalne urzędy pracy. Konferencja stanowi okazję do spotkań, omówienia i podzielenia się aktualnymi osiągnięciami i ustaleniami⁷¹.

⁷⁰ Jest to systematyczne porównanie, nawet przy pomocy narzędzi statystycznych, większej liczby ewaluacji wpływu opartych na metodach kontrfaktycznych podobnych interwencji. Metaewaluacja dostarcza wskazówek na temat skuteczności danego rodzaju interwencji dzięki badaniu skutków podobnych działań w wielu różnych kontekstach. W przypadku metaewaluacji nie można jednak zagłębiać się w szczegóły dotyczące każdej pojedynczej interwencji wykorzystanej do porównania, a ich implikacje dla polityki odnoszą się do ogólnych kierunków polityki, a nie do pojedynczych szczegółów dotyczących każdej interwencji.

⁷¹ Informacje na temat konferencji są dostępne pod adresem: <https://www.ewaluacja.gov.pl/strony/xiv-miedzynarodowa->

Rozdział 3. Jak wybrać odpowiednią metodę do przeprowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych

W niniejszym rozdziale przedstawiono metody empiryczne stosowane w ewaluacji wpływu opartej na metodach kontrfaktycznych. W szczególności omówiono podejście eksperymentalne (badanie randomizowane z grupą kontrolną), a także najczęściej stosowane metody quasi-eksperymentalne: parowanie (metoda *propensity score*), różnica w różnicach, model regresji nieciągłej i zmienne instrumentalne. Każda metoda charakteryzuje się specyficznym sposobem tworzenia grupy kontrolnej, aby odpowiedzieć na pytanie kontrfaktyczne: „Co by się stało z grupą objętą interwencją pod względem wyników, gdyby nie uczestniczyła w interwencji/nie była objęta interwencją?”. W praktyce wybrany projekt można dostosować do danego kontekstu ustalonego na podstawie rodzaju interwencji i danych, które są dostępne lub mogą być zgromadzone⁷².

Z powodu bardzo zróżnicowanych okoliczności, z jakimi mają do czynienia instytucje zarządzające, dostarczenie szczegółowych wytycznych dotyczących wyboru najwłaściwszego projektu oceny nie jest możliwe. Wybierając najodpowiedniejsze podejście do ewaluacji wpływu opartej na metodach kontrfaktycznych w konkretnych okolicznościach, instytucje zarządzające powinny rozważyć, co sprawdziło się w poprzednich ewaluacjach, zarówno w ramach samej instytucji zarządzającej, jak i w państwach członkowskich – instytucje zarządzające mogą wyciągnąć wnioski na podstawie wcześniejszych osiągnięć w ramach ich programu oraz obszarów, w których wystąpiły podobne okoliczności. W tym względzie przydatnym źródłem informacji mogą być fora wymiany doświadczeń zdobytych przy opracowywaniu ewaluacji i jej wdrażaniu. Ważnym źródłem informacji pomocnym w procesie projektowania może być także wyszukiwanie w literaturze ewaluacji podobnych interwencji. Eksperti, którzy otrzymali zlecenie od instytucji zarządzającej, będą również mieli sugestie dotyczące najlepszych podejść do projektu ewaluacji. Należy pamiętać, że instytucja zarządzająca może posiadać rozległą wiedzę ekspercką i doświadczenie, z których można skorzystać.

Wybór odpowiedniego podejścia

3.1. Randomizacja – podejście eksperymentalne

Najważniejsze cechy randomizacji

- Osoby kwalifikujące się do uczestnictwa są losowo przydzielane do grupy objętej interwencją lub grupy kontrolnej.
- Randomizacja zapewnia, aby obie grupy były identyczne (w ujęciu uśrednionym) pod względem wszystkich istotnych cech.

⁷² W odniesieniu do innych technicznych aspektów metodologicznych zob. również [Komisja Europejska, 2019](#).

- W związku z tym grupa kontrolna stanowi odpowiedź na pytanie kontrfaktyczne, a różnica w wynikach między grupą objętą interwencją a grupą kontrolną stanowi efekt przyczynowy interwencji.

Projekty ewaluacji opartej na randomizacji mogą przybierać różne formy. W tym przypadku skupiono się na prostym podejściu dwugrupowym – jedna grupa objęta interwencją i jedna grupa kontrolna – w celu uściślenia kluczowych zasad. Na wykresie 1 przedstawiono prosty projekt randomizowany.

Kluczową kwestią jest to, że randomizacja zapewnia, aby dwie grupy były statystycznie równoważne we wszystkich aspektach w momencie, w którym poddaje się je randomizacji. Następnie grupa objęta interwencją zostaje objęta interwencją, która stanowi przedmiotem ewaluacji i której wpływ lub skutki należy zmierzyć.

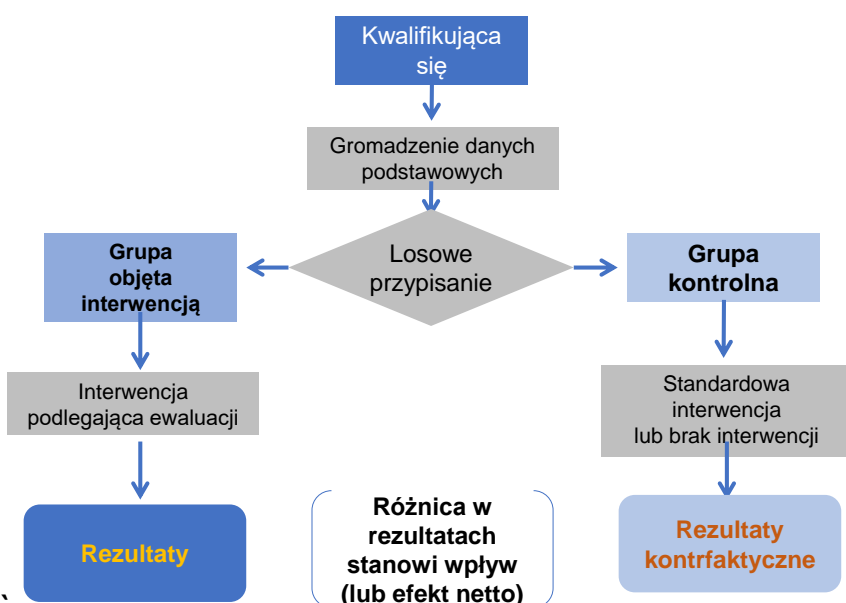
*Grupy
statystycznie
równoważne*

W zależności od kwestii politycznej o zasadniczym znaczeniu, grupa kontrolna może być wyznaczona jako w ogóle nie objęta żadną interwencją albo grupa objęta interwencją może zostać porównana z grupą objętą inną interwencją (można ją postrzegać jako reprezentującą standardową interwencję), bądź też obok grupy kontrolnej może być wiele grup objętych interwencją. Na przykład przedmiotem zainteresowania może być porównanie wyników programu szkoleniowego finansowanego w ramach EFS z innymi szkoleniami finansowanymi z funduszy krajowych lub ze świadczeniem innych usług dla tych samych grup.

*Brak interwencji lub
inna interwencja
dla grupy
kontrolnej*

Ponieważ grupa objęta interwencją i grupa kontrolna są statystycznie równoważne w momencie, w którym poddaje się je randomizacji, a objęcie kolejnymi programami jest kontrolowane, różnice w wynikach można przypisać interwencji poddawanej ewaluacji (z zastrzeżeniem standardowej niepewności statystyczne), a alternatywne wyjaśnienia wykluczyć jako przyczyny wszelkich zaobserwowanych różnic (zob. przykład w ramce poniżej).

Wykres 6. Dwugrupowa kontrola randomizowana



Ze względu na swoje nieodłączne cechy projektowe, projekty randomizowane, o ile są prawidłowo realizowane, oferują możliwość uzyskania mocnych dowodów na efekty interwencji. Z tego powodu są one bardzo cenione. Wymagają one jednak wczesnego i szczegółowego planowania, a także mogą być skomplikowane pod względem projektowania i zarządzania. Ponadto zarządzający programami napotykają poważne trudności w ich prawidłowym wdrażaniu. Na przykład sam proces randomizacji może spowodować zmianę struktury osób uczestniczących w interwencji, tzn. niektórych potencjalnych uczestników może zniechęcać idea randomizacji i mogą odmówić udziału w interwencji. Ponadto osoby objęte randomizacją nie zawsze mogą pogodzić się ze statusem przypisania, a także istnieje szereg innych wyzwań, z którymi należy się zmierzyć. W niektórych okolicznościach realizacja schematów badań randomizowanych z grupą kontrolną może być kosztowna.

Mocne dowody...

Z tych i innych przyczyn wydaje się mało prawdopodobne, aby oceny instrumentów i interwencji finansowanych z EFS były przeprowadzane z zastosowaniem podejścia opartego na randomizacji. W niniejszym przewodniku przestrzegamy jednak przed wykluczeniem randomizacji we wszystkich przypadkach bez odpowiedniego rozważenia. Podejście to było szeroko stosowane, a przykłady, poza tymi z Wielkiej Brytanii omówionymi w poniższej ramce, obejmują również eksperymenty GAIN ze Stanów Zjednoczonych⁷³ (istnieje wiele innych przykładów z Ameryki Północnej), eksperymenty przeprowadzone w Szwecji⁷⁴ w ramach interwencji w ramach EFS dla okresu 2014–2020 (zob. następna ramka), a także badanie przeprowadzone w Niemczech w celu oceny skutków aktywnych usług rynku pracy świadczonych przez prywatnych usługodawców w porównaniu z usługami świadczonymi przez publiczne służby zatrudnienia⁷⁵ i wiele innych.

... ale trudne do zaprojektowania

Ramka 14. Przykład badania randomizowanego projektu EFS dotyczącego młodzieży

Ewaluacja projektu „Ung framtid” („Młoda przyszłość”) w Szwecji

Szwedzki projekt „Ung framtid” („Młoda przyszłość”) to projekt finansowany z EFS, który realizowano w latach 2015–2018. Miał on na celu wspieranie młodych osób w wieku 18–24 lat poszukujących pracy, które wchodzili na rynek pracy w Środkowej Norrlandii, Północno-Środkowej Szwecji i Południowej Szwecji. W ramach tego projektu zintensyfikowano i zindywidualizowano działania dotyczące wsparcia na rzecz młodych osób poszukujących pracy; został on zrealizowany przez szwedzkie publiczne służby zatrudnienia; obejmował indywidualne planowanie, informacje, doradztwo i konkretne wsparcie w zakresie dopasowywania, szkoleń i innych możliwych działań. Ogółem projekt objął swoim zasięgiem prawie 17 000 młodych ludzi.

Ewaluacja została zlecona organizacji Arbetsförmedlingen przez szwedzką Radę EFS jako część ram ewaluacji opracowanych w ramach projektu EFS na lata 2014–2020 mającego na celu poprawę zdolności ewaluacyjnych publicznych służb zatrudnienia („Evidence-based EU fund Project 2014–2020”; zob. ramka w ostatnim rozdziale przewodnika).

W eksperymencie przeprowadzonym w okresie od czerwca 2017 r. do stycznia 2018 r. obejmującym osiem lokalnych urzędów pracy spośród 90 biorących udział w projekcie, zakwalifikowane osoby młode zostały losowo przydzielone do grupy objętej interwencją (osoby otrzymały zintensyfikowane wsparcie) i grupy kontrolnej (osoby otrzymały standardowe wsparcie)⁷⁶. W eksperymencie 4 689 młodych ludzi zostało losowo przydzielonych do grupy objętej interwencją i grupy kontrolnej, przy czym 2 972 z nich przydzielono

⁷³ Zob. [Riccio J., Friedlander, D., Freedman S., 1994.](#)

⁷⁴ Zob. [Hagglund, P., 2006.](#)

⁷⁵ Zob. [Krug G., Stephan G., 2011.](#)

⁷⁶ Etap pilotażowy został zrealizowany w okresie od kwietnia do maja 2017 r. w celu przetestowania szeregu działań, takich jak zarządzanie randomizacją, zaangażowanie młodych uczestników, sprawozdawczość.

do grupy kontrolnej. W wyniku procesu losowego przydziału utworzono grupę objętą interwencją i grupę kontrolną, które były bardzo podobne w momencie przydziału. W rezultacie wszelkie różnice między dwiema grupami w zakresie zmiennych wynikowych (wyjście z bezrobocia, odsetek bezrobotnych i średnia liczba dni pozostawania bez pracy) mierzonych po objęciu interwencją w ramach EFS można przypisać działaniom w ramach „Młodej przyszłości”.

Ustalenia wynikające z badania pokazują, że projekt „Młoda przyszłość” miał pozytywny wpływ na kobiety, ale nie na mężczyzn. W przypadku kobiet pozytywne efekty stwierdzono we wszystkich rozważanych zmiennych wynikowych: liczba kobiet poddanych interwencji, które wyszły z bezrobocia, była wyższa o około 7 punktów procentowych w ciągu pierwszych dwóch miesięcy od przystąpienia do projektu, odsetek bezrobotnych zmniejszył się bardziej w grupie poddanej interwencji niż w grupie kontrolnej, a ponadto kobiety poddane interwencji pozostawały bez pracy przez mniejszą liczbę dni niż kobiety należące do grupy kontrolnej⁷⁷.

Projekty randomizowane można odróżnić od innych podejść głównie ze względu na mocny nacisk na kontrolowanie wystąpienia potencjalnego błędu systematycznego między grupą objętą interwencją a grupą kontrolną w ramach projektu badania. Ten nacisk na projektowanie sprawia, że podejście jest bardzo intuicyjne, ale wymaga wcześniejszego planowania. Projekty randomizowane są często najskuteczniej realizowane w ramach ewaluacji nowych interwencji pilotażowych, a nie tych istniejących. Wynika to stąd, że wymagają one pewnego stopnia kontroli nad sposobem, w jaki uczestnicy są rekrutowani do ewaluowanej interwencji. Taka „kontrola” jest często trudniejsza do osiągnięcia w istniejących programach niż w nowych interwencjach, które charakteryzują się otwartością na nowe pomysły.

Randomizacja w ramach projektu badania

Wdrożenie projektu randomizowanego z grupą kontrolną oznacza, że części kwalifikującej się populacji docelowej nie oferuje się udziału w projekcie, ale przydziela się ją do grupy kontrolnej. Przydział ten jest całkowicie losowy i nie dokonuje się go na życzenie wnioskodawcy ani administratorów interwencji. Z tego powodu decydenci mogą sprzeciwiać się badaniu randomizowanemu z grupą kontrolną ze względów etycznych, zanim rozważą, czy są one wykonalne z praktycznego i analitycznego punktu widzenia.

Zastrzeżenia natury etycznej

Istnieją jednak mocne argumenty przemawiające za stosowaniem projektów randomizowanych. Jeżeli randomizacja zapewnia najlepszą jakość, najbardziej wiarygodne dowody na skuteczność interwencji finansowanych ze środków publicznych, ważne jest, aby były one szerzej stosowane w ocenie wpływu interwencji. Ponadto, jeśli wpływ danej interwencji jest nieznaną z wyprzedzeniem, wykluczenie jednostek nie jest nieetyczne, ponieważ nie można założyć, że jednostki te odniosłyby korzyści. Co więcej, takie podejścia są szeroko stosowane w medycynie i innych dziedzinach nauki, takich jak badania edukacyjne. Ponadto w niektórych sytuacjach, w których usługi i wsparcie oferowane w ramach interwencji są nadsubskrybowane (tj. jest więcej osób kwalifikujących się niż osób, które faktycznie mogą być beneficjentami interwencji), przydzielanie osób do interwencji losowo z puli osób, które się kwalifikują, może być najbardziej etycznym sposobem przydzielenia ograniczonych zasobów.

...ale i silne argumenty przemawiające za stosowaniem projektów randomizowanych

⁷⁷ Axdorph E, Egebark J., Lundström T., Özcan G., 2019.

3.2. Projekty nierandomizowane lub quasi-eksperymentalne

Jeśli randomizacja jest niewykonalna, w zestawie narzędzi ewaluacji wpływu opartej na metodach kontryfaktycznych występuje szereg alternatywnych metod, które wykorzystują różne podejścia do tworzenia grupy kontrolnej w celu uzyskania odpowiedzi na pytanie kontryfakcyjne. W niniejszej podsekcji nakreślono najpierw główne wyzwanie dotyczące metod quasi-eksperymentalnych, a następnie przedstawiono cztery podejścia, które są powszechnie stosowane w praktyce empirycznej dotyczącej ewaluacji wpływu opartej na metodach kontryfaktycznych.

Brak możliwości randomizacji

3.2.1. Grupa docelowa i grupa kontrolna bez randomizacji

W projektach quasi-eksperymentalnych grupy docelowe objęte interwencją są porównywane z grupą kontrolną⁷⁸, do której należą nielosowo przydzielone grupy docelowe lub potencjalne grupy docelowe, które nie są objęte interwencją. Podobnie jak w przypadku eksperymentu, celem jest uzyskanie nieobciążonego błędem systematycznym oszacowania zmiany wyników spowodowanej przedmiotową interwencją. Ponieważ grupa objęta interwencją i grupa kontrolna nie są tworzone losowo, projekty quasi-eksperymentalne wymagają poświęcenia znacznie większej uwagi metodom uwzględniania potencjalnych różnic pomiędzy członkami grupy objętej interwencją i potencjalnymi grupami kontrolnymi, które prawdopodobnie wpłyną na decyzję o uczestnictwie, a tym samym na wyniki. W tym przypadku wybór wiarygodnej grupy kontrolnej ma zasadnicze znaczenie. Niewybranie odpowiedniej grupy kontrolnej i uwzględnienie w analizie pozostałych różnic pomiędzy obiema grupami osłabia wiarygodność szacunków i może utrudniać dążenie do wykluczenia alternatywnych wyjaśnień w odniesieniu do wszelkich zaobserwowanych efektów.

W przypadku interwencji współfinansowanych z EFS najczęściej stosowanym quasi-eksperymentalnym projektem ewaluacji jest projekt z dwiema grupami, projekt odniesienia/projekt działań następczych. Takie projekty charakteryzują się określeniem grupy kontrolnej i grupy objętej interwencją, podobnie jak w przypadku randomizacji, z tym wyjątkiem, że grupa kontrolna zostaje utworzona (bez użycia randomizacji) z istniejących grup nieuczestniczących w programie, przy założeniu, że ma ona być jak najbardziej podobna do grupy poddanej interwencji.

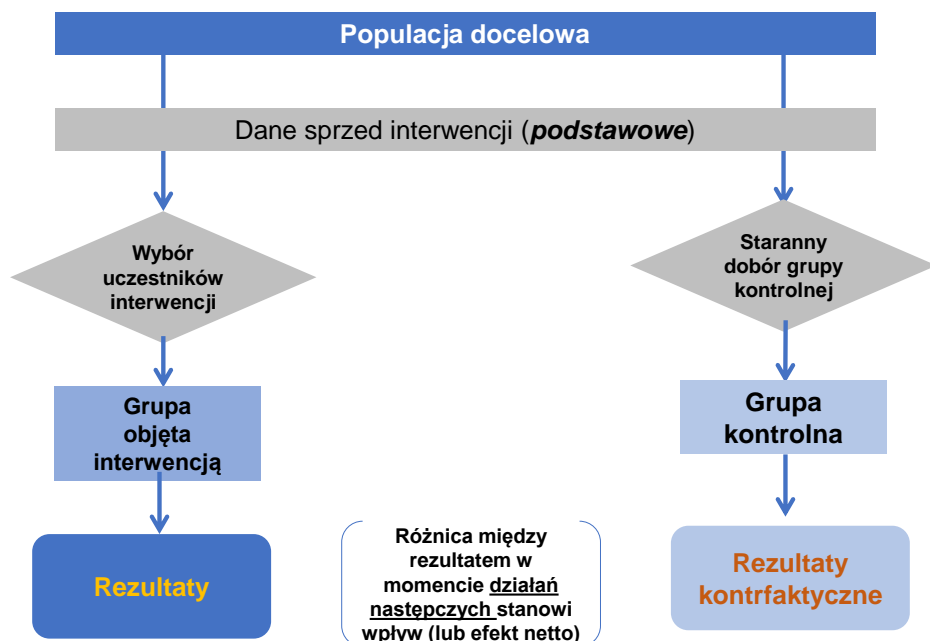
Grupa kontrolna i grupa objęta interwencją muszą być do siebie podobne

Dobłą strategią znalezienia wiarygodnej grupy kontrolnej w warunkach quasi-eksperymentalnych jest wybranie tych osób z grup kontrolnych, które zostały wykluczone z udziału w interwencji na podstawie czynników niezwiązanych z ich charakterystyką i potencjalnymi wynikami. W niektórych okolicznościach może istnieć powód, aby twierdzić, że chociaż grupy kontrolne nie zostały utworzone w sposób jednoznacznie losowy, to można dokonać identyfikacji *ex post* osób fizycznych lub przedsiębiorstw, w przypadku których brak objęcia interwencją okazał się losowy w odniesieniu do potencjalnych rezultatów. Jeśli takie okoliczności wystąpią, będą one zbliżone do ideału w kontekście podejścia quasi-eksperymentalnego. Na

⁷⁸ Ścisłe rzecz ujmując, „grupa kontrolna” jest terminem używanym w projektach eksperymentalnych, a w projektach quasi-eksperymentalnych zwykle mówi się o „grupach porównawczych”. Jednak w praktyce terminy te często stosuje się zamiennie. W niniejszym przewodniku stosuje się zatem termin „grupa kontrolna” w odniesieniu do grupy wykorzystywanej do oszacowania scenariusza kontryfaktycznego, niezależnie od projektu ewaluacji wpływu opartej na metodach kontryfaktycznych.

przykład niektórzy członkowie grupy docelowej interwencji mogą zostać wykluczeni z udziału w niej w wyniku niedopatrzenia lub błędu administracyjnego. Zrozumienie procesu wyboru jednostek do interwencji jest zatem niezwykle ważne dla określenia wiarygodnej grupy kontrolnej – nie da się tego bardziej podkreślić.

Wykres 7. Stylizowany projekt quasi-eksperymentalny z grupą objętą interwencją i grupą kontrolną



Wiarygodną grupę kontrolną można stworzyć na kilka sposobów. Po pierwsze, można zastosować podejście oparte na dopasowaniu statystycznym: tj. dane są gromadzone zarówno od osób poddanych interwencji, jak i od (zazwyczaj bardzo licznej) próby osób niepoddanych interwencji. Następnie grupę kontrolną tworzy się z grupy osób niepoddanych interwencji, wybierając osoby, które są najbardziej podobne do osób należących do grupy objętej interwencją. „Podobieństwo” dotyczy zbioru cech społeczno-demograficznych – np. wieku, płci, wykształcenia, zatrudnienia, bezrobocia itp. – mierzonych zanim grupa objęta interwencją dołączyła do programu. W rezultacie osoby niepoddane interwencji są w rzeczywistości „parowane” z osobami poddanymi interwencji. W praktyce potencjalnie długą listę cech społeczno-demograficznych można podsumować przy użyciu wartości „propensity score”, co ułatwia stosowanie.

Parowanie osób poddanych interwencji z osobami niepoddanymi interwencji

3.2.2. Propensity score matching

Najważniejsze cechy metody propensity score matching

- Dopasowanie imituje randomizację poprzez skonstruowanie *ex post* grupy kontrolnej, która jest jak najbardziej podobna do grupy objętej interwencją pod względem wszystkich istotnych cech.
- W przeciwieństwie do randomizacji dopasować można wyłącznie cechy możliwe do zaobserwowania (wiek, płeć, wykształcenie itp.), natomiast cech niemożliwych do zaobserwowania (np. motywacja) nie można brać pod uwagę.

- Zasadność tego podejścia w dużym stopniu zależy od dostępności danych.

Propensity score matching (PSM) polega na oszacowaniu modelu statystycznego dla całej próby (grupy objętej interwencją i potencjalnej grupy kontrolnej), co pozwala oszacować skłonność do uczestnictwa każdej osoby lub przedsiębiorstwa – niezależnie od tego, czy rzeczywiście biorą udział czy nie⁷⁹. Osoby lub przedsiębiorstwa poddane interwencji są następnie parowane – z jedną osobą lub przedsiębiorstwem niepoddanymi interwencji albo z kilkoma osobami lub przedsiębiorstwami niepoddanymi interwencji – na podstawie wartości *propensity score*⁸⁰. Procedura ta umożliwia identyfikację grupy kontrolnej, którą można następnie wykorzystać do oszacowania scenariusza kontrfaktycznego. Dopasowanie w ten sposób zapewnia, aby oszacowania wpływu uwzględniały możliwe do zaobserwowania różnice między grupą poddaną interwencji a grupą kontrolną oraz – zakładając, że wszystkie istotne różnice występujące przed dopasowaniem są możliwe do zaobserwowania – zapewnia możliwość uzyskania oszacowania skutków interwencji nieobarczonego błędem systematycznym. Jeżeli jednak wybór jednostek do objęcia interwencją odbywa się na podstawie niezaobserwowanych czynników, pozostanie wątpliwość co do adekwatności dopasowania pod względem jego zdolności do wyeliminowania błędu systematycznego. Niezwykle ważnym założeniem leżącym u podstaw podejścia opartego na parowaniu jest to, że cechą charakterystyczną procesu wyboru są możliwe do zaobserwowania dane.

Skłonność do uczestnictwa jako sposób na zdefiniowanie grupy poddanej interwencji i grupy kontrolnej

Poniższy wykres w sposób intuicyjny i uproszczony przedstawia podejście *propensity score matching*. Na osi Y przedstawiono liczbę osób w grupie poddanej interwencji i w grupie niepoddanej interwencji, uporządkowanych według wartości *propensity score* na osi X. Zazwyczaj osoby poddane interwencji charakteryzują się stosunkowo wyższymi wartościami *propensity score*, natomiast osoby niepoddane interwencji – niższymi wartościami *propensity score*. Obszar, na którym wartości *propensity score* tych dwóch grup pokrywają się, jest znany jako obszar wspólnego przedziału określoności (ang. *common support*).⁸¹ Na tym obszarze jednostki poddane interwencji są parowane z jednostkami niepoddanymi interwencji. Na wykresie przedstawiono dwa przykłady, jednak proces zasadniczo powtarza się do momentu, gdy na obszarze wspólnego przedziału określoności każda jednostka poddana interwencji zostanie przypisana do jednostki niepoddanej interwencji. Na wykresie parowanie to przeprowadzono, korzystając z techniki parowania z „najbliższym sąsiadem” (ang. *nearest neighbour*). „Najbliższym sąsiadem” każdego członka grupy objętej interwencją jest członek grupy kontrolnej o wartości *propensity score* najbliższej jego własnej wartości. Po utworzeniu dwóch grup można porównać ich średnie wyniki w celu oszacowania wpływu. W praktyce przeprowadzenie procesu *propensity score matching* może stać się bardzo skomplikowane i wymagać rozważenia szeregu kwestii. Wiele z tych kwestii zostało tutaj pominiętych, aby zapewnić zrozumiałość kluczowych zasad. W następnym ramce przedstawiono

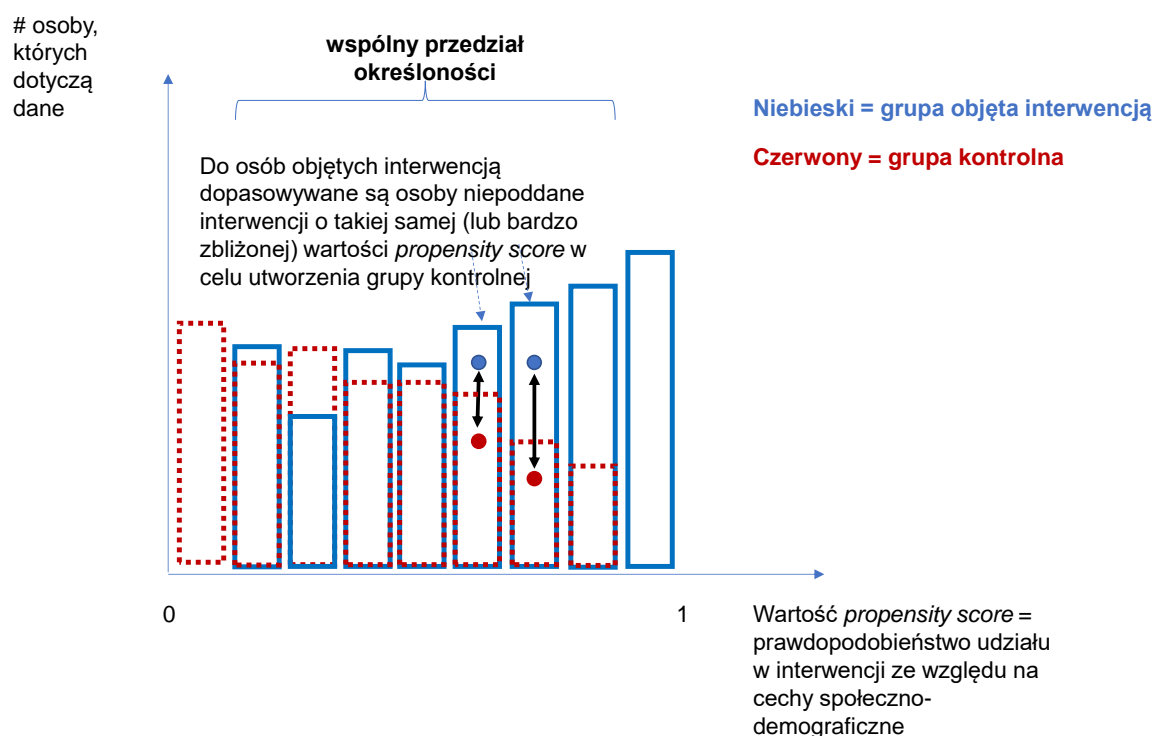
⁷⁹ W celu uproszczenia tych rozważań zakłada się, że decydenci chcą wiedzieć, jak interwencja wpływa na tych, którzy faktycznie są beneficjentami usług świadczonych w ramach programu (w wielu przypadkach jest to podzbiór grupy docelowej, której zaoferowano taką możliwość). Nazywa się to analizą „oddziaływania na jednostki poddane interwencji”.

⁸⁰ Istnieje szeroki zakres potencjalnych podejść do parowania na podstawie wartości *propensity score*. Przegląd w przystępnej formie można znaleźć tutaj: [Caliendo M., Kopeinig S., 2008](#)

⁸¹ Zakres obszaru wspólnego przedziału określoności ma wpływ na liczebność próby i przydatność wyników dla kształtowania polityki, w szczególności w przypadku, gdy duża liczba jednostek poddanych interwencji znajduje się poza obszarem wspólnego przedziału określoności.

praktyczny przykład, w którym w ewaluacji EFS zastosowano podejście polegające na dopasowywaniu.

Wykres 8. Ilustracja podejścia opartego na wartościach *propensity score*



Wiarygodność podejścia opartego na wartościach *propensity score* opiera się między innymi na założeniu, że możliwe do zaobserwowania dane w pełni charakteryzują wybór do objęcia interwencją. Innymi słowy, zakłada się, że między grupą objętą interwencją a grupą kontrolną nie ma żadnych niezabobserwowanych różnic związanych z wynikami lub decyzją o udziale w interwencji. Wiarygodność tego założenia jest tym większa, im więcej zmiennych uwzględnia się w oszacowaniu wartości *propensity score*, przy czym wyboru zmiennych dokonuje się na podstawie wcześniejszej wiedzy i teorii. W szczególności w kontekście interwencji na rynku pracy przy sprawdzaniu potencjalnych niezabobserwowanych różnic zdecydowanie zaleca się włączenie dotychczasowych doświadczeń poszczególnych osób na rynku pracy⁸².

Wybór na podstawie możliwych do zaobserwowania danych

Ramka 15. Przykład ewaluacji, w której zastosowano podejście oparte na parowaniu⁸³

Wpływ interwencji w ramach EFS finansowanych w latach 2014–2020 na osoby długotrwale bezrobotne w regionie Marche⁸⁴

Na potrzeby ewaluacji wpływu szeregu interwencji w ramach EFS (staży, praktyk zawodowych, krótkich praktyk zawodowych w gminach i bonów na szkolenia) sfinansowanych we włoskim regionie Marche na osoby długotrwale bezrobotne zastosowano podejście oparte na parowaniu. Interwencje nie były specjalnie ukierunkowane na osoby długotrwale bezrobotne, jednak stanowiły one większość uczestników wszystkich przeanalizowanych interwencji w ramach EFS.

⁸² Zob. [Caliendo, Mahlstedt i Mitnik, 2017](#) oraz [Kluve, Lehmann i Schmidt, 2008](#).

⁸³ Inne przykłady praktycznego zastosowania technik (PSM, modelu regresji nieciągłej i różnicy w różnicach) można znaleźć tutaj: [Komisja Europejska, 2020](#).

⁸⁴ Zob. [Pompili M., Giorgetti I., 2020a](#)

Celem ewaluacji było określenie wpływu środków wprowadzanych w ramach EFS na prawdopodobieństwo zatrudnienia po 6, 9, 12, 15 i 18 miesiącach od rozpoczęcia interwencji. Zastosowano kilka zmiennych wynikowych: prawdopodobieństwo znalezienia zatrudnienia w określonym czasie po interwencjach, prawdopodobieństwo znalezienia zatrudnienia na podstawie umowy na czas nieokreślony oraz liczbę dni przepracowanych w okresie po interwencjach.

W celu zidentyfikowania różnic między grupą poddaną interwencji a grupą niepoddaną interwencji przyjęto podejście oparte na wartościach *propensity score* (parowanie z najbliższym sąsiadem), a cztery interwencje w ramach EFS wymienione powyżej przeanalizowano oddzielnie. W przypadku każdej interwencji grupa poddana interwencji składała się z uczestników, którzy rozpoczęli udział w niej przed końcem sierpnia 2019 r. (526 w przypadku staży, 1 058 w przypadku praktyk zawodowych, 236 w przypadku krótkich praktyk zawodowych w gminach i 241 w przypadku bonów na szkolenia); w grupie kontrolnej znalazły się osoby bezrobotne zarejestrowane przez publiczne służby zatrudnienia w latach 2016–2018, których okres bezrobocia wyniósł co najmniej 12 miesięcy (77 255 przypadków).

Przy obliczaniu PSM wzięto pod uwagę szereg zmiennych należących do cech społeczno-demograficznych osób poddanych i niepoddanych interwencji, takie jak: płeć, wiek, obywatelstwo, wykształcenie, miejsce zamieszkania, data rozpoczęcia udziału w interwencji oraz historia zatrudnienia do 36 miesięcy przed interwencją.

Do pomiaru zmiennych wynikowych dotyczących grupy poddanej interwencji i grupy kontrolnej (przed interwencjami i po interwencjach) wykorzystano dane z COB, archiwum obowiązkowych powiadomień przedsiębiorstw o umowach o pracę przesyłanych do publicznych służb zatrudnienia.

Wyniki badania były niejednoznaczne, przy czym pozytywny wpływ odnotowano w przypadku staży i praktyk zawodowych, a negatywny – w przypadku krótkich praktyk zawodowych w gminach, natomiast brak znaczącego wpływu – w przypadku bonów na szkolenia⁸⁵.

3.2.3. Różnica w różnicach

Najważniejsze cechy różnicy w różnicach

- Różnica w różnicach to intuicyjne podejście, w ramach którego porównuje się różnicę w wynikach przed interwencją i po niej w grupie objętej interwencją z tymi w grupie kontrolnej.
- Ponieważ zmiana w czasie w grupie kontrolnej odzwierciedla, co stałoby się z grupą objętą interwencją w przypadku braku interwencji (scenariusz konfaktyczny), każda dodatkowa różnica w wynikach przed interwencją lub po niej w grupie objętej interwencją odzwierciedla efekt przyczynowy interwencji.
- Jest to prosta metoda, którą można zastosować w wielu sytuacjach.

Czy to oddzielnie, czy też w połączeniu z parowaniem, podstawowe (lub dokonane przed interwencją) pomiary zmiennych wynikowych można wykorzystać do przeprowadzenia oszacowań różnic w różnicach. W tym przypadku różnicę w wynikach przed interwencją i po jej zakończeniu w grupie kontrolnej odejmuje się od tej samej różnicy zaobserwowanej w grupie poddanej interwencji, aby uzyskać szacunkowy wpływ tej interwencji. Również w tym przypadku wybór wiarygodnej grupy kontrolnej ma zasadnicze znaczenie. Wpływ obliczony na podstawie różnicy w różnicach można zwykle zaobserwować w ramach regresji, która uwzględnia również inne zaobserwowane różnice między grupą objętą interwencją a grupą kontrolną. Ponadto podejście to uwzględnia niezauważalne różnice pomiędzy tymi dwiema grupami, które są niezmiennie w czasie, jak również te, które

Przed interwencją i po jej zakończeniu

⁸⁵ W przypadku bonów na szkolenia – jak podkreślono w sprawozdaniu oceniającym – ustalenia można uznać jedynie za wstępne, ponieważ przebadana próba jest zbyt mała w stosunku do całkowitej liczby uczestników.

zmieniają się w czasie, ale wpływają w równym stopniu na grupę kontrolną i grupę objętą interwencją (na przykład czynniki wpływające na całą gospodarkę). Dzięki tej zdolności do uwzględnienia niektórych aspektów niezaobserwowanych różnic między grupą objętą interwencją a grupą kontrolną, podejście oparte na różnicy w różnicach stanowi w większości przypadków ulepszenie w stosunku do strategii dopasowywania przekrojowego. Na wykresie 4 przedstawiono wizualizację podejścia opartego na różnicy w różnicach.

Oś X przedstawia upływ czasu, a oś Y skalę, na której zapisuje się wyniki. Wynikami w tym przypadku mogą być wynagrodzenia. Średnie wynagrodzenia w przypadku grupy objętej interwencją w okresie przed interwencją wynoszą YT1, a w przypadku grupy kontrolnej – YC1. W okresie po interwencji wynagrodzenia wynoszą YT2 i YC2 odpowiednio w przypadku grupy objętej interwencją i grupy kontrolnej. Górna ciągła linia przedstawia zatem zmianę w wynagrodzeniach w grupie objętej interwencją, natomiast dolna ciągła linia – w grupie kontrolnej.

Przybliżone oszacowanie wpływu interwencji wynikałoby z porównania wynagrodzeń w grupach objętej interwencją i grupie kontrolnej w okresie po zakończeniu interwencji, tj. YT2 – YC2. Oszacowanie to byłoby jednak błędne, ponieważ nie uwzględniałoby różnic w wynagrodzeniach przed interwencją. Estymator różnicy w różnicach można rozumieć między innymi jako odjęcie szacowanego błędu systematycznego przed interwencją od różnicy wyników po jej zakończeniu. W rezultacie różnica w wynagrodzeniach po interwencji (YT2 – YC2) jest korygowana poprzez odjęcie od niej różnicy w wynagrodzeniach przed interwencją (YT1 – YC1), a zatem estymator wpływu różnicy w różnicach można wyrazić – w dużym uproszczeniu – w następujący sposób:

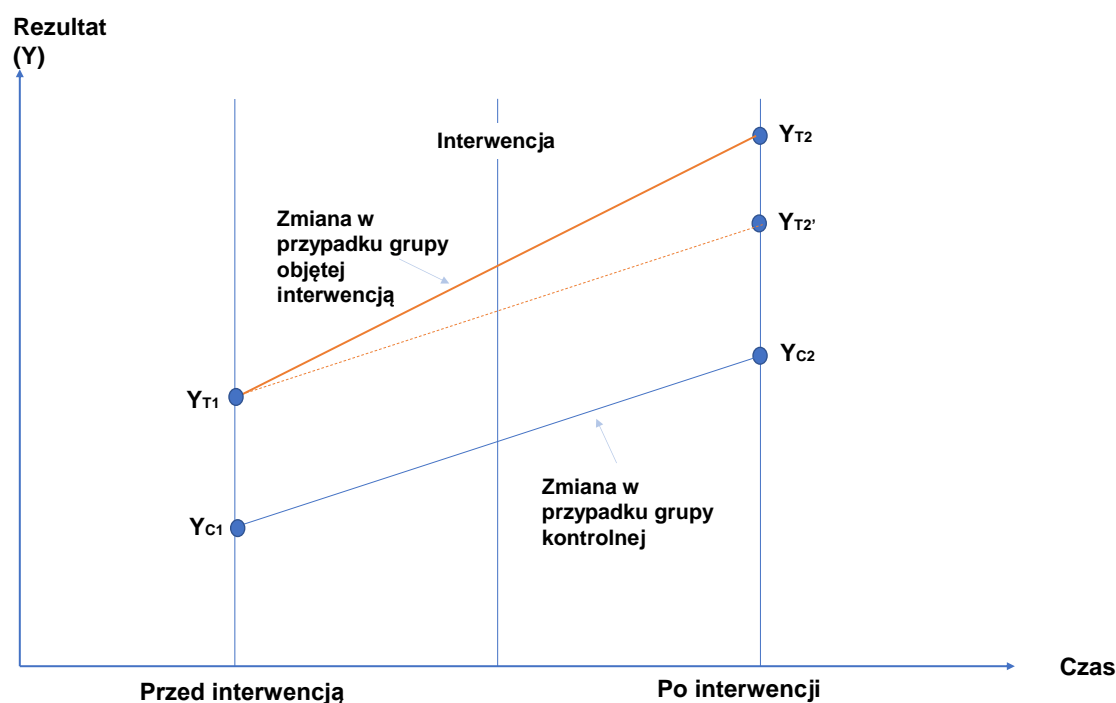
$$(YT2 - YC2) - (YT1 - YC1)$$

Logika podwójnej różnicy

Jeżeli różnice w wynagrodzeniach po zakończeniu interwencji nie zostają skorygowane o wcześniej istniejące różnice między grupą objętą interwencją i grupą kontrolną, otrzymane oszacowania mogą być obarczone błędem systematycznym. Alternatywnie, jak wspomniano wcześniej, podejście oparte na różnicy w różnicach można rozumieć jako odjęcie zmiany w rezultatach grupy kontrolnej od takiej samej zmiany zaobserwowanej w odniesieniu do grupy objętej interwencją. Zaobserwowaną zmianę w grupie kontrolnej postrzega się jako tę, która wystąpiłaby w grupie objętej interwencją w przypadku braku interwencji.

W najprostszym scenariuszu głównym założeniem, na którym opiera się podejście polegające na różnicy w różnicach, jest założenie o wspólnych tendencjach, tj. wykazywanych w czasie tendencjach w zakresie rezultatów grupy objętej interwencją i grupy kontrolnej, które są takie same w przypadku braku interwencji. Nie można bezpośrednio sprawdzić tego założenia, chociaż w przypadkach, gdy dostępne są wielokrotne pomiary zmiennej wpływającej na wyniki przed interwencją zarówno dla grupy objętej interwencją, jak i grupy kontrolnej, a pomiary te wykazują równoległe tendencje, sprzyja to większej wiarygodności podejścia. Przykład podejścia opartego na różnicy w różnicach można znaleźć w ramce poniżej.

Główne założenie: wspólne tendencje są takie same w przypadku braku interwencji

Wykres 9. Ilustracja podejścia opartego na różnicy w różnicach

Ramka 16. Przykład ewaluacji, w której zastosowano podejście oparte na różnicy w różnicach

Ocena działań na rzecz pracowników w starszym wieku w województwie lubelskim finansowanych ze środków EFS w okresie programowania 2007–2013⁸⁶

Głównym celem badania była ewaluacja interwencji w ramach EFS na rzecz wydłużania aktywności zawodowej osób starszych w kontekście niekorzystnej sytuacji demograficznej w województwie lubelskim w Polsce. Interwencje finansowano w ramach Programu Operacyjnego Kapitał Ludzki na lata 2007–2013.

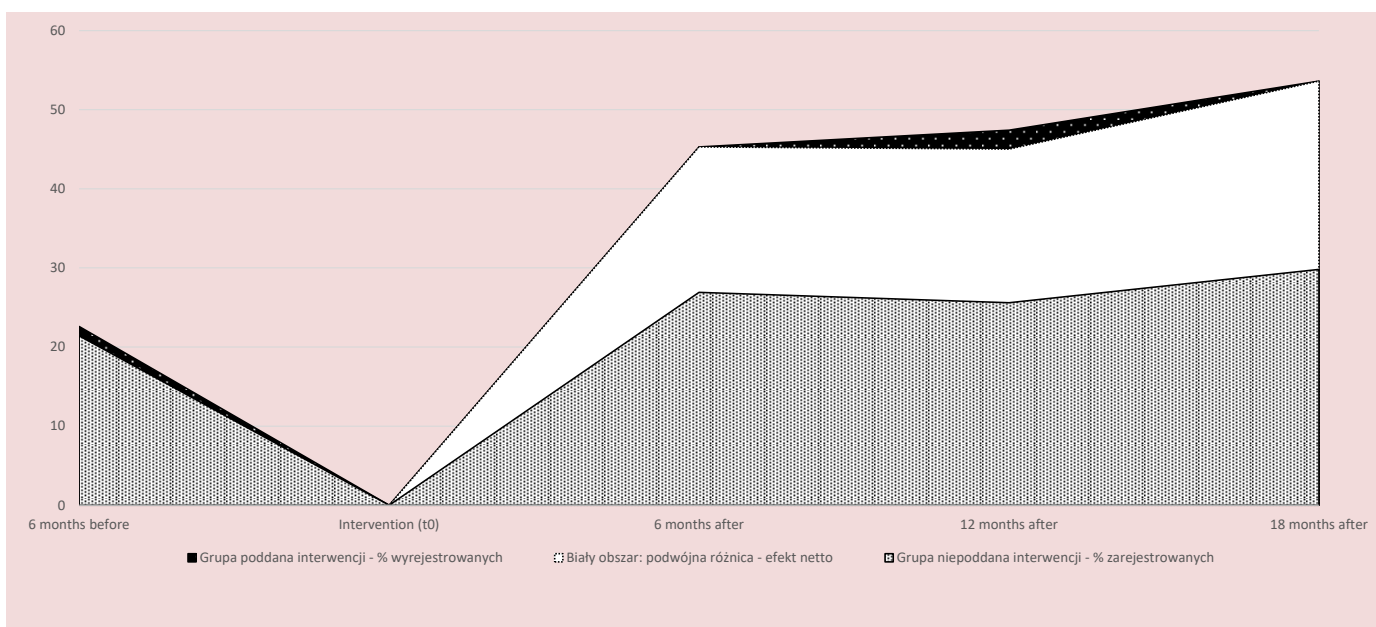
Grupę poddaną interwencji stanowiło 145 osób bezrobotnych w chwili udziału w interwencjach. Grupę kontrolną stanowiły zanonimizowane dane rejestrowe udostępniane przez urzędy pracy, zawierające dane społeczno-demograficzne oraz historię zdarzeń rejestrowych dotyczących statusu zatrudnienia konkretnych osób (dane uzyskano z ośmiu urzędów pracy i zawierały one informacje o 67 102 osobach). Zmienną wynikową był „status wpisu do rejestru bezrobotnych”⁸⁷, który można było zaobserwować zarówno przed interwencją, jak i po jej zakończeniu, co pozwoliło na zastosowanie podejścia opartego na różnicy w różnicach. Status zatrudnienia sprawdzano w 5 różnych momentach: 6 miesięcy przed rozpoczęciem udziału w projekcie, w trakcie udziału w projekcie oraz 6, 12 i 18 miesięcy po zakończeniu projektu.

Wybór grupy kontrolnej na potrzeby podejścia opartego na różnicy w różnicach uległ poprawie dzięki technikom parowania mającym na celu zidentyfikowanie osoby niepoddanej interwencji przypadającej na każdą osobę poddaną interwencji na podstawie szeregu możliwych do zaobserwowania zmiennych, takich jak płeć, wiek, wykształcenie, stopa bezrobocia w miejscu zamieszkania.

Analiza wykazała, że wsparcie na rzecz osób bezrobotnych w wieku 50+ przyniosło pozytywny efekt netto dla zatrudnienia: osoby bezrobotne 50+ poddane interwencji miały ponad 2,5-krotnie większą szansę na wyrejestrowanie się z rejestru osób bezrobotnych niż osoby niepoddane interwencji. Na poniższym wykresie przedstawiono zmiany zmiennej wynikowej w przypadku grupy poddanej interwencji i grupy kontrolnej oraz szacowany efekt netto po uwzględnieniu założenia dotyczącego podejścia opartego na różnicy w różnicach w zakresie równoległości.

⁸⁶ Zob. [Re-source Pracownia Badań i Doradztwa, 2015](#).

⁸⁷ Choć bycie niezarejestrowanym jako „osoba bezrobotna” nie jest tożsame z byciem „zatrudnionym”, założenie takie było konieczne do przeprowadzenia badania.



3.2.4. Model regresji nieciągłej

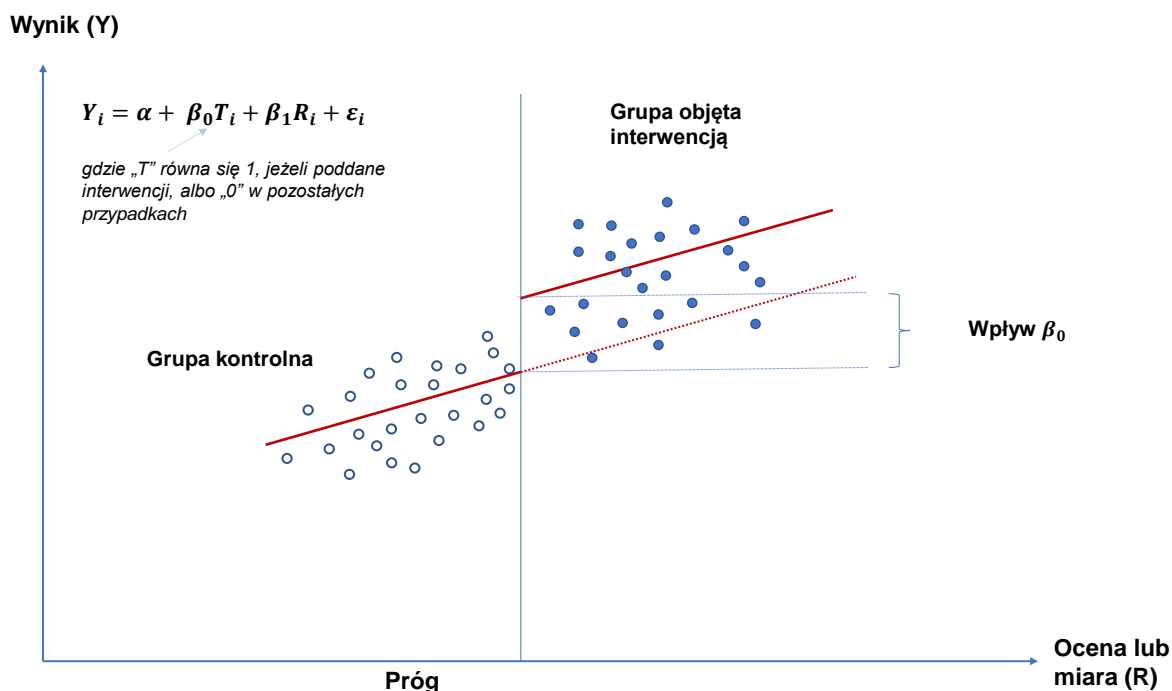
Najważniejsze cechy regresji nieciągłej

- Model regresji nieciągłej opiera się na założeniu, że określona wartość („próg”) punktacji lub oceny jest wyznacznikiem, czy dana osoba uczestniczy w interwencji czy nie.
- Osoby charakteryzujące się wartością zbliżoną do tej wartości uważa się za porównywalne. Różnica polega na tym, że osoby po jednej stronie progu uczestniczą w interwencji (grupa objęta interwencją), a osoby po drugiej stronie nie biorą w niej udziału (grupa kontrolna).
- Porównanie tych grup daje zwykle precyzyjną i intuicyjną miarę efektu interwencji, jednak technika ta ma zastosowanie wyłącznie w przypadku, gdy proces wyboru opiera się na punktacji lub ocenie.

Podejście oparte na **regresji nieciągłej** można przyjąć, gdy dostęp do interwencji określa się za pośrednictwem progu wyznaczonego dla oceny, skali lub miary o charakterze ciągłym. Na przykład dostęp do szkolenia może być uzależniony od wyników testu umiejętności, przy czym osoby, które uzyskały wynik powyżej określonego progu (lub wartości granicznej) biorą udział w szkoleniu, a osoby, które uzyskały wynik poniżej tego progu, nie uczestniczą w nim. Aby podejście to było ważne, próg należy określić bez znajomości wyników potencjalnych uczestników szkolenia; kandydaci z wynikiem bezpośrednio sąsiadującym z progiem będą do siebie bardzo podobni, a jedyna różnica będzie polegała na tym, że ci powyżej progu wezmą udział w interwencji, podczas gdy ci tuż poniżej nie będą w niej uczestniczyli. Wyniki osób powyżej i poniżej progu można porównać w celu uzyskania szacunkowego wpływu interwencji na progę.

Próg oddziela osoby poddane interwencji od osób należących do grupy kontrolnej

Wykres 10. Ilustracja podejścia opartego na regresji nieciągłej



Model regresji nieciągłej można stosować w sytuacji, gdy próg umożliwia pełną identyfikację grupy objętej interwencją (przy pełnej zgodności), w którym to przypadku uzyskuje się *ostrą* nieciągłość, albo gdy pod pewnymi warunkami nie wszystkie osoby po danej stronie progu ściśle przestrzegają przypisania do interwencji (nieciągłość *rozmyta*).

*Nieciągłość ostra
lub rozmyta*

Wykres 10 powyżej przedstawia stylizowany przykład modelu regresji nieciągłej. Jest to najprostszą ilustracją ostrej nieciągłości. Interwencja wywołuje stałe skutki w odniesieniu do każdej wartości oceny, a wpływ szacuje się przy użyciu modelu regresji liniowej (nie ma żadnych problemów dotyczących formy funkcyjnej regresji wpływu). W praktyce analiza będzie musiała niezmiennie charakteryzować się znacznie większym stopniem zaawansowania niż ta przedstawiona na wykresie 10.

Punkty na wykresie 10 reprezentują poszczególne jednostki, na przykład uczestników szkolenia. Na osi X oznacza się ocenę lub miarę wykorzystywane do przydzielania uczestnikom szkolenia godzin w harmonogramie kursu szkoleniowego. Osoby, które uzyskały wynik na prawo od pionowej linii ciągłej wskazującej próg w przypadku tej oceny lub miary (na przykład w teście umiejętności), kwalifikują się do szkolenia i wchodzi w skład grupy objętej interwencją. Potencjalni uczestnicy szkolenia, którzy uzyskali wynik poniżej progu oceny lub miary, nie kwalifikują się do szkolenia i wchodzi w skład grupy kontrolnej.

Kluczową kwestią jest to, że ocena stosowana do przydzielenia grupy docelowej do warunków interwencji albo kontrolnych jest ciągłą zmienną ilościową mierzoną przed interwencją, a dana osoba zostaje dopuszczona lub nie zostaje dopuszczona do programu szkoleniowego w zależności od tego, czy jej wynik jest powyżej czy poniżej wcześniej określonego progu.

Wynik nanosi się na osi Y. Zasadniczo wpływ interwencji identyfikuje się poprzez oszacowanie modelu regresji liniowej dla konkretnych danych (biorąc

pod uwagę powyższe założenia); tj. regresję zmiennej wyniku względem miary oceny wraz ze zmienną zero-jedynkową (wskaźnikiem interwencji), która określa, czy wynik jest poniżej czy powyżej progu (tj. czy dana osoba jest przypisana do grupy objętej interwencją czy do grupy kontrolnej).

Takie równanie regresji wpływu przedstawiono na wykresie 10. Efekt lub wpływ szkolenia w tym przykładzie uzyskuje się na podstawie współczynnika wskaźnika interwencji, tj. β_0 ⁸⁸. Dzięki temu można skutecznie stwierdzić, czy istnieje przerwa lub nieciągłość wokół progu, na co na wykresie 10 wskazuje przesunięcie w górę linii regresji w miejscu progu lub punkcie granicznym. W tym przykładzie zaobserwowano pozytywny wpływ szkolenia na wynik.

Alternatywnie oszacowanie wpływu można rozumieć jako przedłużenie linii przerywanej do linii grupy kontrolnej przedstawionej na wykresie 10. Można je uznać za oszacowanie scenariusza kontrfaktycznego w odniesieniu do grupy objętej interwencją – relacji między oceną a miarą wyniku, którą można by było zaobserwować przy braku interwencji – przy czym różnica między tą przerywaną linią a linią reprezentującą trend grupy objętej interwencją ilustruje efekt lub wpływ interwencji. Należy zauważyć, że przy braku interwencji nie pojawia się nieciągłość linii i zakłada się, że w takim przypadku wynik zmienia się w sposób ciągły wraz z oceną lub miarą.

Podejście oparte na regresji nieciągłej sprawdza się, ponieważ obserwacje w grupie objętej interwencją i grupie kontrolnej w pobliżu progu są do siebie podobne, z wyjątkiem faktu, że osoby powyżej progu w tym przykładzie uczestniczą w szkoleniu, podczas gdy osoby poniżej tego punktu nie biorą w nim udziału. Sytuacja nie różni się zatem od randomizacji w przypadku obserwacji w pobliżu progu. Istnieje jednak jedno istotne ograniczenie. W przypadku większości zastosowań oszacowanie wpływu przy użyciu podejścia opartego na modelu regresji nieciągłej może stanowić dla decydenta jedynie informację o efektach dotyczących osób znajdujących się blisko progu lub punktu granicznego. Stopień, w jakim można dokonać uogólnień na temat osób znajdujących się dalej od progu, może być ograniczony.

Wysoka wiarygodność wewnętrzna przy ograniczonej możliwości uogólnienia

Model regresji nieciągłej może być użytecznym podejściem, jeżeli osoby przydziela się do interwencji na podstawie potrzeb mierzonych za pomocą oceny lub wyniku o charakterze ciągłym. Analiza może jednak stać się złożona, w przypadku gdy próg jest rozmyty i występuje niezgodność, a także gdy występują problemy związane z formą funkcji modelu regresji wpływu. W rzeczywistości konieczne jest przyjęcie szeregu założeń, których słuszność nie zawsze można potwierdzić.

W kolejnej ramce przedstawiono praktyczny przykład, w którym podejście oparte na regresji nieciągłej zastosowano do ewaluacji gwarancji dla młodzieży.

⁸⁸ W prostym przypadku byłby to efekt zamiaru poddania oddziaływaniu w punkcie progu (zob. [Bloom H. S., 2009](#))

Ramka 17. Przykład ewaluacji, w której zastosowano podejście oparte na regresji nieciągłej

Vocational Training for Unemployed Youth in Latvia: Evidence from a Regression discontinuity design [Szkolenie zawodowe dla bezrobotnych osób młodych na Łotwie: dowody na podstawie modelu regresji nieciągłej]⁸⁹

Naukowcy ze Wspólnego Centrum Badawczego Komisji Europejskiej zastosowali podejście oparte na regresji nieciągłej, aby ocenić wpływ programu szkolenia zawodowego na wyniki na rynku pracy bezrobotnych osób młodych na Łotwie. Program ten był finansowany w ramach gwarancji dla młodzieży na lata 2014–2020 i skierowany do osób młodych w wieku od 15 do 29 lat, które nie kształcą się, nie pracują ani nie szkolą się (młodzież NEET).

Wykorzystane dane pochodziły z łotewskiej Państwowej Agencji Pracy (SEA), która przekazała informacje zarówno o uczestnikach, jak i osobach nieuczestniczących w programie, zarejestrowanych jako osoby bezrobotne w określonych okresach, i zostały dopasowane do danych z krajowego urzędu skarbowego, które zawierają informacje na temat dochodów osób w określonych terminach przed rozpoczęciem udziału w programie i po jego zakończeniu (styczeń 2012 r. i czerwiec 2017 r.). Po oczyszczeniu zbiorów danych ostateczna próba grupy poddanej interwencji wyniosła 898 osób, a grupy kontrolnej – 10 717.

Ewaluatorzy wykorzystali zasadę rządu łotewskiego, zgodnie z którą pierwszeństwo uczestnictwa w programie szkolenia zawodowego przyznaje się osobom bezrobotnym poniżej 25 roku życia. Wiek jest zatem zmienną bieżącą szkolenia zawodowego, która określa prawdopodobieństwo uczestnictwa w tym programie, przy czym wiek 25 lat stanowi próg, poniżej którego uczestnictwo w programie drastycznie wzrasta ze względu na zasadę pierwszeństwa. Ponieważ wiek osób nie zależy od nich samych, przydział do programu szkolenia zawodowego w okolicach progu można uznać za losowy. Zasada pierwszeństwa oraz fakt, że udział w badaniu był dobrowolny, wymagały od badających zastosowania modelu regresji nieciągłej w postaci rozmytej. Strategia identyfikacji opierała się na fakcie, że osoby w okolicy progu (wiek 25 lat) były do siebie podobne z wyjątkiem uczestnictwa w programie szkolenia zawodowego.

Wyniki pokazały, że wpływ prawdopodobnego przyszłego zatrudnienia i miesięcznego dochodu do 3,5 roku po przystąpieniu do programu jest pozytywny, ale statystycznie nieistotny. Stwierdzono jednak pozytywny wpływ zasady pierwszeństwa na uczestnictwo w programie. Ponieważ wiarygodność modelu regresji nieciągłej w postaci rozmytej opiera się na tym, że potencjalni uczestnicy nie mogą kontrolować zmiennej bieżącej (wiek), wprowadzono kilka testów w celu potwierdzenia tego efektu.

3.2.5. Zmienne instrumentalne

Najważniejsze cechy zmiennych instrumentalnych

- Koncepcja zmiennych instrumentalnych polega na tym, że pewien (z góry ustalony/zewnętrzny) proces determinuje uczestnictwo, jednak nie jest to w rzeczywistości procedura wyboru.
- Wstępnie ustalony proces i jego zmienne kluczowe można wykorzystać do uwzględnienia wszelkiej autoselekcji w decyzji osób o uczestnictwie.

W przypadku podejścia opartego na **zmiennych instrumentalnych** (IV) wybór do interwencji powinien być przynajmniej częściowo zdeterminowany przez czynnik zewnętrzny, który nie jest związany z rezultatami innymi niż te, uzyskane w drodze interwencji. W związku z tym czynnik zewnętrzny wpływa na uczestnictwo, ale nie bezpośrednio na rezultaty. Zazwyczaj takimi czynnikami zewnętrznymi mogą być błędy lub przeoczenia administracyjne lub inne losowe różnice w przeprowadzonej interwencji.

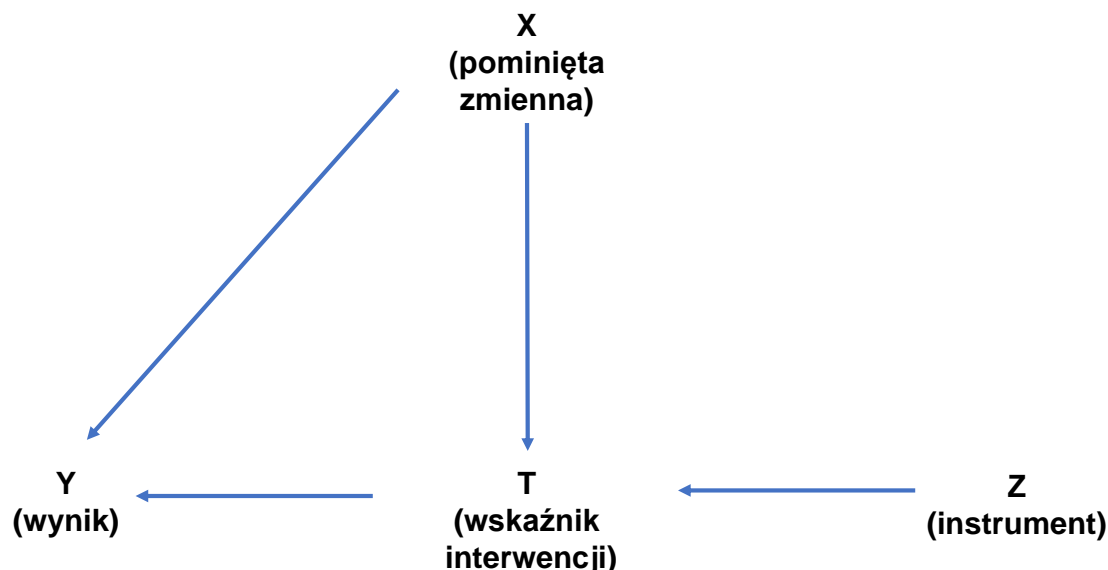
Na wykresie 11 poniżej przedstawiono podejście oparte na zmiennych instrumentalnych. Cztery zmienne zaprezentowano w bardzo uproszczonym

Wpływ czynników zewnętrznych na uczestnictwo

⁸⁹ Zob. [Bratti M. i in., 2018.](#)

systemie przyczynowo-skutkowym. Zmienne reprezentują dane zebrane wśród populacji hipotetycznie objętej interwencją szkoleniową (zarówno osób, które odbywają szkolenie, jak i tych, które należą do grupy kontrolnej).

Wykres 11. Ilustracja podejścia opartego na zmiennych instrumentalnych



„Y” oznacza rozpatrywany wynik. W przypadku interwencji szkoleniowej mogą to być zarobki. „T” to wskaźnik tego, czy dana osoba podjęła szkolenie⁹⁰.

„X” jest pominiętą zmienną, która nie jest obserwowana, ale jest związana zarówno z „Y” (wynik), jak i z „T” (wskaźnik interwencji), rozszerzając koncepcję programu szkoleniowego na przykład o podstawową miarę zdolności. W tym przypadku zdolność wiąże się zarówno z uczestnictwem w szkoleniu, jak i z zarobkami. Na przykład zdolniejsi członkowie grupy docelowej mogą zdecydować się na podjęcie szkolenia, a także korzystać z wyższych wynagrodzeń.

Istnienie „X” motywuje do poszukiwania instrumentu i zakłóca wpływ szkolenia na zarobki. Innymi słowy, oszacowanie jest obciążone błędem systematycznym z powodu istnienia „X” oraz faktu, że nie jest ono obserwowane i nie można go bezpośrednio uwzględnić w analizie.

Ponadto zmienna „Z” jest instrumentem, który według Morgana i Winshipa⁹¹ można traktować jako wstrząs dotyczący „T”, które jest niezależne od „X”. Z tego powodu na wykresie 11 nie ma linii, która łączyłaby „Z” z „X”. Co więcej, „Z” wpływa na „Y” wyłącznie poprzez „T”; „Z” nie wpływa na „Y” w żaden inny sposób. Oznacza to, że „Z” można wykorzystać do spowodowania zmienności „T” (interwencji), która nie jest skorelowana z zakłócającą zmienną „X”. W rezultacie dzięki wykorzystaniu tej zmienności, można uzyskać miarę wpływu „T” na „Y” nieobciążoną błędem systematycznym⁹².

Poniżej opisano najprostsze okoliczności, w których można przyjąć podejście IV oparte na zmiennych instrumentalnych, pomijając w tym miejscu wiele powiązanych zawiłości. W praktyce często trudno jest znaleźć przekonujący

*Identyfikacja
ważnego
instrumentu ma
zasadnicze
znaczenie*

⁹⁰ Innymi słowy, istnieje pełna zgodność, a wszystkie osoby z grupy objętej interwencją uczestniczą w szkoleniu.

⁹¹ Morgan S.L., Winship C., 2014.

⁹² Efekt przyczynowy między „T” i „Y” oblicza się, uwzględniając instrument, poprzez oszacowanie relacji między „Z” i „Y” oraz podzielenie jej przez szacowaną relację między „Z” i „T”.

instrument. Wiarygodność różnych potencjalnych instrumentów jest w dużym stopniu uzależniona od kontekstu, a leżących u ich podstaw założeń identyfikujących zasadniczo nie można przetestować statystycznie (w rzeczywistości wymaganą korelację między „Z” i „T” można przetestować statystycznie, natomiast „niezależności” „Z” i „X” nie można). Na przykład jedną ze strategii może być wykorzystanie odległości od ośrodków, w których prowadzone jest szkolenie (fizyczna lokalizacja szkolenia), do domu uczestnika szkolenia jako instrumentu służącego do oszacowania wpływu szkolenia na zarobki netto jego uczestników. Można zauważyć, że uczestnicy szkoleń, którzy mieszkają bliżej ośrodków szkoleniowych, częściej biorą udział w interwencji szkoleniowej oraz że odległość między domem uczestnika szkolenia a ośrodkiem szkoleniowym nie ma związku z innymi czynnikami determinującymi zarobki netto i uczestnictwo w szkoleniu (na przykład miarami kapitału ludzkiego). Odległość może zatem wpływać na zarobki netto wyłącznie poprzez jej wpływ na szkolenie⁹³.

Zmienne instrumentalne można stosować w szerokim zakresie kontekstów. Wartości szacunkowe można uzyskać poprzez różne podejścia do szacowania, w zależności od zmiennej odpowiedzi. Dotychczas podejście to nie było często stosowane w ewaluacjach EFS. W ramce 18 poniżej przedstawiono przykład analizy efektów przyczynowych w polityce z zastosowaniem podejścia IV opartego na zmiennych instrumentalnych.

Ramka 18. Przykład badania, w którym przyjęto podejście oparte na zmiennych instrumentalnych

Wpływ szkolenia językowego dla bezrobotnych imigrantów na zatrudnienie⁹⁴

Nabycie kompetencji w zakresie języka kraju przyjmującego jest dla imigrantów ważnym czynnikiem osiągnięcia wysokiego poziomu integracji na rynku pracy.

W najnowszym artykule (Lang J., 2021) autorka wykorzystuje oparte na zmiennych instrumentalnych podejście do oszacowania efektu przyczynowego programu szkolenia z języka niemieckiego do celów zawodowych dla wyników w zakresie zatrudnienia imigrantów uczestniczących w kursach dwa lata po zapisie. Program był realizowany przez niemiecki Federalny Urząd ds. Migracji i Uchodźców (BAMF) i finansowany ze środków EFS; program BAMF w ramach EFS był realizowany do końca 2017 r. Kursy składały się z kilku elementów: nauki języka niemieckiego, podnoszenia kwalifikacji zawodowych i staży zawodowych. Do udziału kwalifikowały się osoby z historią migracji, aktywne na niemieckim rynku pracy, bez względu na narodowość i datę imigracji.

Aby rozwiązać problem niemożliwej do zaobserwowania heterogeniczności pod względem umiejętności językowych, autorka wykorzystuje zewnętrzną zmienność w intensywności lokalnych szkoleń z języka na poziomie urzędu pracy. Instrument ten uznaje się za ważny, ponieważ urzędy pracy mają dużą swobodę we wdrażaniu różnych kombinacji programów w zależności od swoich „stylów prowadzenia polityki”, a zmienna ma charakter zewnętrzny w stosunku do wyników zatrudnienia osób poszukujących pracy. Analizie poddano kursy realizowane w 2014 r.

Ewaluacja opiera się na bogatym zbiorze danych administracyjnych pochodzących ze Zintegrowanych Historii Zatrudnienia, połączonej bazy danych zawierającej informacje administracyjne niemieckiej Federalnej Agencji Pracy. Zintegrowane Historie Zatrudnienia zawierają informacje na temat udziału w szkoleniach językowych, szczegółowe dane dotyczące historii zatrudnienia (z wyjątkiem samozatrudnienia), poszukiwania pracy, przelewów dokonanych w okresie bezrobocia oraz dane osobowe poszczególnych osób. Wykorzystano również inne dane (WGH, Werdegangshistorie), aby uzyskać szczegółowe informacje na temat uczestnictwa w kursach językowych prowadzonych przez BAMF oraz aby

⁹³ Interpretacja wyników takiej analizy może stać się bardziej złożona w zależności od tego, czy instrument jest skorelowany ze zmiennością efektu interwencji (zob. [Bryson i in., 2002](#)).

⁹⁴ Zob. [Lang, J., 2021](#).

uzupełnić brakujące dane ze Zintegrowanych Historii Zatrudnienia – tam gdzie było to możliwe – informacjami na temat samozatrudnienia i okresów urlopu rodzicielskiego.

Z ustaleń wynika, że po okresie efektu „lock-in” trwającym zasadniczo nieco ponad rok, dwa lata po udziale w interwencji imigranci uczestniczący w programie szkoleniowym mają o 9 punktów procentowych wyższe prawdopodobieństwo zatrudnienia niż grupa kontrolna, przy czym prawdopodobieństwo to nie ogranicza się do prac niewymagających kwalifikacji. Stwierdzono również, że wczesne zapewnienie szkolenia językowego (wkrótce po przybyciu do Niemiec) korzystnie wpływa na integrację imigrantów na rynku pracy.

Tabela 5. Porównanie najważniejszych cech głównych podejść do CIE

Podejście	Najważniejsze cechy	Zalety	Wymogi dotyczące danych	Ograniczenia
Randomizacja – podejście eksperymentalne	<ul style="list-style-type: none"> - Osoby kwalifikujące się do uczestnictwa są losowo przydzielane do grupy objętej interwencją i grupy kontrolnej - Randomizacja zapewnia, aby obie grupy były identyczne (co do zasady) pod względem wszystkich istotnych cech - W związku z tym grupa kontrolna oferuje odpowiedź na pytanie kontrafakcyjne, a różnica w wynikach między grupą objętą interwencją a grupą kontrolną stanowi efekt przyczynowy interwencji 	<ul style="list-style-type: none"> - W przypadku prawidłowego obliczenia wartości szacunkowe dotyczące wpływu nie są obciążone błędem systematycznym - Wyniki są przejrzyste i zrozumiałe - Ustalenia są w mniejszym stopniu narażone na zastrzeżenia i wątpliwości - Różnorodność dostępnych wariantów modeli umożliwiającą dostosowanie do wielu kontekstów politycznych i okoliczności interwencji 	<ul style="list-style-type: none"> - Podstawowy wymóg kontroli wyboru do interwencji poprzez randomizację - Zapis, kogo przydzielono do jakiej grupy - Zasadnicze znaczenie ma gromadzenie danych podstawowych - Pomiarów wyników należy rejestrować zarówno w przypadku grupy objętej interwencją, jak i grupy kontrolnej 	<ul style="list-style-type: none"> - Często grupa poddana interwencji i grupa kontrolna nie są zgodne z zasadami dotyczącymi przydziału do interwencji - Często wymagana jest zgoda uczestników - Randomizacja może wpływać na wybór osób uczestniczących w interwencji lub zgłaszających się do niej - Świadomość statusu przypisania może zmienić zachowanie uczestników i wpłynąć na wyniki - Wątpliwości natury etycznej - Znaczne wymagania dotyczące planowania i modelowania - Może być kosztowne (choć niekoniecznie)
Parowanie (wartość <i>propensity score</i>)	<ul style="list-style-type: none"> - Dopasowanie imituje randomizację przez skonstruowanie <i>ex post</i> grupy kontrolnej, która (co do zasady) jest jak najbardziej podobna do grupy objętej interwencją pod względem wszystkich istotnych cech - W przeciwieństwie do randomizacji dopasować można wyłącznie cechy możliwe do zaobserwowania (wiek, płeć, wykształcenie itp.), natomiast cech niemożliwych do zaobserwowania (np. motywacja) nie można brać pod uwagę - Zasadność tego podejścia w sposób krytyczny zależy zatem od dostępności danych 	<ul style="list-style-type: none"> - Wymaga dobrej znajomości procesów wyboru, ale nie wymaga bezpośredniego nadzoru nad wyborem do interwencji - Może być stosowane wstecznie, jeśli dostępne są odpowiednie dane, i w różnych kontekstach - Z technicznego punktu widzenia jest to semiparametryczna metoda szacowania; wymaga mniejszej liczby założeń parametrycznych (np. nie ma potrzeby stosowania standardowych założeń regresji) - Można je wykorzystać do oszacowania wielu skutków interwencji 	<ul style="list-style-type: none"> - Dokładna identyfikacja uczestników interwencji - Źródła danych, z których należy dobrać próbę - Zrozumiała koncepcja uczestnictwa i dobre zrozumienie wyboru do interwencji - Obszerne dane, najlepiej zebrane na etapie wyjściowym, na podstawie których można wybrać parę - Pomiarów rezultatów interwencji w przypadku jej uczestników i grupy kontrolnej 	<ul style="list-style-type: none"> - Wymaga znacznej ilości danych, które pozwolą na pełną charakterystykę procesu wyboru - Ważność zależy od jakości grupy kontrolnej i jej starannego wyboru oraz od skali wspólnego przedziału określoności - Opiera się na założeniu, że możliwe do zaobserwowania dane mogą odpowiednio charakteryzować wybór do interwencji - Zakres różnych dostępnych podejść do parowania wymaga dokładnej analizy - Wyniki mogą być trudne do wyjaśnienia i interpretacji, a także

Podejście	Najważniejsze cechy	Zalety	Wymogi dotyczące danych	Ograniczenia
				potencjalnie niejednoznaczne
Różnica w różnicach	<ul style="list-style-type: none"> - Różnica w różnicach to intuicyjne podejście, w ramach którego porównuje się różnicę w wynikach przed interwencją i po jej zakończeniu w grupie objętej interwencją z różnicą w wynikach przed interwencją i po jej zakończeniu w grupie kontrolnej - Ponieważ zmiana w czasie w grupie kontrolnej odzwierciedla, co stałoby się z grupą objętą interwencją w przypadku braku interwencji (scenariusz kontrfaktyczny), każda dodatkowa różnica w wynikach przed interwencją lub po niej w grupie objętej interwencją stanowi efekt przyczynowy interwencji - Jest to prosta metoda, którą można w praktyce zastosować w wielu sytuacjach 	<ul style="list-style-type: none"> - Uwzględnia niektóre aspekty niemożliwych do zaobserwowania różnic między uczestnikami interwencji a członkami grupy kontrolnej - Można stosować w połączeniu z parowaniem - Wykorzystuje dane z okresu przed interwencją i po jej zakończeniu, takie jak dane panelowe (dane indywidualne na przestrzeni czasu) lub wielokrotnie zbierane dane przekrojowe (dane dotyczące osób zebrane w różnym czasie) 	<ul style="list-style-type: none"> - Wymagania dotyczące danych są podobne do innych podejść, ale dodatkowo istnieje dotyczący wyników wymóg w zakresie pomiarów przed interwencją - Aby przetestować główne założenia potrzebne są liczne obserwacje wyników przed interwencją zarówno dla grupy objętej interwencją, jak i grupy kontrolnej 	<ul style="list-style-type: none"> - Wymaga przyjęcia założenia o wspólnych tendencjach w zakresie rezultatów wśród uczestników interwencji i członków grupy kontrolnej, aby móc się na nie powołać - Analiza może stać się dość złożona i podatna na błędną interpretację - Obszerne dane dotyczące wyników uzyskanych przed interwencją niezbędne do sprawdzenia założenia o wspólnych tendencjach - Nie można wykorzystać do oszacowania wielu skutków interwencji⁹⁵
Model regresji nieciągłej	<ul style="list-style-type: none"> - Model regresji nieciągłej opiera się na założeniu, że określona wartość („próg”) punktacji lub oceny jest wyznacznikiem, czy dana osoba uczestniczy w interwencji czy nie - Osoby charakteryzujące się wartością zbliżoną do tej wartości uważa się za porównywalne. Różnica polega na tym, że osoby po jednej stronie progu uczestniczą w interwencji (grupa objęta interwencją), a osoby po drugiej stronie nie biorą w niej udziału (grupa kontrolna) - Porównanie tych grup daje zasadniczo precyzyjną i intuicyjną miarę efektu interwencji, jednak technika ta ma zastosowanie wyłącznie w przypadku, gdy proces wyboru opiera się na punktacji lub ocenie. 	<ul style="list-style-type: none"> - Dostępne są zarówno ostre, jak i rozmyte podejścia do modelu regresji nieciągłej - Pod pewnymi warunkami umożliwia uzyskanie wpływu skutków interwencji nieobarczonego błędem systematycznym 	<ul style="list-style-type: none"> - Wybór progu musi być niezależny od wartości oceny przyznanej każdemu członkowi grupy docelowej - Wymagane są dane dotyczące osób zarówno z grupy objętej interwencją, jak i grupy kontrolnej w zakresie oceny lub miary, progu lub wartości granicznej oraz rezultatów 	<ul style="list-style-type: none"> - To podejście nie jest ważne bez ciągłej miary lub oceny, która określa interwencję - Analizy mogą stać się złożone i obarczone wątpliwościami, gdy kwestie formy funkcji regresji wpływu stają się powszechne, gdy występuje niezgodność lub gdy wielkość próby wokół progu jest ograniczona - Interpretacja wyników badań i ekstrapolacja uogólnień mogą nieść ze sobą pewne niebezpieczeństwa
Zmienne	- Koncepcja zmiennych instrumentalnych	- Umożliwia uzyskanie wysokiej	- Wymaga danych podstawowych,	- Znalezienie wiarygodnego

⁹⁵ Zob. [Frolich, M., 2004.](#)

Podejście	Najważniejsze cechy	Zalety	Wymogi dotyczące danych	Ograniczenia
instrumentalne (IV)	<p>polega na tym, że istnieje pewien (z góry ustalony/zewnętrzny) proces, który determinuje uczestnictwo, jednak nie jest on w rzeczywistości procesem wyboru</p> <p>- Następnie ten wstępnie ustalony proces można wykorzystać do uwzględnienia wszelkiej autoselekcji w rzeczywistym podejmowaniu decyzji o uczestnictwie przez poszczególne osoby</p>	<p>jakości oszacowań lub dowodów na istnienie efektów przyczynowych</p> <ul style="list-style-type: none"> - Rozwiązuje problem błędu systematycznego pominiętej zmiennej (lub błędu selekcji) - Można stosować wstecznie 	<p>danych o rezultatach i uczestnictwie w interwencji, a ponadto wymaga, aby można było zidentyfikować instrument</p> <ul style="list-style-type: none"> - Instrument musi być związany z uczestnictwem w interwencji i tylko na tej podstawie wpływać na wyniki <p>Instrument nie powinien być skorelowany z żadnymi innymi wyznacznikami wyników</p>	<p>instrumentu może być trudne</p> <ul style="list-style-type: none"> - Wytlumaczenie osobom niebędącym ekspertami może być trudne - Interpretacja wyników nie jest prosta; ograniczona możliwość przetestowania założeń identyfikacyjnych

Rozdział 4. Dalsze promowanie ewaluacji wpływu opartych na metodach kontrfaktycznych

Niniejszy przewodnik ma na celu zachęcenie IZ do częstszego przeprowadzania dobrej jakości ewaluacji wpływu opartych na metodach kontrfaktycznych i wsparcie ich w tym procesie. Stanowi on wytyczne dla osób odpowiedzialnych za planowanie i zlecanie ewaluacji wpływu interwencji współfinansowanych w ramach EFS+. Do tej pory skupiono się na planowaniu i wdrażaniu ewaluacji wpływu opartej na metodach kontrfaktycznych oraz omówiono szereg kluczowych pytań, które wymagają rozważenia. W okresie programowania 2014–2020 zwrócono jednak uwagę na szereg innych „szerszych kwestii” oraz wyzwania, z którymi należy się zmierzyć, aby wzmocnić ewaluacje EFS+ i wykorzystanie ewaluacji wpływu opartych na metodach kontrfaktycznych.

Ta sekcja przewodnika zawiera pewne sugestie dotyczące rozwiązywania tych „szerszych” kwestii. W szczególności omówiono w niej działania, które należy podjąć w celu rozwiązania następujących problemów:

Szersze kwestie do rozwiązania

- brak wiedzy IZ oraz szerszej społeczności odpowiedzialnej za kształtowanie polityki w państwach członkowskich na temat podejść do ewaluacji wpływu opartej na metodach kontrfaktycznych;
- brak zewnętrznych, odpowiednio wykwalifikowanych i doświadczonych wykonawców w państwach członkowskich, którzy byliby w stanie przeprowadzić ewaluacje wpływu oparte na metodach kontrfaktycznych;
- kwestia barier prawnych, z którymi należy się zmierzyć w ujęciu ogólnym we wszystkich ewaluacjach wpływu opartych na metodach kontrfaktycznych;
- postępy w kierunku lepszego planowania ewaluacji wpływu opartych na metodach kontrfaktycznych;
- rozszerzenie zakresu ewaluacji wpływu opartych na metodach kontrfaktycznych.

4.1. Poprawa poziomu zrozumienia wśród zainteresowanych stron

W przypadku okresu programowania 2021–2027 art. 44 ust. 1 rozporządzenia w sprawie wspólnych przepisów stanowi, że „[w] celu poprawy jakości opracowywania programów i ich wdrażania [...] instytucja zarządzająca przeprowadza ewaluacje programów pod kątem co najmniej jednego z następujących kryteriów: skuteczności, efektywności, trafności, spójności i unijnej wartości dodanej. Ewaluacje mogą również obejmować inne odpowiednie kryteria, takie jak poziom włączenia społecznego, niedyskryminację i widoczność, i mogą obejmować więcej niż jeden program”. Jak omówiono w poprzednich sekcjach, oznacza to, że odpowiednie zdolności w zakresie ewaluacji muszą być dostępne zarówno w IZ zlecającej

ewaluację wpływu opartą na metodach kontrfaktycznych, jak i u ewaluatorów ubiegających się o zamówienie.

W niektórych przypadkach służby prowadzące Punkt Informacyjny ds. Oceny odnotowały w niektórych administracjach brak zdolności do przeprowadzania ewaluacji wpływu opartych na metodach kontrfaktycznych – tj. organizowania przetargów i współpracy przy ich realizacji. Utrudniło to ewaluatorom przeprowadzenie ewaluacji wpływu opartych na metodach kontrfaktycznych, ponieważ nie określono z wyprzedzeniem jasnych pytań ewaluacyjnych, dostępności podstawowych danych ani dobrze przemyślanego planowania realizacji.

Czasami, szczególnie w małych krajach, na rynku usług doradczych brakuje możliwości przeprowadzenia ewaluacji wpływu opartej na metodach kontrfaktycznych; brakuje również fachowej wiedzy technicznej. W związku z tym istnieje powszechna potrzeba stymulowania zarówno popytu na ewaluacje wpływu oparte na metodach kontrfaktycznych, jak i ich podaży. Kwestia podaży może ulec poprawie, ponieważ IZ i państwa członkowskie zaczynają zlecać ewaluacje wpływu oparte na metodach kontrfaktycznych lub zgłaszać zapotrzebowanie na takie badania. Szybkość reakcji na zwiększone zapotrzebowanie na ewaluacje wpływu oparte na metodach kontrfaktycznych będzie zależała od wcześniejszych umiejętności, doświadczenia i istnienia w państwach członkowskich instytucji zdolnych do wdrożenia takich podejść. Stymulowanie popytu można jednak również częściowo osiągnąć poprzez poprawę znajomości i zrozumienia metod ewaluacji wpływu opartej na metodach kontrfaktycznych wśród osób pracujących w IZ.

Stymulowanie popytu na CIE i jej podaży

Jednym z rozwiązań tego problemu jest prowadzenie przez IZ kursów szkoleniowych z zakresu metod ewaluacji wpływu opartej na metodach kontrfaktycznych dla swoich pracowników. Szkolenia powinny skupiać się na korzyściach dla IZ wynikających z przyjęcia metod ewaluacji wpływu opartej na metodach kontrfaktycznych. Ponadto należy podkreślić kwestie odpowiedzialności i uczenia się tego, co się sprawdza. W załączniku 2 zamieszczono sugerowany konspekt kursu.

Opracowanie szkoleń z zakresu metod kontrfaktycznych ewaluacji wpływu

4.2. Budowanie zdolności

Kolejną kwestią poruszoną w okresie 2014–2020 i wspomnianą w powyższych sekcjach była potrzeba budowania zdolności do przeprowadzania ewaluacji wpływu opartych na metodach kontrfaktycznych w społecznościach badawczych/akademickich/doradczych państw członkowskich. W niektórych przypadkach było oczywiste, że umiejętności wymagane do tego typu działań były dostępne w państwach członkowskich, ale osoby posiadające te umiejętności napotykały przeszkody w ich zastosowaniu w kontekście ewaluacji (np. ograniczony dostęp do możliwych do wykorzystania danych lub problemy z określeniem rozsądnej grupy kontrolnej).

Istnieje szereg działań, które można podjąć w celu rozwoju podaży usług ewaluacyjnych. Wiele z poruszonych kwestii ma zastosowanie zarówno do ewaluacji wpływu opartej na metodach kontrfaktycznych, jak i do ewaluacji w bardziej ogólnym ujęciu. Zazwyczaj w celu poprawy podaży ewaluacji podejmuje się trzy działania:

Wzmacnianie instytucji i budowanie wspólnot praktyków

- budowanie relacji z instytucjami akademickimi, w szczególności z uczelniami;
- budowanie i wzmacnianie niezależnej społeczności konsultantów;
- wspieranie rozwoju profesjonalnej społeczności ewaluacyjnej.

Uczelnie

Rozwijanie powiązań z uczelniami jest ważne z dwóch powodów. Po pierwsze, nauczyciele akademicy mogą posiadać umiejętności i wiedzę wymagane do prowadzenia ewaluacji wpływu opartych na metodach kontrfaktycznych. Na przykład wielu mikroekonomistów, ekonometryków, socjologów ilościowych lub psychologów posiada niezbędne umiejętności. W wielu państwach członkowskich wymagane umiejętności mogą być dostępne, ale osoby, które je posiadają, nie pomyślały wcześniej o zastosowaniu ich do ewaluacji interwencji. Być może brakuje im bodźców, którymi należałoby się zająć.

*Rozwój
umiejętności
akademickich*

W niektórych państwach członkowskich istnieje tradycja aktywnego angażowania się pracowników naukowych w badania stosowane dotyczące polityki. W takim otoczeniu pracownicy akademicy orientują się w kwestiach współpracy z władzami i IZ. W innych państwach członkowskich, w których uczelnie i pracownicy akademicy nie są tak zaangażowani w prace stosowane, może być konieczna zmiana kulturowa. Jedną ze skutecznych metod rozwijania bazy dostawców w ramach sektora uczelni jest zapewnienie przez organy państw członkowskich i IZ podstawowego finansowania kosztów ośrodków badań naukowych zajmujących się metodami ewaluacji wpływu opartej na metodach kontrfaktycznych.

Uczelnie i pracownicy akademicy mogą również odegrać rolę w szkoleniu następnej generacji ewaluatorów. Ścisłe współpracując z uczelniami, można zachęcić je do włączenia metod ewaluacji programów do ich programów nauczania, a w ramach tej zmiany zapewnić, aby metody ewaluacji wpływu opartej na metodach kontrfaktycznych były uwzględniane w tych programach. W niektórych państwach członkowskich uczelnie mogą również odgrywać rolę w prowadzeniu kursów ustawicznego doskonalenia zawodowego w zakresie ewaluacji wpływu i metod ewaluacji wpływu opartej na metodach kontrfaktycznych. Może to być skierowane do decydentów, specjalistów technicznych w IZ, jak również innych potencjalnych dostawców, takich jak niezależni konsultanci. Państwa członkowskie mogą rozważyć zapewnienie finansowania takich szkoleń.

*Szkolenie
kolejnego
pokolenia*

Niezależni konsultanci

W przypadku niektórych ewaluacji przeprowadzanych na dużą skalę istnieje rynek międzynarodowy. Jest tak z pewnością w przypadku dużych ewaluacji wpływu opartych na metodach kontrfaktycznych. Wiele państw członkowskich będzie jednak chciało rozwijać krajowy potencjał w zakresie przeprowadzania ewaluacji wpływu opartych na metodach kontrfaktycznych. Jedną ze strategii prowadzących do osiągnięcia tego celu może być tworzenie strategicznych sojuszy między potencjalnymi dostawcami krajowymi i międzynarodowymi firmami konsultingowymi.

Rozwój rynku

Poniżej przedstawiono kilka sugestii dotyczących rozwoju bazy dostawców krajowych, którzy mogliby przeprowadzać ewaluacje wpływu oparte na metodach kontrfaktycznych, z których to sugestii mogą korzystać IZ (lub inne

organy) zlecające przeprowadzenie ewaluacji wpływu opartych na metodach kontrfaktycznych:

- naleganie, aby w ofertach konsorcjów lub partnerstw zawsze uwzględniano lokalnych konsultantów;
- dostosowywanie umów dotyczących ewaluacji w taki sposób, aby nowe, krajowe podmioty na rynku ewaluacji mogły przeprowadzać stosunkowo niewielkie ewaluacje o niskim ryzyku;
- zapewnienie, aby wymogi techniczne i finansowe związane z ubieganiem się o przeprowadzenie ewaluacji nie były zbyt restrykcyjne i umożliwiały uczestnictwo nowym podmiotom;
- przy dokonywaniu wyboru położyć nacisk na kryteria techniczne i wiedzę ekspercką, a nie na złożone procedury administracyjne, których mogą nie znać mniej doświadczeni konsultanci;
- organizowanie briefingów z potencjalnymi konsultantami, aby odpowiedzieć na pytania i zachęcić do składania ofert w konkurencyjnym środowisku;
- wspieranie tworzenia sieci kontaktów wśród stosunkowo odizolowanych konsultantów zajmujących się ewaluacją, aby zachęcać do budowania zespołów, tworzenia konsorcjów oraz innych sieci i stowarzyszeń zawodowych, również na poziomie międzynarodowym;
- przyznanie przez zleceniodawców ewaluacji, że być może będą musieli przyjąć bardziej bezpośrednie podejście do zarządzania nowymi wykonawcami, aby przyspieszyć zdobywanie przez nich wiedzy i doświadczenia.

Ramka 19. Przykład projektu mającego na celu wzmocnienie kultury i zdolności w zakresie ewaluacji wpływu opartej na metodach kontrfaktycznych

W okresie programowania 2014–2020 w ramach EFS w Szwecji realizowano projekt mający na celu poprawę kultury i zdolności w zakresie ewaluacji po stronie publicznych służb zatrudnienia. Projekt pod nazwą „Evidence based EU funded projects” [„Oparte na dowodach projekty finansowane przez UE”] realizowano w latach 2016–2018. W ramach tego projektu w warsztatach i innych działaniach szkoleniowych w zakresie ewaluacji uczestniczyli kierownicy publicznych służb zatrudnienia oraz inni pracownicy.

Działania nie skupiały się na kwestiach technicznych, ale na znaczeniu ewaluacji, głównych etapach ewaluacji wpływu, znaczeniu rzetelnych danych i zapewnieniu jakości ewaluacji. Utworzono stronę internetową i zamieszczono na niej materiały edukacyjne.

Głównymi rezultatami projektu była poprawa wiedzy kierowników publicznych służb zatrudnienia w zakresie ewaluacji, ich otwartość na praktykę ewaluacji oraz nastawienie na współpracę. W szczególności, według osoby, z którą przeprowadzono wywiad, „wpływ projektów na liczbę przeprowadzonych ewaluacji wpływu” nie jest jeszcze duży, ale widać pewien postęp, biorąc pod uwagę, że w poprzednim okresie programowania 2007–2013 w Szwecji nie przeprowadzono żadnych ewaluacji wpływu opartych na metodach kontrfaktycznych.

Ewaluacja projektu „Ung framtid” („Młoda przyszłość”), w której zastosowano podejście oparte na randomizacji, jest konkretnym przykładem zastosowania tej metody w praktyce przy zaangażowaniu w proces ewaluacji kierowników publicznych służb zatrudnienia⁹⁶.

⁹⁶ Informacje na podstawie wywiadu z urzędnikiem państwowym szwedzkich publicznych służb zatrudnienia – Arbetsförmedlingen.

Spółeczność zawodowa

Ważne jest, aby w państwach członkowskich budować społeczności zawodowe zajmujące się ewaluacją, w których jest miejsce na dyskusję o metodach ewaluacji wpływu opartej na metodach kontrfaktycznych i wymianę doświadczeń. Rozwój społeczności zawodowych jest ważny dla wzajemnego wsparcia i uczenia się, a także dla utrzymania standardów jakości. Przydatną strategią mogłoby być nawiązanie kontaktów z odpowiednimi krajowymi stowarzyszeniami ewaluacyjnymi i zachęcanie ich do promowania ewaluacji wpływu opartych na metodach kontrfaktycznych dotyczących interwencji w ramach EFS+ podczas szkoleń, specjalnych konferencji lub seminariów bądź sesji informacyjnych.

Tworzenie społeczności zawodowych

Wymiana doświadczeń

Komisji Europejskiej zależy na przeprowadzaniu bardziej rygorystycznych ewaluacji wpływu EFS+, a ewaluacja wpływu oparta na metodach kontrfaktycznych była powszechnie zalecana w okresie 2014–2020⁹⁷. Obecnie można stwierdzić, że wykorzystanie ewaluacji wpływu opartych na metodach kontrfaktycznych zwiększyło się, ale nadal w wielu IZ i państwach członkowskich jest ograniczone. Wymiana doświadczeń w zakresie metod ewaluacji wpływu opartej na metodach kontrfaktycznych jest jednym z najważniejszych sposobów budowania zdolności i wspierania rozpowszechniania ewaluacji wpływu opartych na metodach kontrfaktycznych w całej UE. W tym celu należy wykorzystać istniejące fora wzajemnego uczenia się, takie jak wzajemne oceny polityki zatrudnienia i włączenia społecznego oraz wspólnoty praktyków w ramach EFS+. Inicjatywy poszczególnych IZ lub państw członkowskich, takie jak międzynarodowe konferencje lub seminaria, mogłyby również sprzyjać wymianie doświadczeń.

Wykorzystanie istniejących forów

4.3. Pokonywanie barier prawnych

Jednym z najbardziej znaczących i istotnych problemów napotykanym przez naukowców prowadzących ewaluacje wpływu oparte na metodach kontrfaktycznych w państwach członkowskich jest uzyskanie dostępu do danych. W szczególności często napotykają oni bariery prawne, które mają na celu ochronę poufności danych osób uwzględnionych w zbiorach danych. Jak wykazano powyżej w odniesieniu do cech RODO, rozwiązanie tych problemów polega na wprowadzeniu szerszych reform i zawarciu umów umożliwiających udostępnianie ewaluatorom odpowiednich danych w sposób kontrolowany i na bieżąco.

Usuwanie barier prawnych w dostępie do danych

Analityczne wersje zbiorów danych administracyjnych mogłyby być regularnie tworzone na podstawie danych administracyjnych będących w posiadaniu organów państw członkowskich, dokumentowane i deponowane w archiwum, do którego dostęp byłby kontrolowany. Upoważnieni wykonawcy mogą pobierać dane z takich zbiorów na podstawie upoważnienia. Jeżeli nie byłoby możliwe uzyskanie konkretnej zgody, dane zostałyby w pełni zanonimizowane przy użyciu zaszyfrowanych identyfikatorów osobowych w celu przestrzegania przepisów RODO. Takie zbiory danych utworzono w kilku

Tworzenie analitycznych zbiorów danych

⁹⁷ W załączniku XI do rozporządzenia w sprawie wspólnych przepisów na lata 2014–2020 (rozporządzenie (UE) nr 1303/2013) zwrócono się o stworzenie skutecznego „systemu wskaźników rezultatu niezbędnych [...] do monitorowania postępów w osiągnięciu rezultatów oraz do podejmowania ewaluacji oddziaływania”. Wymóg ten stanowił element warunku wstępnego dotyczącego okresu programowania 2014–2020.

krajach. W przypadku gdy dostęp nadal jednak stanowi przeszkodę ze względu na różne interpretacje przepisów dotyczących prywatności, inicjatywa krajowa na szczeblu rządowym powinna promować porozumienia i systemy umożliwiające przekazanie danych do celów badawczych w stosunkowo krótkim czasie.

Jeżeli nadal istnieją obawy co do poufności danych osobowych, należy rozważyć utworzenie laboratoriów danych. W tym przypadku ewaluatorzy wykorzystujący w pracy administracyjne zbiory danych otrzymywaliby dostęp do rejestrów tylko w bezpiecznych miejscach, gdzie dostęp do danych podlega ścisłemu monitoringowi i kontroli. Dane musiałyby być przetwarzane i analizowane na miejscu, a na zewnątrz można by było wnosić jedynie wyniki ewentualnych analiz.

Tworzenie laboratoriów danych

4.4. W kierunku bardziej prospektywnych podejść

Wspólną cechą niewielkiej jak dotąd liczby ewaluacji wpływu opartych na metodach kontrfaktycznych przeprowadzonych w odniesieniu do interwencji finansowanych z EFS jest ich retrospektywny raczej niż prospektywny charakter. Oznacza to, że fachowym ewaluatorom zleca się przeprowadzanie ewaluacji interwencji, które zostały opracowane bez uwzględnienia ewaluacji, a w niektórych przypadkach interwencji, w odniesieniu do których w ogóle nie zaplanowano ewaluacji wpływu lub zaplanowano ją tylko w niewielkim stopniu. Oznacza to, że ewaluatorzy muszą tworzyć źródła danych w sposób czasochłonny, kosztowny i nieoptymalny, reagując na dane, które akurat były dostępne, a nie na źródła danych stworzone z myślą o ewaluacji wpływu.

Podejście prospektywne wiązałoby się natomiast z zaangażowaniem ewaluatorów w planowanie ewaluacji wpływu opartej na metodach kontrfaktycznych przy najbliższej okazji i umożliwiałoby wpływanie na interwencje (nowe albo istniejące), często w dość subtelny sposób, czyniąc je bardziej podatnymi na ewaluację wpływu opartą na metodach kontrfaktycznych. To właśnie planowanie ewaluacji wpływu opartej na metodach kontrfaktycznych z wyprzedzeniem może zdecydować o tym, czy możliwe będzie przeprowadzenie rygorystycznej ewaluacji, czy w ogóle nie będzie takiej możliwości. Zaangażowanie odpowiednio przeszkolonych pracowników wewnętrznych lub zewnętrznych wykonawców o statusie ekspertów na wczesnym etapie prowadzenia interwencji lub w momencie podejmowania decyzji w sprawie finansowania oznacza, że:

W ramach podejść prospektywnych od samego początku angażuje się ewaluatorów

- odpowiednie prowadzenie dokumentacji można włączyć do realizacji programów i interwencji;
- niezbędne źródła danych można zidentyfikować na wczesnym etapie, a kwestie dostępu i ochrony danych można rozwiązać w odpowiednim czasie;
- można określić dane podstawowe i w razie potrzeby przeprowadzić ankiety;
- praktyczne kwestie związane ze sposobem rekrutacji uczestników do interwencji można rozwiązać w taki sposób, aby procesy rekrutacji były bardziej spójne z rygorystyczną ewaluacją.

Zaangażowanie ewaluatorów przeszkolonych w zakresie metod ewaluacji wpływu opartej na metodach kontrfaktycznych (niezależnie od tego, czy są to

ewaluatorzy wewnętrzni IZ, czy eksperci zewnętrzni) w proces opracowywania nowych interwencji w ramach EFS+ lub w podejmowanie decyzji dotyczących istniejących interwencji umożliwiłoby rozpoczęcie planowania ewaluacji wpływu na początku okresu programowania, a także przyniosłoby znaczące korzyści w zakresie podejmowania decyzji politycznych na podstawie dowodów.

4.5. Rozszerzenie zakresu ewaluacji wpływu opartej na metodach kontrfaktycznych

Ostatnia, bardzo ważna kwestia dotyczy zakresu politycznego ewaluacji wpływu opartej na metodach kontrfaktycznych i ma różne wymiary: zakres polityk objętych ewaluacją, przeanalizowane rezultaty i kompletność interpretacji ustaleń.

W latach 2014–2020 upowszechnianie ewaluacji wpływu opartych na metodach kontrfaktycznych koncentrowało się niemal wyłącznie na aktywnej polityce rynku pracy i skutkach związanych ze statusem zatrudnienia. Prawdopodobnie można to wyjaśnić faktem, że dane administracyjne dotyczące zatrudnienia są zasadniczo łatwiej dostępne, a dane te stanowią kluczową zmienną wynikową potrzebną do analizy wielu aktywnych polityk rynku pracy.

Większy nacisk na CIE w polityce włączenia społecznego i polityce edukacyjnej

Polityk w zakresie edukacji i włączenia społecznego nie poddano ewaluacji wpływu opartej na metodach kontrfaktycznych lub poddawano je takiej ewaluacji bardzo rzadko, pomimo ich znaczenia dla strategii EFS i przydziału środków; w ramce 20 poniżej przedstawiono niektóre z niewielu dostępnych przykładów z Hiszpanii, Polski i Portugalii.

Aby przeciwdziałać temu niedoborowi ewaluacji wpływu opartych na metodach kontrfaktycznych i pogłębić wiedzę na temat tych ważnych obszarów polityki, konieczne jest podjęcie szeregu połączonych inicjatyw:

Angażowanie dodatkowych podmiotów i rozwijanie odpowiednich źródeł danych

- planowanie i przygotowywanie ewaluacji wpływu opartych na metodach kontrfaktycznych dotyczących środków w zakresie włączenia społecznego lub edukacji z wyprzedzeniem, tak aby ocenić dostępność danych w czasie; rozpoczęcie niezbędnej współpracy między różnymi podmiotami w celu zwiększenia tej dostępności; oraz określenie na wczesnym etapie grupy objętej interwencją i grupy kontrolnej;
- promowanie zaangażowania właścicieli danych w projektowanie ewaluacji wpływu opartej na metodach kontrfaktycznych; szersze zaangażowanie w ewaluację wpływu oparte na metodach kontrfaktycznych organów administracji odpowiedzialnych za edukację i politykę społeczną, w przypadku gdy nie jest to IZ, mogłoby zaowocować większym i bardziej powszechnym zaangażowaniem w ewaluację i udostępnianie danych;
- stymulowanie zawierania porozumień instytucjonalnych i stosowania oprogramowania komputerowego w celu umożliwienia wykorzystania administracyjnych danych dotyczących warunków społecznych i edukacji na potrzeby ewaluacji wpływu opartych na metodach kontrfaktycznych. W tym celu ważne może być zaangażowanie krajowego urzędu statystycznego, jako instytucji „pomostowej” i umożliwiającej przeprowadzenie ewaluacji, ponieważ urzędy te posiadają wymagane umiejętności i opracowują już wiele zbiorów danych administracyjnych na potrzeby tworzenia statystyk krajowych.

Wspieranie gromadzenia odpowiednich danych od uczestników należących do grupy objętej interwencją i do grupy kontrolnej na początku interwencji, aby mieć spójny zbiór danych „przed i po”. W przypadku gdy dane administracyjne nie są dostępne lub adekwatne, niezbędne dane można uzyskać w drodze specjalnego badania, jednak musi się ono rozpocząć na początku interwencji i obejmować również grupę kontrolną, aby zapewnić konieczne porównania.

Ramka 20. Przykłady ewaluacji w obszarze edukacji

Interwencje w ramach EFS przeciwko wczesnemu kończeniu nauki prowadzone w Asturii⁹⁸

W ramach ogólnej ewaluacji PO Principado de Asturias ESF na lata 2014–2020 przeprowadzono specjalną analizę wpływu programu Diversificación Curricular y de Mejora del Aprendizaje y del Rendimiento (PMAR). PMAR to działanie w ramach EFS realizowane w ramach PI10.i, mające na celu przeciwdziałanie wczesnemu kończeniu nauki. W ramach PMAR uczniów dzieli się na określone grupy (od 8 do 15 uczniów, jednak dopuszcza się także inne wielkości, jeśli zaistnieją szczególne okoliczności), które podejmą naukę następujących przedmiotów: lingwistyki, nauk społecznych, nauk ścisłych i matematyki, a także języków obcych (pozostałe przedmioty są nauczane w klasie głównej). Działanie realizowano w roku akademickim 2016–2017, w drugiej i trzeciej klasie szkoły średniej (ESO). Do udziału w programie kwalifikują się uczniowie, którzy powtarzali co najmniej jeden kurs na dowolnym etapie, oraz ci, którzy ukończyli pierwszą klasę ESO, ale nie mogą przejść do następnej klasy. Ostatecznego wyboru dokonuje się na podstawie indywidualnej oceny uczniów (akademickiej i psychopedagogicznej) przeprowadzonej przez zespoły dydaktyczne.

Łącznie w programie wzięło udział 1 053 uczniów: 512 z drugiej i 541 z trzeciej klasy ESO. Uczniów tych porównano z innymi osobami z trudnościami⁹⁹ w uczeniu się w celu wybrania grupy kontrolnej porównywalnej pod względem kryteriów kwalifikacyjnych ze uczniami objętymi interwencją (łącznie 3 852 osoby).

Zastosowano metodę parowania, wykorzystując dane administracyjne z bazy danych SAUCE (Sistema Informático para la Administración Unificada de Centros Educativos) przekazywane w formie zanonimizowanej przez Ministerstwo Edukacji. Dane te wykorzystano do obliczenia wartości *propensity score*¹⁰⁰ oraz do pomiaru zmiennej wynikowej, którą stanowi pomyślne zaliczenie roku akademickiego. Ewaluacja wykazała pozytywny i statystycznie istotny efekt interwencji: objęci interwencją uczniowie drugiej klasy wykazują wskaźnik sukcesu o 18 punktów procentowych wyższy niż grupa kontrolna, a w przypadku poddanych interwencji uczniów trzeciej klasy różnica wynosi około 16 punktów procentowych. W obu grupach wpływ ten jest bardziej pozytywny w przypadku kobiet niż mężczyzn.

Interwencje w ramach EFS na rzecz poprawy kształcenia zawodowego w województwie podlaskim¹⁰¹

W ramach badania zbadano skutki projektu „Dobry zawód-fajne życie”, którego celem była popularyzacja kształcenia zawodowego w województwie podlaskim w Polsce. Grupą docelową projektu była kadra szkół gimnazjalnych, ich uczniowie i rodzice, a działania realizowano zarówno na rzecz instytucji, jak i osób (dostosowane strategie marketingowe i komunikacyjne dla szkół, szkolenia dla nauczycieli, współpraca z przedsiębiorstwami i lokalnymi liderami branż, mentoring dla uczniów i rodziców, doradztwo zawodowe i edukacyjne dla uczniów). Projekt realizowano w latach 2017–2019.

Populację poddaną interwencji stanowili uczniowie, którzy uczestniczyli w projekcie poddanym ewaluacji. Spośród 9 403 uczniów 1 500 wybrano w drodze losowania warstwowego, a 200 wzięło udział w badaniu ankietowym. Informacje na temat uczniów, którzy uczestniczyli w projekcie, pochodzą z danych z monitoringu. Wskazano, że grupa kontrolna składa się z uczniów, którzy wykazują cechy podobne do uczniów należących do grupy poddanej interwencji, ale którzy nie brali udziału w projekcie. Liczebność grupy kontrolnej mniej więcej dwukrotnie przewyższała liczbę członków grupy objętej interwencją, a

⁹⁸ Zob. [Diaz J.M. i in., 2019.](#)

⁹⁹ Określono ich jako uczniów, którzy w roku szkolnym 2015–2016 byli w pierwszej lub drugiej klasie szkoły średniej, ukończyli kurs, nie zaliczając co najmniej jednego przedmiotu, i powtarzali co najmniej jeden kurs, gdy uczęszczali do szkoły.

¹⁰⁰ Zmienne, na podstawie których dopasowano obie grupy, były następujące: wyniki w nauce w poprzednim roku szkolnym, rodzaj szkoły (publiczna czy prywatna), zapisanie się na inny, dodatkowy kurs w poprzednim roku szkolnym (tak lub nie), płeć, średni dochód w gminie zamieszkania, miejsce zamieszkania, kraj urodzenia (Hiszpania czy reszta świata).

¹⁰¹ Zob. [Zub M. i in., 2020.](#)

planowanie i wybór jej członków przeprowadzono losowo z użyciem Powszechnego Elektronicznego Systemu Ewidencji Ludności (PESEL). Ze względu jednak na trudności spowodowane pandemią COVID-19 zastosowanie zaplanowanej metodyki było niemożliwe. Ostatecznie jako główną metodę doboru próby zastosowano metodę kuli śnieżnej. W rezultacie przeprowadzono 401 wywiadów, z których 384 wykorzystano w badaniu.

Dane dotyczące zmiennych wynikowych¹⁰² zarówno w przypadku grupy objętej interwencją, jak i grupy kontrolnej zgromadzono w drodze badania kwestionariuszowego przeprowadzanego w drodze wywiadu bezpośredniego, telefonicznego lub internetowego. W badaniu zastosowano technikę *propensity score weighting* jako metodę kontryfaktyczną służącą ocenie wpływu (jedynie na poziomie ogólnym, a nie w odniesieniu do poszczególnych podgrup). W wyniku analiz kontryfaktycznych przeprowadzonych w ramach badania nie udało się sformułować jednoznacznych wniosków. Poziom istotności rezultatów był niezadowolający.

Ewaluacja uczelnianego systemu stypendialnego dla studentów w trudnej sytuacji materialnej w Portugalii¹⁰³

Druga ewaluacja, jeszcze oficjalnie nieopublikowana, dotyczy programu stypendialnego dla studentów z gospodarstw domowych o niskich dochodach. Celem tego programu jest sprzyjanie dostępowi do szkolnictwa wyższego i podniesienie poziomu uczestnictwa w kształceniu. Program jest objęty wsparciem z EFS (od okresu 2007–2013) w regionach Północ, Centrum i Alentejo oraz wsparciem państwa w pozostałych regionach. Do otrzymania stypendium kwalifikują się studenci, którzy nie osiągnęli określonego progu dochodowego (dochód na głowę na poziomie 6,8–7,9 tys. euro) i którzy zdobyli minimalną liczbę punktów w poprzednim roku studiów. To drugie kryterium nie dotyczy studentów pierwszego roku akademickiego. Zasady kwalifikowalności są takie same zarówno w przypadku uczelni publicznych, jak i uczelni prywatnych. Kwota stypendium jest proporcjonalna do dochodu gospodarstwa domowego.

Od 2011 r. każdego roku w programie udział bierze około 70 000 studentów. Ewaluacja obejmuje jednak okres począwszy od roku 2012, gdyż od tego roku dane stały się dostępne. Ponadto w ramach ewaluacji skoncentrowano się na studentach pierwszego roku studiów magisterskich lub licencjackich, ubiegających się o przyznanie stypendium po raz pierwszy. O kwalifikowalności decyduje kryterium dochodowe¹⁰⁴.

Zastosowano dwa administracyjne zbiory danych: dane dotyczące studentów ubiegających się o stypendium w latach 2012–2018 (dane przekazane przez Dyрекcję Generalną ds. Szkolnictwa Wyższego w Portugalii) połączono za pomocą niepowtarzalnego identyfikatora z innym zbiorem danych zawierającym informacje na temat przebiegu studiów i postępów (dane przekazane przez Dyрекcję Generalną ds. Statystyki i Nauki). Ostateczna próba objęta analizą składa się z 156 002 studentów, z których 130 602 należy do grupy poddanej interwencji, a pozostałe 25 400 nie należy do tej grupy. Za sprawą progu dochodowego (zmienna bieżąca) ewaluator mógł zastosować model regresji nieciągłej i dokonano pomiaru szeregu zmiennych wynikowych: w perspektywie krótkoterminowej odsetek studentów pozostający na liście studentów na koniec pierwszego roku, oraz punkty zdobyte na koniec pierwszego roku; w perspektywie długoterminowej – prawdopodobieństwo ukończenia studiów, ostateczna uzyskana ocena oraz liczba lat pozostałych do ukończenia studiów.

W początkowym okresie zaobserwowano pozytywny wpływ pod względem współczynników skolaryzacji, ale nie pod względem liczby uzyskanych punktów. W dłuższym okresie większa liczba studentów poddanych interwencji ukończyła studia w krótszym czasie i z większą liczbą punktów niż studenci w grupie kontrolnej. Większy wpływ zaobserwowano wśród mężczyzn, studentów zamieszkałych na terytoriach objętych wsparciem z EFS oraz studentów studiów magisterskich i studentów uczących się na uczelniach państwowych.

¹⁰² Zmienne wynikowe (na podstawie odpowiedzi udzielonych w ramach badania) były następujące: uczęszczanie w momencie badania do technikum, uczęszczanie w tym czasie do szkoły branżowej I stopnia oraz uczęszczanie w tym czasie do którejkolwiek z tych szkół, udział w rekrutacji na uczelnię, do technikum, do szkoły branżowej I stopnia, do szkoły kształcenia zawodowego (technikum lub szkoły branżowej I stopnia), szkoła w dużym stopniu spełnia oczekiwania ucznia, kształcenie zawodowe zdecydowanie odpowiada zainteresowaniom ucznia.

¹⁰³ Zob. Guthmuller S., Meroni E.C., dotychczas nie opublikowano.

¹⁰⁴ Przewidziano analizę całej próby studentów (nie tylko studentów pierwszego roku). Analiza ta jest jednak jeszcze w toku.

Już wcześniej podkreślono przydatności oraz charakter objaśniający metaewaluacji. Ogólny wzrost liczby ewaluacji wpływu opartych na metodach kontrfaktycznych i w rezultacie ewentualny wzrost liczby metaewaluacji spowodowałby intensyfikację debaty na temat konkretnych działań, w których udział brałyby jednocześnie kilka instytucji zarządzających i państw członkowskich. Jak wskazano powyżej, problem metodycznej złożoności tych ewaluacji należy rozwiązać w ramach lepszego planowania ewaluacji wpływu opartych na metodach kontrfaktycznych na szczeblu krajowym i międzynarodowym.

Zwiększenie liczby metaewaluacji

W przypadku interwencji w ramach EFS do pomiaru wpływu w drodze ewaluacji wpływu opartych na metodach kontrfaktycznych wykorzystuje się zmienne ilościowe lub binarne (tak/nie), takie jak zarobki lub status zatrudnienia. Ważnymi miarami powodzenia interwencji są jednak „miękkie wyniki” odnoszące się do samooceny i zdolności uczestników. Na przykład warunki wstępne znalezienia zatrudnienia lub aktywności na rynku pracy obejmują „umiejętność przystosowania zawodowego”, samoocenę lub zdobyte kompetencje. W wielu kontekstach analizy miękkich wyników byłyby jeszcze przydatniejsze niż dane dotyczące opuszczenia rynku pracy oraz dałyby obraz dopasowania interwencji do indywidualnych potrzeb.

Uwzględnianie „miękkich wyników”

W odniesieniu do tych miękkich wyników można stosować ewaluacje wpływu oparte na metodach kontrfaktycznych, pod warunkiem że informacje gromadzi się przed interwencją i po jej zakończeniu zarówno w grupie objętej interwencją, jak i w grupie kontrolnej zgodnie z teorią zmiany interwencji. Takie badania należy planować na wczesnym etapie i muszą one obejmować wszystkie istotne podmioty, w szczególności beneficjentów, którzy pozostają w stałym bezpośrednim kontakcie z uczestnikami. Przykład ewaluacji wpływu opartej na metodach kontrfaktycznych dotyczącej „miękkich wyników” przedstawiono w ramce 21 poniżej.

Ramka 21. Przykład oceny wpływu na „miękkie wyniki” w Niemczech

Czy programy tworzenia miejsc pracy zwiększają integrację społeczną i dobrostan długotrwale bezrobotnych?¹⁰⁵

W ramach badania analizuje się wpływ programu tworzenia miejsc pracy skierowanego do grupy szczególnie wrażliwej, tj. osób długotrwale bezrobotnych, które korzystają z pomocy społecznej od co najmniej 4 lat oraz są dotknięte dolegliwościami zdrowotnymi lub mają na utrzymaniu dzieci lub też spełniają obie te przesłanki. Program realizowano w latach 2015–2018 i udział w nim wzięło około 20 000 uczestników. Zastosowany środek polegał na subsydiowaniu przez maksymalny okres 36 miesięcy zwykłych umów o pracę w wymiarze 30 godzin pracy w tygodniu (głównie u pracodawców z sektora publicznego lub w organizacjach charytatywnych).

Mając na uwadze, że ten środek w zakresie subsydiowania zatrudnienia przez maksymalnie 36 miesięcy wyraźnie służył propagowaniu społecznej integracji grup docelowych, autorzy starali się ocenić wpływ w tym zakresie.

Do celów analizy zastosowano obszerny zbiór danych łączący dane administracyjne (niemieckie Zintegrowane Historie Zatrudnienia) z panelowym badaniem wśród uczestników programu i osób w grupie kontrolnej. Przeprowadzenie badania panelowego umożliwiło zgromadzenie informacji na temat subiektywnych miar jakości życia: zadowolenie z życia, zdrowie psychiczne, przynależność społeczna i status społeczny. Informacje na temat tych miar gromadzono z wykorzystaniem skali Likerta spójnych z innymi badaniami (w tym krajowym badaniem PASS), aby uzyskać porównywalne ustalenia. Badanie składało się z następujących głównych etapów: identyfikacja i dopasowanie danych administracyjnych

¹⁰⁵ Zob. [Ivanov B. i in., 2020.](#)

dotyczących osób poddanych interwencji i osób niepoddanych interwencji; przeprowadzenie wywiadów (w trzech falach) wśród członków utworzonych w ten sposób grupy objętej interwencją i grupy kontrolnej; stworzenie ostatecznego zbioru danych, wyczyszczenie danych i usunięcie przypadków, w których brakuje danych. Na próbę składało się 2 531 dopasowanych par w fazie 1; 1 191 w fazie 2 oraz 450 w fazie 3 (spośród próby liczącej łącznie 62 000 osób, z których 12 400 włączono do grupy poddanej interwencji, a 49 600 do grupy kontrolnej).

Stosując technikę *propensity score matching* i mierząc wpływ po upływie 7, 18 i 29 miesięcy od rozpoczęcia udziału w programie, autorzy ustalili, że interwencje miały pozytywny wpływ na miary dobrostanu, choć w różnym stopniu: na przykład zaobserwowano znaczny wzrost zadowolenia z życia, podczas gdy status społeczny poprawił się jedynie w umiarkowanym stopniu. W toku realizacji programów zwykle obserwuje się jednak pogorszenie wpływu. Możliwe, że wynika to z wzrostu liczby uczestników kończących udział w programie i członków grupy kontrolnej znajdujących zatrudnienie. Znaczniejszy pozytywny wpływ odnotowano wśród osób w najtrudniejszej sytuacji.

Jak podkreślono na początku niniejszego przewodnika, ewaluacja wpływu oparta na metodach kontryfaktycznych jest skutecznym narzędziem analitycznym dającym odpowiedź na pytanie: „jakie” efekty miała interwencja w ramach EFS+. Aby poznać odpowiedź na pytanie: „w jaki sposób” lub „dlaczego” dana interwencja miała dany mierzony efekt, należy użyć innych narzędzi.

CIE w połączeniu z TBE

Bywa, że w przypadku dobrze znanych interwencji, które zostały już poddane szczegółowej ewaluacji, wystarczy przeprowadzenie stosunkowo ograniczonej analizy realizacji i kilku wywiadów z beneficjentami, aby zidentyfikować główne elementy stanowiące podstawę mierzonych efektów. W przypadku bardziej złożonych, mniej znanych lub innowacyjnych interwencji, często ewaluację wpływu opartą na metodach kontryfaktycznych trzeba łączyć z innymi rodzajami ewaluacji.

Dzięki ewaluacji wdrożenia (lub procesu) ukończonej przed ewaluacją wpływu opartą na metodach kontryfaktycznych można poznać przebieg interwencji, problemy związane z jej wdrożeniem i wkład poszczególnych podmiotów w ramach współpracy przyczyniający się do pomyślnego ukończenia interwencji. Ustalenia z tej ewaluacji można wykorzystać do stwierdzenia, jakie efekty należy zbadać w ramach ewaluacji wpływu opartej na metodach kontryfaktycznych, jaki powinien być skład grupy kontrolnej, jeżeli chodzi o cechy uczestników, oraz kiedy należy przeprowadzić ewaluację wpływu opartą na metodach kontryfaktycznych.

Równolegle do ewaluacji wpływu opartej na metodach kontryfaktycznych można przeprowadzić ewaluację opartą na teorii (Theory-Based Evaluation, TBE). Te dwa rodzaje ewaluacji mogą wzajemnie się uzupełniać: Ewaluacja oparta na teorii daje obraz kontekstowych, społecznych i indywidualnych mechanizmów, które miały wpływ na efekty i umożliwiły ich osiągnięcie, natomiast ewaluacja wpływu oparta na metodach kontryfaktycznych służy do obliczenia efektów netto i wskazania, w jakim stopniu wyniki interwencji są znaczące. W ramach ewaluacji wpływu zaleca się łączyć ewaluację wpływu opartej na metodach kontryfaktycznych z ewaluacją opartą na teorii, aby uzyskać pełny przegląd efektów i ich przyczyn.

Glosariusze

Akronimy

BAEL	Badanie aktywności ekonomicznej ludności
CIE	Ewaluacja wpływu oparta na metodach kontrfaktycznych
DG EMPL	Dyrekcja Generalna ds. Zatrudnienia, Spraw Społecznych i Włączenia Społecznego
DG REGIO	Dyrekcja Generalna ds. Polityki Regionalnej i Miejskiej
EFRR	Europejski Fundusz Rozwoju Regionalnego
EFS	Europejski Fundusz Społeczny
IEB	Zintegrowane Historie Zatrudnienia
IP	Instytucja(-e) pośrednicząca(-e)
IV	Zmienna instrumentalna
IŻ	Instytucja(-e) zarządzająca(-e)
JLD	Zbiór danych wzdłużnych dotyczących osób poszukujących pracy
KE	Komisja Europejska
MŚP	Małe i średnie przedsiębiorstwa
NGO	Organizacje pozarządowe
PI	Priorytet inwestycyjny
PO	Program(y) operacyjny(-e)
PSM	Metoda <i>propensity score matching</i>
RDD	Model regresji nieciągłej
RODO	Ogólne rozporządzenie o ochronie danych (UE) 2016/679
RWP	Rozporządzenie w sprawie wspólnych przepisów
UE	Unia Europejska
YEI	Inicjatywa na rzecz zatrudnienia ludzi młodych

Definicje

Termin	Definicja
Analiza kontrfaktyczna	Porównanie między sytuacją, która faktycznie zaistniała, a sytuacją, która miałaby miejsce, gdyby nie doszło do interwencji, jeżeli chodzi o rezultaty. Ponieważ różnica między rezultatami rzeczywistymi a rezultatami kontrfaktycznymi stanowi efekt przyczynowy interwencji, analiza kontrfaktyczna obejmuje wszystkie podejścia służące ocenie kwestii, jaką część obserwowanej zmiany można przypisać interwencji podlegającej ewaluacji.
Beneficjent	Zgodnie z art. 2 rozporządzenia (UE) 2021/1060 (rozporządzenie w sprawie wspólnych przepisów) „beneficjent” oznacza „podmiot publiczny lub prywatny, podmiot mający osobowość prawną lub niemający osobowości prawnej lub osobę fizyczną, odpowiedzialne za inicjowanie operacji lub inicjowanie i wdrażanie operacji”. W kontekście instrumentów finansowych (stosowanych czasem na przykład w przypadku mikrokredytów udzielanych osobom samozatrudnionym w kontekście EFS+) beneficjentem jest podmiot, który wdraża fundusz powierniczy, lub, jeżeli nie istnieje struktura z funduszem powierniczym, podmiot wdrażający fundusz szczegółowy, lub, jeżeli instytucja zarządzająca zarządza instrumentem finansowym – instytucję zarządzającą.
Dane podstawowe	Dane dotyczące zmiennych, których pomiar odbywa się przed realizacją interwencji skierowanej do danej jednostki (osoby fizycznej lub przedsiębiorstwa). W wielu przypadkach wskaźniki wyników interwencji mierzone przed interwencją gromadzi się zarówno w przypadku grupy objętej interwencją, jak i grupy kontrolnej.
Dopasowywanie	Jest to metoda polegająca na dopasowywaniu interwencji do prób kontrolnych na podstawie obserwowanych cech.
Efektywność	Efektywność definiuje się jako uzyskanie danego produktu minimalnym kosztem albo, według równoważnej definicji, jako maksymalizacja produktu w przypadku danego poziomu zasobów. Poziom efektywności można ustalić w drodze analizy kosztów i korzyści lub analizy efektywności kosztowej.
Ewaluacja procesu	Ewaluacja procesu dotyczy realizacji programu, w tym między innymi sposobu świadczenia usług, różnic między populacją zamierzoną a populacją poddaną interwencji, dostępu do programu oraz praktyk w zakresie zarządzania.
Ewaluacja wewnętrzna	Ewaluacja przeprowadzona wewnątrz, tj. zlecona bezpośrednio niezależnej instytucji lub jednostce publicznej (instytucji zarządzającej lub instytucji pośredniczącej) bez organizacji procedury przetargowej lub w formie rozszerzonego procesu monitorowania i analizy.
Ewaluacja wpływu oparta na metodach kontrfaktycznych	Rodzaj ewaluacji wpływu, w ramach której podejmuje się próbę identyfikacji efektów przyczynowych interwencji. Taka ewaluacja polega na oszacowaniu średnich rezultatów kontrfaktycznych, a następnie na ich odjęciu od średnich wyników odnotowanych u jednostek poddanych interwencji. Szacunkowe rezultaty kontrfaktyczne zwykle pochodzą z obserwacji grup kontrolnych starannie dobranych tak, aby były jak najbardziej podobne do grupy poddanej interwencji.
Ewaluacja zewnętrzna	Ewaluacja przeprowadzona przez niezależnego ewaluatora zewnętrznego wybranego w drodze procedury przetargowej.
Grupa kontrolna	Grupa osób, przedsiębiorstw lub innych jednostek, które wykazują maksymalne podobieństwo do członków grupy objętej interwencją, ale pozostają niepoddane interwencji. Grupa ta stanowi źródło kontrfaktycznych szacunkowych wartości rezultatów. Ściśle mówiąc, termin „grupa kontrolna” odnosi się do warunków eksperymentalnych (zob. „randomizacja” poniżej), a termin „grupa odniesienia” – do warunków quasi-eksperymentalnych. W praktyce jednak terminy te stosuje się zamiennie.
Grupa objęta interwencją	Grupa osób fizycznych, przedsiębiorstw lub innych jednostek, które korzystają z interwencji lub odczuwają jej wpływ (może to oznaczać propozycję objęcia interwencją lub faktyczne poddanie interwencji).
Interwencje	Interwencje zasadniczo oznaczają operacje realizowane w ramach programów operacyjnych EFS lub projekty współfinansowane w ramach EFS.

Termin	Definicja
Model nielosowy lub quasi-eksperymentalne	<p>Podejścia w zakresie ewaluacji wpływu opartej na metodach kontrfaktycznych, w przypadku których grupy kontrolne tworzy się z wykorzystaniem metod innych niż randomizacja.</p>
Model regresji nieciągłej	<p>Metodę tę można zastosować, gdy dostęp do interwencji określa się za pomocą proggu wyznaczonego dla oceny, skali lub miary o charakterze ciągłym. W podejściu tym wykorzystuje się fakt, że jednostki z wynikiem bezpośrednio sąsiadującym z progiem będą do siebie bardzo podobni, a jedyna różnica będzie polegała na tym, że ci po jednej stronie proggu wezmą udział w interwencji, podczas gdy ci po drugiej stronie nie będą w niej uczestniczyli. Wyniki osób powyżej i poniżej proggu można porównać w celu uzyskania wpływu interwencji.</p>
Plan ewaluacji	<p>Zgodnie z art. 44 rozporządzenia (UE) 2021/1060 w sprawie wspólnych przepisów państwo członkowskie lub instytucja zarządzająca sporządzają plan ewaluacji, który może obejmować więcej niż jeden program i który należy przedłożyć komitetowi monitorującemu nie później niż rok po decyzji zatwierdzającej program.</p>
Podejście oparte na zmiennych instrumentalnych (IV)	<p>W przypadku stosowania tej metody wybór do interwencji powinien być przynajmniej częściowo zdeterminowany przez czynnik zewnętrzny (lub instrument), który nie jest związany z rezultatami innymi niż te, uzyskane w drodze interwencji. W związku z tym czynnik zewnętrzny wpływa na uczestnictwo, ale nie bezpośrednio na rezultaty.</p>
Produkt	<p>Za produkt uznaje się wszystko to, co bezpośrednio wytworzono/dostarczono wskutek realizacji operacji wspieranej z EFS, mierzone w jednostkach fizycznych lub pieniężnych. Główną miarą produktów jest liczba objętych wsparciem osób, podmiotów, dostarczonych towarów i usługi oraz zrealizowanych projektów.</p>
<i>Propensity score matching</i> (PSM)	<p>Metoda PSM polega na oszacowaniu modelu statystycznego dla całej próby (grupa objęta interwencją i potencjalna grupa kontrolna), który to model pozwala oszacować skłonność do uczestnictwa poszczególnych osób lub przedsiębiorstw – niezależnie od tego, czy rzeczywiście biorą udział czy nie. Osoby lub przedsiębiorstwa poddane interwencji są następnie dopasowywane do jednej osoby lub jednego przedsiębiorstwa niepoddanych interwencji albo do kilku takich osób lub przedsiębiorstw na podstawie wartości <i>propensity score</i>.</p>
Randomizacja	<p>Metoda polegająca na losowym przydziale uczestników grupy docelowej do grupy objętej interwencją albo grupy kontrolnej. Randomizacja stanowi sposób na zapewnienie, aby grupy były statystycznie równoważne we wszystkich aspektach w momencie, w którym poddaje się je randomizacji.</p>
Rezultat	<p>Wpływ operacji na uczestników lub podmioty, na przykład na ich sytuację w kontekście zatrudnienia, zarobki, wyniki osiągane w standardowych testach stosowanych w edukacji, zyski itp. Wpływ ten można zmierzyć w okresie krótko- lub długoterminowym. W kontekście EFS+, mierząc wskaźniki rezultatów, za okres krótkoterminowy przyjmuje się okres bezpośrednio po zakończeniu uczestnictwa (cztery tygodnie), a za okres długoterminowy – sześć miesięcy po zakończeniu udziału¹⁰⁶. W kontekście analizy wpływu okres długoterminowy często jest jednak dłuższy i oznacza co najmniej 24 miesiące od zakończenia interwencji.</p>
Różnica w różnicach	<p>Jest to najprostszy sposób, w jaki różnicę w wynikach przed interwencją i po jej zakończeniu w grupie kontrolnej odejmuje się od tej samej różnicy zaobserwowanej w grupie poddanej interwencji, aby uzyskać szacunkowy wpływ danej interwencji. Wpływ obliczony na podstawie różnicy w różnicach zwykle obserwuje się w ramach regresji.</p>
Skuteczność	<p>Skuteczność oznacza „osiągnięcie celów”. Ewaluacja skuteczności polega na porównaniu wyników uzyskanych z wynikami planowanymi (lub z sytuacją odniesienia) lub na porównaniu sytuacji obserwowanej po przeprowadzeniu działania z sytuacją, która miałyby miejsce bez podjęcia danego działania (sytuacja kontrfaktyczna).</p>
Trafność	<p>Trafność oznacza adekwatność wyraźnych celów interwencji względem problemów społeczno-gospodarczych, które dana interwencja ma rozwiązać¹⁰⁷.</p>
Uczestnicy	<p>Uczestnik oznacza osobę fizyczną, która bezpośrednio korzysta z operacji, ale nie odpowiada ani za rozpoczęcie, ani za realizację takiej operacji. W kontekście EFS</p>

¹⁰⁶ Zob. [Komisja Europejska, 2021a](#) – i [Komisja Europejska, 2018](#).

¹⁰⁷ Zob. [Komisja Europejska, 2013](#).

Termin	Definicja
	uczestnikami są osoby otrzymujące wsparcie w drodze interwencji w ramach EFS.
Wpływ	W kontekście ewaluacji wpływu opartej na metodach kontrfaktycznych wpływ oznacza efekt netto definiowany jako różnica między średnimi rezultatami interwencji a średnimi rezultatami kontrfaktycznymi. Do celów niniejszego przewodnika terminy „wpływ” i „efekt netto” stosuje się zamiennie.
Wynik	Prawdopodobny lub osiągnięty krótko- i średnioterminowy wpływ produktów interwencji ¹⁰⁸ . Wynik przypomina „rezultat”, ale termin ten stosuje się częściej w przypadku ewaluacji wpływu.

¹⁰⁸ Zob. [OECD, 2010](#).

Bibliografia

- Ashenfelter, O. (1978). Estimating the effect of training programmes on earnings, *Review of Economics and Statistics*, 6, s. 47–57. <https://www.jstor.org/stable/1924332?origin=crossref>
- Axdorph E, Egebark J., Lundström T., Özcan G. (2019) Effekter av förstärkta förmedlingsinsatser för unga arbets sökande – Resultat från utvärderingen av Ung framtid. Arbetsförmedlingen
- Baran J. i in. (2016). Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. I Raport Wskaźnikowy. <http://files.evaluationhelpdesk.eu/Evaluations/PLE49.pdf>
- Baran J. i in. (2017). Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój II. Raport wskaźnikowy. <http://files.evaluationhelpdesk.eu/Evaluations/PLE196.pdf>
- Baran J. i in. (2018). Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój III. Raport wskaźnikowy. <http://files.evaluationhelpdesk.eu/Evaluations/PLE197.pdf>
- Baran J. i in. (2018a). Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. II Raport Tematyczny. <http://files.evaluationhelpdesk.eu/Evaluations/PLE221.pdf>
- Bazzoli M., De Poli S., Rettore E., Schizzerotto A. (2018). Are Vocational Training Programmes Worth Their Cost? Evidence from a Cost-Benefit Analysis. w *Politica Economica/Journal of Economic Policy*, t. XXXIV(3), s. 215–240. <https://www.rivisteweb.it/doi/10.1429/92119>
- Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs, *Evaluation Review*, 8(2), s. 225–246. <https://cpb-us-e2.wpmucdn.com/sites.uci.edu/dist/1/1159/files/2021/03/Bloom-MDES-Eval-Rev-1995-Bloom.pdf>
- Bloom, H. S. (2009) Modern regression discontinuity analysis, MDRC Working Papers on Research Methodology, Nowy Jork: MDRC. <https://www.mdrc.org/publication/modern-regression-discontinuity-analysis>
- Boockmann B. i in. (2017). Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales. <http://files.evaluationhelpdesk.eu/Evaluations/DEE189.pdf>
- Boockmann B. i in. (2018). Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales. <http://files.evaluationhelpdesk.eu/Evaluations/DEE52.pdf>
- Boockmann B. i in. (2019). Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales. <http://files.evaluationhelpdesk.eu/Evaluations/DEE94.pdf>
- Boockmann B. i in. (2021). Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales. <http://files.evaluationhelpdesk.eu/Evaluations/DEE202.pdf>
- Borik V. i in. (2015). The net effects of graduate work experience and the promotion of self-employment. Technical Report. <http://files.evaluationhelpdesk.eu/Evaluations/SKE8.pdf>

- Bratti M. i in. (2018). Vocational Training for Unemployed Youth in Latvia: Evidence from a Regression Discontinuity Design IZA DP nr 11870. <http://ftp.iza.org/dp11870.pdf>
- Bratu C. i in. (2014). Knowledge gaps in evaluating labour market and social inclusion policies. W dokumencie wskazano 39 ewaluacji wpływu opartych na metodach kontrfaktycznych EFS w okresie programowania 2007–2013. doi:10.2788/083390. <https://op.europa.eu/en/publication-detail/-/publication/d0e73217-2a40-47af-a5c3-a700dcd47da3/language-en>
- Bredgaard, T. (2015). Evaluating what Works for whom in active Labour market policies. *European Journal of Social Security*, 17(4), s. 436–452. <https://doi.org/10.1177/138826271501700403>
- Bryson, A., Dorsett, R. i Purdon, S. (2002), The use of propensity score matching in the evaluation of active labour market policies, Department for Work and Pensions, Working Paper Number 4. [http://eprints.lse.ac.uk/4993/1/The use of propensity score matching in the evaluation of active labour market policies.pdf](http://eprints.lse.ac.uk/4993/1/The_use_of_propensity_score_matching_in_the_evaluation_of_active_labour_market_policies.pdf)
- Caliendo M., Kopeinig S., (2008), Some practical Guidance for the implementation of propensity score matching. *Journal of Economic Surveys*; t. 22, wyd. 1 – s. 31–72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Caliendo, M., Mahlstedt, R., & Mitnik, O. A. (2017). Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labour market policies. *Labour Economics*, 46, s. 14–25. <https://doi.org/10.1016/j.labeco.2017.02.001>
- Card, D., Ibarra, P. i Villa, J. M. (2011). Building in an evaluation component for active labour market programs: a practitioner's guide, Discussion Paper No. 6085, Bonn, Niemcy: IZA. <http://ftp.iza.org/dp6085.pdf>
- Card, D., Kluve, J. i Weber, A. (2017) What works? A meta-analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, t. 16, wyd. trzecie, czerwiec 2018 r., s. 894–931, <https://doi.org/10.1093/jeaa/jvx028>
- Centre for Disease Control and Prevention CDC (2013). Good Evaluation Questions: Checklist to Help Focus Your Evaluation. Departament Zdrowia i Usług Społecznych Stanów Zjednoczonych. http://www.cdc.gov/asthma/program_eval/AssessingEvaluationQuestionChecklist.pdf
- Danmarks Statistik (2017). Effektmåling og monitorering 2016 - strukturfondsindsatsen i Syddanmark. <http://files.evaluationhelpdesk.eu/Evaluations/DKE7.pdf>
- Danmarks Statistik (2018). Effektmåling af den Virksomhedsrettede Strukturfondsindsats 2007-2013. <http://files.evaluationhelpdesk.eu/Evaluations/DKE10.pdf>
- Diaz J. M. i in. (2019). Evaluación del PO-FSE 2014/2020 del Principado de Asturias para el informe anual a presentar en 2019. <http://files.evaluationhelpdesk.eu/Evaluations/ESE101.pdf>
- Ecorys, Ismeri Europa (2020) Study supporting the evaluation of ESF support to education and training (Thematic Objective 10), dla Komisji, DG ds. Zatrudnienia, Spraw Społecznych i Włączenia Społecznego. <https://op.europa.eu/en/publication-detail/-/publication/d0c1a558-077d-11eb-a511-01aa75ed71a1/language-en/format-PDF/source-173162502>
- Fondazione G. Brodolini, Metis GmbH, Applica, Ockham IPS (2020). Study for the Evaluation of ESF Support to Employment and Labour Mobility (cel tematyczny 8), dla Komisji Europejskiej, DG ds. Zatrudnienia, Spraw Społecznych i Włączenia Społecznego. <https://ec.europa.eu/social/BlobServlet?docId=22899&langId=en>
- Frolich, M (2004) Programme evaluation with multiple treatments, *Journal of Economic Surveys*, 18(2), s. 181–224. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.620.9209&rep=rep1&type=pdf>

- Guthmuller S., Meroni E.C., Evaluation of the higher education grant system for less privileged students in Portugal. sprawozdanie techniczne JRC. Dotychczas nieopublikowane.
- Hagglund, P. (2006). A description of three randomised experiments in Swedish labour market policy, Institute for Labour Market Policy Evaluation, sprawozdanie 2006:4.
<https://www.ifau.se/globalassets/pdf/se/2006/r06-04.pdf>
- Heckman, J. J., Ichimura, H., Smith, J. i Todd, P. (1998) Characterizing selection bias using experimental data. *Econometrica*, t. 66, nr 5. s. 1017–1098.
<http://jenni.uchicago.edu/papers/Heckman Ichimura etal 1998 Econometrica v66 n5 r.pdf>
- Holland, P. W. (1986). Statistics and Causal Inference, *Journal of the American Statistical Association*, 81 (396), 945-960 DOI: 10.1080/01621459.1986.10478354.
<http://people.umass.edu/~stanek/pdffiles/causal-holland.pdf>
https://ec.europa.eu/regional_policy/archive/conferences/evaluation2009/abstracts/martini.doc
- Hunger K. i Sattler K. (2017). Evaluationsbericht zum spezifischen Ziel A 1.1. im Rahmen der Evaluierung des Operationellen Programms des Europäischen Sozialfonds in Baden-Württemberg 2014–2020. <http://files.evaluationhelpdesk.eu/Evaluations/DEE26.pdf>
- IAW Institut für Angewandte Wirtschaftsforschung _ ISG Institut für Sozialforschung und Gesellschaftspolitik, GmbH (2015). Evaluation der Modellprojekte "Bürgerarbeit". Endbericht.
<http://files.evaluationhelpdesk.eu/Evaluations/DEE7.pdf>
- ICF, Cambridge Econometrics and Eurocentre (2020) Study supporting the 2020 evaluation of promoting social inclusion, combating poverty and any discrimination by the European Social Fund (Thematic Objective 9) dla Komisji Europejskiej, DG ds. Zatrudnienia, Spraw Społecznych i Włączenia Społecznego. <https://op.europa.eu/en/publication-detail/-/publication/8788ec85-2308-11eb-b57e-01aa75ed71a1>
- Indecon (2016). Indecon Evaluation of JobBridge Activation Programme.
<http://files.evaluationhelpdesk.eu/Evaluations/IEE2.pdf>
- Instytut Badań Strukturalnych, Imapp, IQS. (2015). Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. I Raport Tematyczny. <http://files.evaluationhelpdesk.eu/Evaluations/PLE34.pdf>
- Ires Piemonte. (2019). Misure di sostegno all'imprenditorialità.
<http://files.evaluationhelpdesk.eu/Evaluations/ITE61.pdf>
- Isfol (2016). Primo rapporto di valutazione del piano italiano Garanzia giovani.
<http://files.evaluationhelpdesk.eu/Evaluations/ITE13.pdf>
- Ismeri Europa – Ecorys – Institute for Employment Studies. (2019). Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations. doi:10.2767/39339. <https://op.europa.eu/it/publication-detail/-/publication/84cc9eb9-b33d-11e9-9d01-01aa75ed71a1>
- Ismeri Europa (2018). I Rapporto tematico di valutazione – I risultati di PIPOL.
<http://files.evaluationhelpdesk.eu/Evaluations/ITE36.pdf>
- Ivanov B. i in., 2020, Do Job creation schemes improve the social integration and well-being of the long-term unemployed?, *Labour Economics*. <https://doi.org/10.1016/j.labeco.2020.101836>.
- Kalinowski H. (2020). Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. V Raport wskaźnikowy.
<http://files.evaluationhelpdesk.eu/Evaluations/PLE296.pdf>
- Kalinowski H. i in., 2020, Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. Raport końcowy.
<http://files.evaluationhelpdesk.eu/Evaluations/PLE339.pdf>

Kluve, J., Lehmann, H., & Schmidt, C. M. (2008). Disentangling treatment effects of Active Labor Market Policies: The role of labor force status sequences. *Labour Economics*, 15(6), s. 1270–1295. <https://doi.org/10.1016/j.labeco.2007.12.002>

Komisja Europejska (2007): Indicative Guidelines on evaluation methods: evaluation during the programming period. Dokument roboczy nr 5. DG ds. Polityki Regionalnej. https://ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd5_ongoing_en.pdf

Komisja Europejska (2013): EVALSED: The resource for the evaluation of Socio-Economic Development. Wersja zaktualizowana. https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf

Komisja Europejska (2017), Zestaw instrumentów służących lepszemu stanowienu prawa, Narzędzie nr 47 *Kryteria ewaluacji i pytania ewaluacyjne*, https://ec.europa.eu/info/sites/default/files/file_import/better-regulation-toolbox-47_en_0.pdf

Komisja Europejska (2018) Monitoring and Evaluation of European Cohesion Policy European Social Fund. Wytyczne. <https://ec.europa.eu/sfc/en/system/files/ged/ESF%20monitoring%20and%20evaluation%20guidance.pdf>

Komisja Europejska (2019) Advanced counterfactual evaluation methods. Wytyczne. doi:10.2767/464242. <https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8254&furtherPubs=yes>

Komisja Europejska (2020) Counterfactual impact evaluation of European Social Fund interventions in practice. Wytyczne dla instytucji zarządzających. doi:10,2767/55495. <https://op.europa.eu/en/publication-detail/-/publication/f82c5fb8-bb40-11ea-811c-01aa75ed71a1/language-en>

Komisja Europejska (2020) How to use administrative data for European Social Funds counterfactual impact evaluations. Przewodnik dla instytucji zarządzających. <https://op.europa.eu/en/publication-detail/-/publication/d96feed3-f30c-11ea-991b-01aa75ed71a1>

Komisja Europejska (2021), rozporządzenie Parlamentu Europejskiego i Rady (UE) 2021/1060 z dnia 24 czerwca 2021 r. ustanawiające wspólne przepisy dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego Plus, Funduszu Spójności, Funduszu na rzecz Sprawiedliwej Transformacji i Europejskiego Funduszu Morskiego, Rybackiego i Akwakultury, a także przepisy finansowe na potrzeby tych funduszy oraz na potrzeby Funduszu Azylu, Migracji i Integracji, Funduszu Bezpieczeństwa Wewnętrznego i Instrumentu Wsparcia Finansowego na rzecz Zarządzania Granicami i Polityki Wizowej

Komisja Europejska (2021a): Common Indicators Toolbox. Dokument roboczy. czerwiec 2021. https://ec.europa.eu/sfc/en/system/files/2021/ged/toolbox_june_2021_final.pdf

Krug, G and Stephan, G. (2011) Is contracting-out intensified placement services more effective than in-house production? Evidence from a randomized field experiment, LASER Discussion Papers – dokument nr 5, <http://doku.iab.de/externe/2011/k110912303.pdf>

Lammers M. i Kok L. (2021). Are active labor market policies (cost-) effective in the long run? Evidence from the Netherlands. W *Empirical Economics* 60:1719–1746. <https://doi.org/10.1007/s00181-019-01812-3>

Lang, J. (2021) Employment effects of language training for unemployed immigrants. *J Popul Econ* (2021). <https://doi.org/10.1007/s00148-021-00832-7>

Martini, A. (2009). Counterfactual impact evaluation: what it can (and cannot) do for cohesion policy, dokument przygotowany na 6. Europejską Konferencję w sprawie Oceny Polityki Spójności, Warszawa, dnia 30 listopada.

Morgan, S. L. and Winship, C. (2014) Counterfactual and causal inference: Methods and

- principles for social research. Wyd. drugie. Cambridge i Nowy Jork: Cambridge University Press.
- OECD (2010), Glossary of Key Terms in Evaluation and Results Based Management, Paryż. <https://www.oecd.org/dac/evaluation/2754804.pdf>
- Openfield (2019). Analiza skuteczności i efektywności dotacji na założenie działalności gospodarczej udzielonych w ramach 8 osi priorytetowej RPO WM – komponent 3. <http://files.evaluationhelpdesk.eu/Evaluations/PLE312.pdf>
- Palczyńska M. i in. (2019) Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. IV Raport wskaźnikowy. <http://files.evaluationhelpdesk.eu/Evaluations/PLE274.pdf>
- Pomatto G. (2017). L'attuazione del Buono per Servizi al Lavoro nella Regione Piemonte. Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE55.pdf>
- Pomatto G. (2019). Buoni per servizi al lavoro nella Regione Piemonte: qualità percepita dai destinatari e meccanismi dell'attuazione. Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE77.pdf>
- Pompili M., Giorgetti I. (2020). Rapporto di Placement. Servizio di attività di valutazione del POR FSE 2014/2020 – Regione Marche. <http://files.evaluationhelpdesk.eu/Evaluations/ITE115.pdf>
- Pompili M., Giorgetti I. (2020a). Rapporto tematico „Disoccupazione di lunga durata”. Servizio di attività di valutazione del POR FSE 2014/2020 – Regione Marche. <http://files.evaluationhelpdesk.eu/Evaluations/ITE214.pdf>
- Poy S (2020). Nuove evidenze sull'effetto occupazionale del buono per servizi al lavoro. Target persone disoccupate da almeno 6 mesi. Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE245.pdf>
- Poy S. (2019). Gli effetti occupazionali del buono per servizi al lavoro nella Regione Piemonte: prime evidenze. Misura per disoccupati da almeno 6 mesi, anno 2017. Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE123.pdf>
- Re-source Pracownia Badań i Doradztwa (2015). Ocena działań na rzecz wydłużenia aktywności zawodowej osób starszych w kontekście niekorzystnej sytuacji demograficznej w województwie lubelskim. <http://files.evaluationhelpdesk.eu/evaluations/PLE126.pdf>
- Riccio J., Friedlander, D., Freedman S. (1994). GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program, MDRC, NYC <https://www.mdrc.org/publication/gain-benefits-costs-and-three-year-impacts-welfare-work-program>
- Scheller F. i Seidel K. (2020). Zweiter Evaluationsbericht zum spezifischen Ziel A 1.1: Teilnehmer*innenperspektive und Wirkungsanalyse. im Rahmen der Evaluierung des Operationellen Programms des Europäischen Sozialfonds in Baden-Württemberg 2014–2020. <http://files.evaluationhelpdesk.eu/Evaluations/DEE171.pdf>
- Shadish, W. R., Cook, T. D. i Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalised causal inference, Boston, Stany Zjednoczone: Houghton Mifflin Company.
- Stern E. i in. (2012). Broadening the range of designs and methods for impact evaluations, Department for International Development of the UK, Working Paper 38. <https://assets.publishing.service.gov.uk/media/57a08a6740f0b6497400059e/DFIDWorkingPaper38.pdf>
- WK Kellogg Foundation (2004). Logic Model Development Guide. <https://www.wkcf.org/resource-directory/resources/2004/01/logic-model-development-guide>
- Zub M. i in. (2020). Ocena wpływu wsparcia RPOWP na popularyzację szkolnictwa zawodowego w województwie podlaskim. Raport końcowy.

<http://files.evaluationhelpdesk.eu/Evaluations/PLE359.pdf>

Załączniki

Załącznik 1. Dodatkowe informacje

Pracownikom instytucji zarządzających chcącym poszerzyć swoją wiedzę w zakresie kwestii podjętych w niniejszym przewodniku polecamy następujące pozycje. W literaturze istnieje wiele opracowań poświęconych zagadnieniu ewaluacji. Poniżej wymieniono wiarygodne, istotne omówienia zawierające informacje, które można bezpośrednio wykorzystać przy planowaniu ewaluacji wpływu opartej na metodach kontrfaktycznych. Po każdym opisie bibliograficznym podano krótki opis większości źródeł.

Ogólna ewaluacja

- Rossi, Peter H., Mark W. Lipsey, i Gary T. Henry. (2018). *Evaluation: A Systematic Approach*. Wyd. ósme. Thousand Oaks, CA: SAGE Publications.

Klasyczny podręcznik na temat praktyk i metod ewaluacyjnych. Zawiera opisy metod i przykłady.

- Gertler, Paul J.; Martinez, Sebastian; Premand, Patrick; Rawlings, Laura B.; Vermeersch, Christel M. J. (2016). *Impact Evaluation in Practice*, wyd. drugie. Waszyngton D.C.: Inter-American Development Bank i Bank Światowy. © World Bank.
<https://openknowledge.worldbank.org/handle/10986/25030> (dokument dostępny w języku angielskim, portugalskim i hiszpańskim)

Podobnie jak w niniejszym przewodniku w podręczniku tym najpierw omówiono klasyczny rodzaj ewaluacji (badanie randomizowane z grupą kontrolną), a następnie metody alternatywne. Chociaż podręcznik jest skierowany do zarządzających programami w państwach o niskich dochodach, omówienie jest istotne i ma bezpośrednie zastosowanie w kontekście państw członkowskich UE.

- White, H., & Raitzer, D. A. (2017). *Impact evaluation of development interventions: A practical guide*. Asian Development Bank.
<https://www.adb.org/sites/default/files/publication/392376/impact-evaluation-development-interventions-guide.pdf>

Podobnie jak w opracowaniu Gertler i in. (2016) wymienionym powyżej w podręczniku tym – który jest poświęcony tematyce ewaluacji wpływu interwencji w kraj o niski i średnim dochodzie – w sposób kompleksowy omówiono ogólne metody ewaluacji wpływu opartej na metodach kontrfaktycznych. W porównaniu z opracowaniem Gertler i in. (2016) jest to bardziej techniczny tekst, w którym można jednak znaleźć wiele praktycznych uwag, które mogą przydać się zarządzającym programami EFS+.

- Komisja Europejska (2019) *Advanced counterfactual evaluation methods*. Wytoczne. doi:10.2767/464242. <https://op.europa.eu/en/publication-detail/-/publication/11968bbb-fac9-11e9-8c1f-01aa75ed71a1/language-en>

W dokumencie tym przedstawiono nowe i bardziej zaawansowane metody ewaluacji wpływu opartej na metodach kontrfaktycznych, takie jak analiza sekwencji, dopasowywanie dynamiczne i kontrole syntetyczne.

- Csillag Marton, Kreko Judi i Scharle Agota. (2020). *Ewaluacja kontrfaktyczna programów zatrudnienia osób młodych*. Sporządzono w ramach projektu „Partnerstwo na rzecz zatrudnienia młodzieży”. <https://ibs.org.pl/app/uploads/2021/03/Przewodnik-metodologiczny.pdf> (dokument dostępny w języku angielskim, hiszpańskim, węgierskim (tłumaczenie w przygotowaniu), włoskim i polskim).

Jest to stopniowy wstęp do ewaluacji polityk rynku pracy skierowanych do osób młodych. Skupiono się na ewaluacjach kontryfakcyjnych, które wykorzystują dane administracyjne. Przedstawione kwestie oparto na praktycznych problemach napotkanych w ewaluacjach dotacji na zatrudnianie młodzieży w czterech krajach: w Hiszpanii, na Węgrzech, we Włoszech i w Polsce.

- HM Treasury (Wielka Brytania) (2020). The Magenta Book: Guidance for evaluation. Londyn: HM Treasury. <https://www.gov.uk/government/publications/the-magenta-book>

Opracowanie to zawiera szczegółowy opis metodyki ewaluacji. Dokumenty te stanowią interesujące przykłady ukazujące perspektywę wewnętrzną ewaluacji rządowej.

Badania randomizowane z grupą kontrolną

- White H. (2013) An introduction to the use of randomised control trials to evaluate development interventions, *Journal of Development Effectiveness*, 5:1, s. 30–49, DOI: 10.1080/19439342.2013.764652
<https://www.tandfonline.com/doi/pdf/10.1080/19439342.2013.764652>
- White, H., Sabarwal S. & T. de Hoop, (2014). Randomized Controlled Trials (RCTs), *Methodological Briefs: Impact Evaluation 7*, UNICEF Office of Research, Florencia.
https://www.unicef-irc.org/publications/pdf/brief_7_randomized_controlled_trials_eng.pdf

Dwa teksty stanowiące nietechniczne wprowadzenie na temat logiki badania randomizowanego z grupą kontrolną zawierające omówienie kilku modeli i krytyczne uwagi dotyczące tego badania.

- Glennerster, R., & Takavarasha, K., (2013). *Running Randomized Evaluations: A Practical Guide*, Princeton University Press, Princeton, NJ.

Opracowanie to stanowi przewodnik, w którym wyjaśnia się sposób opracowania i prawidłowy przebieg badania randomizowanego z grupą kontrolną w zakresie programów społecznych. Omówienia dokonuje się na podstawie konkretnych badań randomizowanych z grupą kontrolną przeprowadzonych przez Abdul Latif Jameel Poverty Action Lab.

Różnica w różnicach

- Card, David, Pablo Ibararán i Juan Miguel Villa. (2011). Building in an Evaluation Component for Active Labor Market Programs: A Practitioner's Guide. IZA Discussion Paper No. 6085. Bonn: Forschungsinstitut zur Zukunft der Arbeit. <http://ftp.iza.org/dp6085.pdf>
- Lechner, M. (2011) The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends in Econometrics*. t. 4, nr 3 (2010), s. 165–224 DOI: 10.1561/08000000014. https://michael-lechner.eu/ml_pdf/journals/2011_Lechner_DiD_2011_ECO%20403%20Lechner_darf%20aufsatz%20Netz.pdf

Opracowanie to zawiera szczegółowe omówienie podejścia „różnicy w różnicach”, a także niektórych głównych kwestii związanych ze stosowaniem tego podejścia. Przedstawiono w nim również rozszerzone zastosowania tego podejścia, takie jak zastosowania nieliniowe i metoda *propensity score matching* w połączeniu z podejściem „różnicy w różnicach”.

- Fredriksson A. i de Oliveira G. M. (2019) Impact evaluation using Difference-in-Differences. *RAUSP Management Journal* t. 54, nr 4, s. 519–532. DOI 10.1108/RAUSP-05-2019-0112
<https://www.emerald.com/insight/content/doi/10.1108/RAUSP-05-2019-0112/full/pdf?title=impact-evaluation-using-difference-in-differences>

Przegląd metod różnicy w różnicach wraz z praktycznymi zaleceniami.

- Card, David i Alan B. Krueger. (1994). *Minimum Wages and Employment: A Case Study of*

the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84 (4), s. 774–775. <https://davidcard.berkeley.edu/papers/njmin-aer.pdf>

Przykład klasycznego zastosowania techniki różnicy w różnicach.

Zmienne instrumentalne

- Morgan, Stephen L. i Christopher Winship. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Wyd. drugie. Cambridge i Nowy Jork: Cambridge University Press.

Jest to raczej techniczny przegląd metod ewaluacji wpływu opartej na metodach kontrfaktycznych, w którym używa się terminologii socjologicznej. Rozdział 9 „Instrumental Variable Estimators of Causal Effects” [„Estymatory zmiennych instrumentalnych efektu przyczynowego”] (s. 291–324) zawiera przegląd logiki i procedur oszacowania zmiennych instrumentalnych.

- Kuhn, Andreas, Jean-Philippe Wuehrich i Josef Zweimüller. 2010. *Fatal Attraction? Access to Early Retirement and Mortality*. IZA Discussion Paper No.5160. Bonn: Forschungsinstitut zur Zukunft der Arbeit. <http://ftp.iza.org/dp5160.pdf>

Badanie wpływu wczesnego przechodzenia na emeryturę na zdrowie pracownika, w którym jako zmienną instrumentalną przyjęto regionalne zróżnicowanie zmiany wieku emerytalnego w Austrii.

- Galiani, S., Rossi, M. A. i Schargrotsky, E. (2011). *Conscription and crime: Evidence from the Argentine draft lottery*. *American Economic Journal: Applied Economics*. <https://www.aeaweb.org/articles?id=10.1257/app.3.2.119>

Innowacyjne i obrazowe zastosowanie randomizowanej loterii poborowej (służącej powołaniu do służby wojskowej) jako zmiennej instrumentalnej. Bardzo dobrze napisane opracowanie, które można służyć jako wzór dla badań, w których jako narzędzia przyjmuje się procedury loterii.

Dopasowywanie

- Heinrich, Carolyn, Alessandro Maffioli i Gonzalo Vázquez. 2010. *A Primer for Applying Propensity Score Matching*. Impact-Evaluation Guidelines Technical Notes No. IDB-TN-161. Waszyngton: Inter-American Development Bank. <https://publications.iadb.org/publications/english/document/A-Primer-for-Applying-Propensity-Score-Matching.pdf>

Podobnie jak niżej wskazany przewodnik dotyczący regresji nieciągłej, opracowanie to jest skierowane do mających szeroką wiedzę zarządzających ewaluacją.

- Caliendo M., Kopeinig S., (2008), *Some practical Guidance for the implementation of propensity score matching*. *Journal of Economic Surveys*; t. 22, wyd. pierwsze – s. 31–72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>

Klasyczny tekst na temat wyborów, których należy dokonać, stosując technikę PSM (jeżeli chodzi o oszacowanie, algorytmy dopasowywania, oceny jakości dopasowywania, wrażliwości oszacowanych efektów interwencji).

- Harris, H. i Horst, S. J. (2016) *A Brief Guide to Decisions at Each Step of the Propensity Score Matching Process, Practical Assessment, Research, and Evaluation*: t. 21, artykuł 4. DOI: <https://doi.org/10.7275/yq7r-4820>. Dokument dostępny pod adresem: <https://scholarworks.umass.edu/pare/vol21/iss1/4>.

Podobny do poprzedniego artykułu.

Model regresji nieciągłej

- Jacob, Robin, Pei Zhu, Marie-Andrée Somers i Howard Bloom. (2012). *A Practical Guide to*

Regression Discontinuity. Nowy Jork: MDRC.

https://www.mdrc.org/sites/default/files/RDD%20Guide_Full%20rev%202016_0.pdf

Wyjątkowo przystępne i szczegółowe omówienie metodyki regresji nieciągłej obejmujące starannie wyselekcjonowaną bibliografię.

- Lee, D. S. i Lemieux T. (2009), Regression discontinuity designs in economics, dokument roboczy NBER nr 14723, National Bureau of Economic Research, Cambridge, MA, 2009.
http://www.nber.org/papers/w14723.pdf?new_window=1

Opracowanie to swojego rodzaju podręcznik dla użytkownika na temat regresji nieciągłej zawierający opis logiki metody oraz różnych sposobów oszacowania modeli regresji nieciągłej.

Załącznik 2. Sugerowany program kursu na temat ewaluacji wpływu opartej na metodach kontrfaktycznych

Kurs wprowadzający poświęcony ewaluacji wpływu opartej na metodach kontrfaktycznych dla instytucji zarządzających i urzędników może obejmować następujące punkty:

- Wprowadzenie do podejść ewaluacyjnych: ewaluacje procesu (dlaczego są ważne i jaki mają związek z ewaluacją wpływu opartą na metodach kontrfaktycznych) i ewaluacje wpływu (podejścia kontrfaktyczne i oparte na teorii).
- Czym są ewaluacje wpływu oparte na metodach kontrfaktycznych? Na jakie pytania ewaluacyjne można uzyskać odpowiedź w ramach skutecznych ewaluacji wpływu opartych na metodach kontrfaktycznych? W przypadku których kryteriów ewaluacji główne dowody uzyskuje się w drodze ewaluacji wpływu opartej na metodach kontrfaktycznych? Na czym polega znaczenie ewaluacji wpływu opartych na metodach kontrfaktycznych?
- Na czym polegają takie ewaluacje? (grupa poddana interwencji i grupa kontrolna, porównanie „przed i po”, zasady wnioskowania).
- Przegląd metodyk ewaluacji wpływu opartej na metodach kontrfaktycznych (charakterystyka, wady i zalety każdej metody):
 - o badanie randomizowane z grupą kontrolną,
 - o dopasowywanie,
 - o różnica w różnicach,
 - o model regresji nieciągłej,
 - o zmienna instrumentalna,
 - o wskazanie innych ewentualnych metod (np. technik analizy przestrzennej).
- Wymogi dotyczące danych
 - o wymogi dotyczące grupy poddanej interwencji i grupy kontrolnej,
 - o potencjalne źródła (monitorowanie EFS, zbiory danych administracyjnych, badania),
 - o przepisy RODO.
- Przegląd etapów realizacji:
 - o planowanie ewaluacji wpływu opartych na metodach kontrfaktycznych (plan ewaluacji, wykonalność ewaluacji wpływu opartej na metodach kontrfaktycznych),
 - o zlecenie ewaluacji wpływu opartych na metodach kontrfaktycznych (specyfikacja istotnych warunków zamówienia, kryteria kwalifikacji i metody),
 - o zarządzanie ewaluacjami wpływu opartymi na metodach kontrfaktycznych (weryfikacja wyników i kontakty z ewaluatorem),
 - o rozpowszechnienie ustaleń z ewaluacji wpływu opartych na metodach kontrfaktycznych (rodzaje odbiorców i kanały komunikacji).

Kurs zorganizowany w ten sposób powinien obejmować analizę praktycznych przykładów w ramach dwu- lub trzydniowych warsztatów. Realizując tego typu kurs, należałoby przyjąć podejście nauczania metodą rozwiązywania problemów (problem-based learning, PBL) i jako analizowany przypadek zastosować środek z zakresu polityki wprowadzony w ramach danego programu.

Załącznik 3. Ewaluacja wpływu oparta na metodach kontrfaktycznych – przykłady podane w przewodniku

Tabela 6. Charakterystyka CIE na podstawie przykładów podanych w niniejszym przewodniku

Kraj	Tytuł – rok	Polityka poddana ewaluacji i badany okres interwencji	Metoda	Grupa poddana interwencji i grupa kontrolna – definicja i wielkość	Zmienne wynikowe	Okres obserwacji	Data
Szwecja	Efekty zintensyfikowanego wsparcia dla młodych osób poszukujących pracy – wyniki ewaluacji projektu „Młoda przyszłość” (2019 r.)	Zintensyfikowane wsparcie ze strony służb zatrudnienia dla osób poszukujących pracy Czerwiec 2017 r. – styczeń 2018 r.	Badanie randomizowane z grupą kontrolną	Grupa poddana interwencji: Osoby młode w wieku 18–24 lat losowo przydzielane do grupy objętej zintensyfikowanym wsparciem Grupa kontrolna: Osoby młode w wieku 18–24 lat losowo przydzielane do grupy objętej standardowym wsparciem Grupa poddana interwencji: 2 792 – grupa kontrolna: 1 897	– odsetek osób bezrobotnych – odsetek osób wychodzących z bezrobocia – średnia liczba dni na bezrobociu	po 12 miesiącach od rozpoczęcia udziału w interwencji	– Publiczne rejestry zatrudnienia na potrzeby zmiennych wyników – Badania wśród pracowników urzędów pracy i pośredników pracy w celu zmierzenia intensywności udzielonego wsparcia
Włochy (Marche)*	Wpływ interwencji w ramach EFS finansowanych w latach 2014–2020 na osoby długotrwale bezrobotne w regionie Marche (2020 r.)	Staże, krótkie praktyki zawodowe, praktyki zawodowe i bony na szkolenia dla bezrobotnych (w tym długotrwale bezrobotnych) 2017 r. – lipiec 2019 r.	Propensity score matching (PSM)	Grupa poddana interwencji: Osoby długotrwale bezrobotne poddane interwencji Grupa kontrolna: Osoby zarejestrowane przez publiczne służby zatrudnienia jako osoby długotrwale bezrobotne w latach 2016–2018 Grupa poddana interwencji: 526 w przypadku staży, 1 058 w przypadku praktyk zawodowych, 236 w przypadku krótkich praktyk zawodowych w gminach i 241 w przypadku bonów na szkolenia – grupa kontrolna: 77 255	– prawdopodobieństwo znalezienia zatrudnienia w określonym czasie po interwencjach – prawdopodobieństwo znalezienia zatrudnienia na podstawie umowy na czas nieokreślony – liczba dni przepracowanych w okresie po interwencjach	6, 9, 12, 15 i 18 miesięcy od rozpoczęcia udziału w interwencji	– Dane z monitorowania EFS w przypadku grupy poddanej interwencji – Dane administracyjne z publicznych służb zatrudnienia dotyczące osób zarejestrowanych – Dane administracyjne dotyczące umów o pracę zawartych w regionie (Comunicazioni Obbligatorie - COB)
Polska (woj. lubelskie)*	Ocena działań na rzecz pracowników w starszym wieku w województwie lubelskim finansowanych ze środków EFS w okresie programowania 2007–2013 (2015 r.)	Kilka rodzajów aktywnej polityki rynku pracy 2007–2013	PSM + różnica w różnicach	Grupa poddana interwencji: Osoby bezrobotne w wieku 50+ otrzymujące wsparcie Grupa kontrolna: Osoby bezrobotne w wieku 50+, które nie uczestniczyły w żadnym działaniu i które były zarejestrowane w wojewódzkim urzędzie pracy w czasie uczestnictwa członków grupy poddanej interwencji w programie poddanym ewaluacji Grupa poddana interwencji: 145 – grupa kontrolna: 67 102	– prawdopodobieństwo wyrejestrowania z urzędu pracy	w okresie 6, 12, 18 miesięcy od rozpoczęcia udziału w interwencjach	– Dane administracyjne z publicznych służb zatrudnienia

Kraj	Tytuł – rok	Polityka poddana ewaluacji i badany okres interwencji	Metoda	Grupa poddana interwencji i grupa kontrolna – definicja i wielkość	Zmienne wynikowe	Okres obserwacji	Data
Łotwa	Szkolenie zawodowe dla bezrobotnych osób młodych na Łotwie: dowody na podstawie modelu regresji nieciągłej (2018 r.)	Szkolenie zawodowe skierowane do młodzieży niekształcącej się, niepracującej ani nieszkolącej się w wieku 15–29 lat 2014–2015	Model regresji nieciągłej	Grupa poddana interwencji: Osoby bezrobotne, które wzięły udział w szkoleniu zawodowym w ciągu jednego roku od daty rejestracji Grupa kontrolna: Osoby bezrobotne zarejestrowane w okresie między czerwcem 2013 r. a grudniem 2015 r., które nie wzięły udział w żadnym działaniu Grupa poddana interwencji: 898 – grupa kontrolna: 10 717	– prawdopodobieństwo znalezienia zatrudnienia w określonym czasie po interwencjach – dochody	po 12–36 miesiącach od zapisu	– Dane administracyjne z łotewskiej Państwowej Agencji Pracy (SEA) – informacje zarówno o uczestnikach, jak i osobach nieuczestniczących w badaniu, zarejestrowanych jako osoby bezrobotne w konkretnym czasie – Dane administracyjne z krajowego urzędu skarbowego – informacje na temat stanu zatrudnienia w różnych terminach oraz dochodów poszczególnych osób
Niemcy	Wpływ szkolenia językowego dla bezrobotnych imigrantów na zatrudnienie (2021 r.)	Kursy językowe (+ krótkie praktyki zawodowe dla niektórych osób poddanych interwencji) dla migrantów poszukujących pracy 2014	Zmienne instrumentalne (IV)	Grupa poddana interwencji: migranci, którzy rozpoczęli udział w kursie językowym do celów zawodowych w 2014 r. Grupa kontrolna: losowo dobrana próba składająca się z osób, które nie odbyły kursu, posiadały co najmniej przez jeden okres obywatelstwo inne niż niemieckie oraz nie uczestniczyły w żadnych środkach z zakresu aktywnej polityki rynku pracy Grupa poddana interwencji: 8 968 – grupa kontrolna: 26 463	– prawdopodobieństwo znalezienia zatrudnienia (regularne zatrudnienie, regularne zatrudnienie w pełnym wymiarze czasu pracy, regularne zatrudnienie przez więcej niż 6 miesięcy, zatrudnienie z dochodem przewyższającym próg zagrożenia ubóstwem, zatrudnienie jako pracownik wykwalifikowany) – wynagrodzenie dzienne i łączne w ramach regularnego zatrudnienia – łączna liczba dni regularnego zatrudnienia	w okresie 24 miesięcy od rozpoczęcia udziału w interwencjach	– Zintegrowane Historie Zatrudnienia do celów identyfikacji obu grup i zmiennych wynikowych – Dane gromadzone retrospektywnie w toku spotkań z osobami poszukującymi pracy i osobami prowadzącymi sprawy (Werdegangshistorie, WGH) w celu uzyskania dalszych informacji na temat okresów samozatrudnienia lub urlopu rodzicielskiego lub ze względów rodzinnych oraz na temat edukacji i zatrudnienia za granicą
Włochy (Trydent)	Czy programy szkolenia zawodowego są warte swojej ceny? Dowody z analizy kosztów i korzyści (2018 r.)	Szkolenie zawodowe dla osób bezrobotnych 2010–2011	PSM	Grupa poddana interwencji: osoby bezrobotne uczestniczące w kursach zawodowych finansowanych ze środków prowincji i z EFS Grupa kontrolna: osoby, które w dniu rozpoczęcia indywidualnych szkoleń były zarejestrowane jako osoby bezrobotne w rejestrach lokalnych publicznych służb zatrudnienia Grupa poddana interwencji: 818 (kursów finansowanych ze środków prowincji), 114 (kursy finansowane w ramach EFS) – grupa kontrolna: 17	– prawdopodobieństwo znalezienia zatrudnienia – zarobki	w okresie 3, 6, 12, 18, 24, 36 miesięcy od rozpoczęcia udziału w interwencjach	– Dane z monitorowania EFS w przypadku grupy poddanej interwencji – Dane administracyjne z publicznych służb zatrudnienia dotyczące osób zarejestrowanych – Dane administracyjne dotyczące umów o pracę zawartych w regionie (Comunicazioni Obbligatorie - COB) – Dane administracyjne dotyczące deklaracji podatkowych z INPS

Kraj	Tytuł – rok	Polityka poddana ewaluacji i badany okres interwencji	Metoda	Grupa poddana interwencji i grupa kontrolna – definicja i wielkość	Zmienne wynikowe	Okres obserwacji	Data
				236 (w przypadku kursów finansowanych przez prowincję), 1 152 (w przypadku kursów finansowanych w ramach EFS)			– Dane administracyjne z Prowincjonalnego Urzędu ds. Edukacji
Niemcy	Czy programy tworzenia miejsc pracy zwiększają integrację społeczną i dobrostan długotrwale bezrobotnych? (2020 r.)	Program tworzenia miejsc pracy (miejsca pracy dotowane nie dłużej niż przez 36 miesięcy) dla osób w trudnej sytuacji (osoby długotrwale bezrobotne) 2015 r. – czerwiec 2017 r.	PSM	Grupa poddana interwencji: próba uczestników, którzy odpowiedzieli na wszystkie pytania zawarte w kwestionariuszu Grupa kontrolna: osoby, które nie uczestniczyły w programie, były zarejestrowane przez publiczne służby zatrudnienia, spełniały te same kryteria kwalifikowalności do grupy kontrolnej i odpowiedziały na wszystkie pytania zawarte w kwestionariuszu Grupa poddana interwencji: 2 531 w fali 1, 1 191 w fali 2, 450 w fali 3 – grupa kontrolna: jw.	– prawdopodobieństwo znalezienia zatrudnienia – subiektywne miary: zadowolenie z życia, zdrowie psychiczne, przynależność społeczna i status społeczny	po 7, 18, 29 miesiącach od rozpoczęcia udziału w programie	– Zintegrowane Historie Zatrudnienia do celów identyfikacji obu grup i zmiennych wynikowych – Badanie subiektywnych zmiennych wynikowych (wyniki miękkie)
Hiszpania (Asturia)*	Ewaluacja PO EFS Asturia (2019 r.)	Metody nauczania grup uczniów zagrożonych wczesnym kończeniem nauki (druga i trzecia klasa szkoły średniej) 2016–2017	PSM	Grupa poddana interwencji: uczniowie uczestniczący w interwencjach Grupa kontrolna: uczniowie porównywalni pod względem kryteriów kwalifikowalności Grupa poddana interwencji: 1 053 – grupa kontrolna: 3 852	– prawdopodobieństwo otrzymania promocji do następnej klasy	po 12 miesiącach od rozpoczęcia udziału w interwencji	– dane administracyjne dotyczące edukacji Baza danych SAUCE (Ministerstwo Edukacji)
Portugalia	Ewaluacja uczelnianego systemu stypendialnego dla studentów w trudnej sytuacji materialnej w Portugalii (2020 r.) – jeszcze nie opublikowano	Stypendia dla studentów pochodzących z rodzin o niskich dochodach 2021–2018	RDD	Grupa poddana interwencji: kandydaci na studentów pierwszego roku studiów magisterskich lub licencjackich Grupa kontrolna: uczniowie niekwalifikujący się ze względu na przekroczenie ustalonego progu dochodowego Grupa poddana interwencji: 130 602 – grupa kontrolna: 25 400	– prawdopodobieństwo pozostania na liście studentów na koniec pierwszego roku – liczba punktów uzyskanych na koniec pierwszego roku – prawdopodobieństwo ukończenia studiów – liczba lat pozostałych do ukończenia studiów – ostateczna ocena ze studiów	po 12–48 miesiącach od uzyskania stypendium	– Dane administracyjne dotyczące uczniów ubiegających się o stypendium (Dyrekcja Generalna ds. Szkolnictwa Wyższego w Portugalii) – Dane administracyjne zawierające informacje na temat przebiegu studiów i postępów (Dyrekcja Generalna ds. Statystyki i Nauki)

Kraj	Tytuł – rok	Polityka poddana ewaluacji i badany okres interwencji	Metoda	Grupa poddana interwencji i grupa kontrolna – definicja i wielkość	Zmienne wynikowe	Okres obserwacji	Data
Polska (woj. podlaskie)*	Ocena wpływu wsparcia PO Województwa Podlaskiego w latach 2014–2020 na popularyzację kształcenie zawodowego (2020 r.)	Projekt „Dobry zawód – fajne życie” objęty wsparciem z EFS miał na celu popularyzację kształcenia i szkolenia zawodowego i był skierowany do szkół średnich I stopnia 2017–2019	PSM	Grupa poddana interwencji: próba uczniów biorących udział w projekcie Grupa kontrolna: uczniowie, którzy wykazują cechy podobne do uczniów należących do grupy poddanej interwencji, ale którzy nie brali udziału w projekcie Grupa poddana interwencji: 200 – grupa kontrolna: 384	– uczeszczanie w momencie badania do technikum – uczeszczanie w momencie badania do szkoły branżowej I stopnia – uczeszczanie w momencie badania do techników albo szkoły branżowej I stopnia – udział w rekrutacji na uczelnię, do technikum, do szkoły branżowej I stopnia, do szkoły kształcenia zawodowego	po 12–36 miesiącach od interwencji	– dane z monitorowania – Powszechny Elektroniczny System Ewidencji Ludności (PESEL) do doboru grupy kontrolnej – badanie służące do pomiaru zmiennych wynikowych
Włochy (Piemont)*	Wpływ na zatrudnienie wywierany przez bony przeznaczone na programy służb zatrudnienia (Buono servizi lavoro – BSL finansowane z PO EFS Piemont na lata 2014–2020) (2019 r.)	Bony dla służb zatrudnienia (wprowadzenie + szkolenie + oferta) dla osób bezrobotnych w trudnej sytuacji mających status osób bezrobotnych od co najmniej 6 miesięcy 2018	PSM	Grupa poddana interwencji: bezrobotni poddani interwencji Grupa kontrolna: bezrobotni zarejestrowani przez publiczne służby zatrudnienia o podobnych cechach (30-latkowie bezrobotni od 6 miesięcy) Grupa poddana interwencji: 8 125 – grupa kontrolna: 130 000	– prawdopodobieństwo znalezienia zatrudnienia – prawdopodobieństwo znalezienia zatrudnienia na podstawie umowy na czas nieokreślony	po 6, 12, 16 miesiącach od interwencji	– Dane z monitorowania EFS w przypadku grupy poddanej interwencji – Dane administracyjne z publicznych służb zatrudnienia dotyczące osób zarejestrowanych – Dane administracyjne dotyczące umów o pracę zawartych w regionie (Comunicazioni Obbligatorie - COB)
Niemcy*	Ewaluacja programu integracji osób długotrwale bezrobotnych w Niemczech w latach 2014–2020 (2019 r.)	Środki w zakresie integracji długotrwale bezrobotnych stosowane przez co najmniej 24 miesiące 2015–2017	Różnica w różnicach (oszacowani e zamiaru poddania interwencji)	Grupa poddana interwencji: Długotrwale bezrobotni potencjalnie kwalifikujący się w okresie realizacji programu (od 2015 r.) Grupa kontrolna: osoby, które spełniają kryteria kwalifikowalności przed realizacją programu (w latach 2010–2012) Grupa poddana interwencji i grupa kontrolna: 134 515 osób na 237 874 okresów bezrobocia (35% w przypadku grupy poddanej interwencji i 65% w przypadku grupy kontrolnej)	– prawdopodobieństwo znalezienia zatrudnienia (zatrudnienie ze składkami na ubezpieczenia społeczne)	po 24 miesiącach od interwencji	– Dane administracyjne z monitorowania z Federalnego Urzędu Administracyjnego (baza danych ZUWES) – Zintegrowane Historie Zatrudnienia do celów identyfikacji zmiennych wynikowych
Niemcy (Badenia-Wirtembergia)*	Drugie sprawozdanie oceniające osiągnięcie celu szczegółowego A1.1 w ramach PO EFS Badenia-Wirtembergia na lata 2014–2020 (2020)	Środki służące integracji osób bezrobotnych, szczególnie długotrwale bezrobotnych i innych osób w trudnej sytuacji Marzec 2016 r. – grudzień 2017 r.	PSM	Grupa poddana interwencji: uczestnicy zidentyfikowani w bazie danych IEB Grupa kontrolna: osoby, które spełniają kryteria kwalifikowalności w tym samym okresie Grupa poddana interwencji: 1 578 (z 1 800) – grupa kontrolna: nd.	– prawdopodobieństwo znalezienia zatrudnienia (zatrudnienie ze składkami na ubezpieczenia społeczne) – prawdopodobieństwo dalszego pobierania zasiłku dla bezrobotnych	po 15 miesiącach od interwencji	– Dane z monitorowania – Zintegrowane Historie Zatrudnienia do celów identyfikacji zmiennych wynikowych

Uwaga: GPI = grupa poddana interwencji, GK = grupa kontrolna, * oznacza ewaluacje zidentyfikowane przez Punkt Informacyjny ds. Oceny.

JAK SKONTAKTOWAĆ SIĘ Z UE

Osobiście

W całej Unii Europejskiej istnieją setki punktów informacji Europejskiej Europe Direct. Adres najbliższego ośrodka można znaleźć na stronie internetowej: https://europa.eu/european-union/contact_pl

Telefonicznie lub pocztą elektroniczną

Europe Direct to serwis informacyjny, przeznaczony do udzielania odpowiedzi na pytania na temat Unii Europejskiej. Kontakt z serwisem jest możliwy:

- pod bezpłatnym numerem telefonu: 00 800 67891011 (niektórzy operatorzy mogą naliczać opłaty za te połączenia);
- pod poniższym standardowym numerem: +32 22999696 lub
- pocztą elektroniczną: https://europa.eu/european-union/contact_pl

WYSZUKIWANIE INFORMACJI O UE

W internecie

Informacje o Unii Europejskiej we wszystkich językach urzędowych UE są dostępne na stronie internetowej Europa pod adresem: https://europa.eu/european-union/index_pl

Publikacje UE

Bezpłatne i płatne publikacje UE można pobrać lub zamówić na stronie: <https://op.europa.eu/pl/publications> Większą liczbę egzemplarzy bezpłatnych publikacji można otrzymać kontaktując się z Europe Direct lub z lokalnym ośrodkiem informacyjnym (zob. https://europa.eu/european-union/contact_pl).

Prawo UE i powiązane dokumenty

Informacje prawne dotyczące UE, w tym wszystkie unijne akty prawne od 1952 r. we wszystkich językach urzędowych UE, są dostępne w portalu EUR-Lex: <http://eur-lex.europa.eu>

Portal Otwartych Danych UE

Portal Otwartych Danych UE (<http://data.europa.eu/euodp/en>) umożliwia dostęp do zbiorów danych UE. Dane można pobierać i wykorzystywać bezpłatnie, zarówno do celów komercyjnych, jak i niekomercyjnych.



Urząd Publikacji
Unii Europejskiej