

EWALUACJA KONTRFAKTYCZNA PROGRAMÓW ZATRUDNIENIA OSÓB MŁODYCH

PRZEWODNIK METODOLOGICZNY

MÁRTON CSILLAG, JUDIT KREKÓ, ÁGOTA SCHARLE

Współpraca:

MARCO CENTRA, MASSIMILIANO DEIDDA, LUCÍA GORJÓN, VALENTINA
GUALTIERI, KAROL MADOŃ, IGA MAGDA, MARTA PALCZYŃSKA,
MATEUSZ SMOTER, FRANCESCO TRENTINI, CLAUDIA VILLOSIO

Listopad 2020 r.



Ewaluacja kontrfaktyczna programów zatrudnienia osób młodych

Przewodnik metodologiczny

Projekt „Partnerstwo na rzecz zatrudnienia młodzieży – badania ewaluacyjne w Hiszpanii, Polsce, na Węgrzech i we Włoszech” finansowany jest przez Islandię, Liechtenstein i Norwegię, w ramach Funduszu na Rzecz Zatrudnienia Młodzieży EEA i Funduszy Norweskich.

REALIZOWANY PRZEZ:



SPIS TREŚCI

1. WPROWADZENIE	4
2. DANE ADMINISTRACYJNE	8
3. PROCES EWALUACJI	13
3.1. Wybór programu poddanego ewaluacji	13
3.2. Wybór zmiennych wynikowych	13
3.3. Strategia identyfikacji: wybór metody ewaluacji kontrfaktycznej	15
3.4. Heterogeniczność efektów	21
4. KWESTIE ISTOTNE DLA DECYDENTÓW	23
4.1. Interpretacja wyników	23
4.2. Powiązanie wyników z zasadami programu i praktyką jego wdrażania	23
4.3. Trafność zewnętrzna, porównywalność z innymi badaniami i zbiorczymi statystykami	24
4.4. Efekt wypierania i zbędna strata społeczna	24
4.5. Analiza kosztów i korzyści oraz efektywności kosztowej	25
Bibliografia	26
Załącznik 1: Skrócony opis ewaluacji w czterech krajach	28
Włochy	28
Węgry	28
Polska (Ewaluacja 1)	29
Polska (Ewaluacja 2)	30
Hiszpania	30
Załącznik 2: Wykaz skrótów i pojęć	32

1. WPROWADZENIE¹

Niniejszy przewodnik zawiera wskazówki dotyczące ewaluacji polityk rynku pracy skierowanych do osób młodych. Skupiono się na ewaluacjach kontrfaktycznych, które wykorzystują dane administracyjne. Przykłady pochodzą z czterech krajów (Hiszpania, Węgry, Włochy i Polska) uczestniczących w projekcie „Partnerstwo na rzecz zatrudnienia młodzieży”. W jego ramach zbadano skuteczność różnych programów rynku pracy, które mają przeciwdziałać bezrobociu osób młodych. Przewodnik może być również pomocny przy ewaluacji innych programów.

Celem tej publikacji jest dostarczenie analitykom – pracującym w różnych instytucjach prowadzących badania ewaluacyjne (ośrodkach akademickich, ministerstwach, publicznych służbach zatrudnienia) – praktycznego zestawu narzędzi, które pomogą im w procesie ewaluacji. Zakładamy, że odbiorcy przewodnika znają podstawowe terminy statystyczne i ewaluacyjne. Niemniej jednak na końcu opracowania zamieściliśmy słowniczek kluczowych pojęć. Ponieważ zarówno analizowane programy, jak i źródła danych, które wykorzystano do ewaluacji, różnią się pomiędzy krajami, niniejsze opracowanie nie zawiera opisu ujednoliconej metodologii². Opisano w nim raczej ogólne zagadnienia i problemy specyficzne dla kontrfaktycznych ewaluacji programów zatrudnienia skierowanych do osób młodych³.

Program „Gwarancje dla młodzieży” wprowadzono we wszystkich państwach członkowskich UE, ale nie poddano go dotychczas rzetelnej ewaluacji. Ewaluacja polityk publicznych jest niezwykle ważna dla zrozumienia ich oddziaływania. Ewaluacje ilościowe, oparte o dane administracyjne, pomagają zrozumieć, czy cele interwencji zostały osiągnięte, a także jak poprawić efektywność danego programu. Dane administracyjne umożliwiają przeprowadzenie ewaluacji ex post również w krajach, w których programy rynku pracy zostały wprowadzone z pominięciem procesu monitorowania i gromadzenia danych.

Celem kontrfaktycznych ewaluacji polityk rynku pracy jest oszacowanie przyczynowego wpływu interwencji na różne aspekty, na przykład na prawdopodobieństwo zatrudnienia uczestników i czas jego trwania. Umożliwiają one oszacowanie skutków danej interwencji w porównaniu do stanu kontrfaktycznego, czyli takiego, w którym interwencja nie została wprowadzona. Ponieważ w praktyce obserwowanie świata kontrfaktycznego nie jest możliwe, kluczową kwestią jest znalezienie wiarygodnej grupy kontrolnej, której członkowie będą podobni pod każdym względem do uczestników danej interwencji, ale nie wezmą w niej udziału.

W ramach projektu „Partnerstwo na rzecz zatrudnienia młodzieży” ocenialiśmy głównie programy rynku pracy, w których pracodawcom zatrudniającym osoby młode oferowano zachęty finansowe. W dwóch krajach (Węgry i Polska) projekt koncentruje się na analizie programów zatrudnienia subsydiowanego, realizowanych w ramach „Gwarancji dla młodzieży”. Program ten zobowiązuje wszystkie państwa członkowskie UE do zapewnienia młodym ludziom – niepracującym, niekształcącym ani nieszkolącym się – dobrej jakości ofert zatrudnienia, kształcenia, szkolenia lub stażu. W dwóch pozostałych krajach (Włochy i Hiszpania) ewaluowano programy, które mają zwiększyć

1 Jesteśmy wdzięczni następującym osobom za ich cenne uwagi: Namita Datta, Ilf Bencheikh, Tibor Keresztély i András Svraha.

2 Niniejszy przewodnik koncentruje się na kontrfaktycznych ewaluacjach realizowanych przy użyciu danych administracyjnych. Aby uzyskać szersze informacje na temat monitorowania i ewaluacji, zob. na przykład OECD (2009) i UNDP (2009), Gertler et al (2016.), Morris et al (2013).

3 Aby uzyskać szczegółowy opis metod kontrfaktycznej ewaluacji ilościowej, zob. na przykład Angrist i Pischke (2009), Abadie i Cattaneo (2018), Gertler et al (2016) oraz Morris et al (2013), Komisja Europejska (2013).

stabilność zatrudnienia młodych ludzi poprzez zachęcanie pracodawców do zatrudnienia ich w oparciu o stabilne umowy. Główne informacje na temat ewaluowanych programów znajdują się w ramce 1 i 2, a bardziej szczegółowe w załączniku 1.

Wszystkie cztery ewaluacje w ramach projektu „Partnerstwo na rzecz zatrudnienia młodzieży” przeprowadzono z wykorzystaniem jednostkowych danych administracyjnych pochodzących z różnych źródeł, np. ewidencji bezrobotnych, ubezpieczeń społecznych, napływów i odpływów z formalnego zatrudnienia itp., które są dostępne w większości krajów europejskich. Niniejszy przewodnik koncentruje się w szczególności na wyzwaniach i trudnościach związanych z wykorzystaniem danych administracyjnych w ewaluacjach polityk publicznych.

Publikacja podzielona jest na części. W pierwszej podsumowano najważniejsze kwestie związane z uzyskiwaniem dostępu do danych administracyjnych i ich wykorzystaniem. Jest to istotne, ponieważ różnego typu ograniczenia w dostępie do danych mają wpływ na proces ewaluacji na wszystkich jej etapach. Druga część zawiera omówienie procesu ewaluacji ilościowej. Szczególny nacisk położono na opisanie metod, które umożliwiają analitykom uzyskanie wiarygodnych rezultatów. W ostatniej omówiono zagadnienia, które mogą być przydatne na etapie prezentowania wyników decydentom. Poszczególne części są oparte na doświadczeniach i zilustrowane konkretnymi przykładami pochodzącymi z czterech krajów uczestniczących w projekcie „Partnerstwo na rzecz zatrudnienia młodzieży”. Wykaz skrótów i pojęć zamieszczony na końcu niniejszego opracowania zawiera definicje kluczowych pojęć technicznych.

Ramka 1. Szczegółowe informacje dotyczące ewaluowanych programów: Włochy

WŁOCHY	
Polityka	Ewaluacja wspólna: 1) Obniżenie kosztów zwolnienia w przypadku umów na czas nieokreślony (umowa z gradacją ochrony). 2) Niższe składki na ubezpieczenia społeczne dla nowych umów na czas nieokreślony i dla konwersji z umowy na czas określony na umowę na czas nieokreślony.
Kryteria kwalifikacji	Osoby, które nie miały umowy na czas nieokreślony w okresie 6 miesięcy poprzedzających ich zatrudnienie.
Okres, rodzaj	1) Zmiana regulacji dotyczących zwolnień na stałe. 2) Subsydia do zatrudnienia przez 36 miesięcy.
Dotacja	1) Zmiana regulacji dotyczących zwolnień. 2) Ulga 100% na pozapłacowe koszty pracownicze.
Dane	Sisco (statystyczny system obowiązkowej informacji online). Sisco to ewidencja administracji publicznej, zawierająca podstawowe informacje na temat zatrudnienia, przepływów pomiędzy różnymi rodzajami umów i przyczyn ich zakończenia.
Próba	Umowy o pracę zarejestrowane w latach 2014–2015.
Metodologia	DiD (metoda różnicy w różnicach) na skorygowanej próbie z wykorzystaniem interakcji zmiennych definiujących grupę wieku ze zmienną wskazującą czy dana osoba otrzymała wsparcie
Rezultat	Udział nowych umów na czas nieokreślony w stosunku do całkowitej liczby umów zarejestrowanych w 2015 r.

Ramka 2. Szczegółowe informacje dotyczące ewaluowanych programów: Polska

POLSKA	
Polityka	1) Zatrudnienie subsydiowane. 2) Staż, szkolenie, roboty publiczne, prace interwencyjne, bon stażowy i bon szkoleniowy.
Kryteria kwalifikacji	Osoby bezrobotne w wieku do 30 lat w przypadku 1 i 2.
Okres, rodzaj	1) 12 miesięcy zatrudnienia subsydiowanego + 12 miesięcy zatrudnienia bez subsydiowania na tych samych warunkach. 2) 1-12 miesięcy w zależności od formy wsparcia.
Dotacja	1) Subsydiowanie po stronie pracodawcy do wysokości płacy minimalnej + składki na ubezpieczenia społeczne (~500 + 100 EUR w 2018 r.). 2) Subsydiowanie robót publicznych i prac interwencyjnych oraz dodatkowa premia dla pracodawcy w przypadku bonów stażowych.
Dane	Dane o osobach bezrobotnych powiatowych urzędów pracy
Próba	2015–2016
Metodologia	1) RDD (model regresji nieciągłej) w połączeniu z DiD. 2) PSM (propensity score matching).
Rezultat	1) wykreślenie z ewidencji bezrobotnych w ciągu 12 do 36 miesięcy od rozpoczęcia udziału w programie. 2) wykreślenie z ewidencji osób bezrobotnych przy jednoczesnym nieuczestniczeniu w innej formie wsparcia w ciągu od 12 do 36 miesięcy od rozpoczęcia udziału w programie.

Ramka 3. Szczegółowe informacje dotyczące ewaluacji kraju: Hiszpania

HISZPANIA	
Polityka	1) Zatrudnienie subsydiowane. Umowa o staż.
Kryteria kwalifikacji	Licencjat lub wykształcenie zawodowe, do 30 r.ż., osoby bezrobotne.
Okres, rodzaj	Minimum 6 miesięcy, maksymalnie 2 lata. Pełny lub częściowy wymiar czasu pracy.
Dotacja	Obniżenie składek na ubezpieczenia społeczne o 50–75%. Premie za przejście na umowę stałą w wys. 500 EUR na rok dla mężczyzny lub 700 EUR dla kobiety, na okres pierwszych trzech lat.
Dane	1) Rejestry PSZ, pełna ewidencja podpisanych umów, powiązane ze 2) zbiorem rejestrów ubezpieczeń społecznych dotyczącym historii zatrudnienia Continuous Sample of Work Histories (CSWH)
Próba	2012–2018.
Metodologia	Zastosowanie modelu logitowego i wielomianowego modelu logitowego do porównania uczestników stażu z kwalifikującymi się osobami, które z niego nie skorzystały i pracowały w oparciu o zwykłe umowy czasowe.
Rezultat	1) Prawdopodobieństwo pozostania danej osoby w firmie po zakończeniu umowy o staż. 2) Prawdopodobieństwo podpisania umowy na czas nieokreślony w tej samej firmie. 3) Prawdopodobieństwo podpisania umowy na czas nieokreślony w innej firmie.

Ramka 4. Szczegółowe informacje dotyczące ewaluacji kraju: Węgry

WĘGRY	
Polityka	Zatrudnienie subsydiowane na 90-dniowy okres próbny
Kryteria kwalifikacji	Do 25 r.ż., osoby zarejestrowane jako bezrobotne.
Okres, rodzaj	Maksymalnie 90 dni.
Dotacja	100% kosztów wynagrodzenia (wynagrodzenie + składki pracodawcy).
Dane	1) Rejestry PSZ powiązane z 2) ewidencją ubezpieczeń społecznych i rejestry edukacyjne.
Próba	2015–2017 (zmienne historyczne są konstruowane od 2003 roku).
Metodologia	1) Metoda najbliższego sąsiada i technika propensity score matching, aby porównać uczestników zatrudnienia subsydiowanego z uczestnikami robót publicznych i szkoleń. 2) Metoda różnicy w różnicach, przy wykorzystaniu faktu rozpoczęcia programu w jednym regionie o 9 miesięcy później.
Rezultat	1) Utrzymanie pracy po 6 i 12 miesiącach od rozpoczęcia subsydiowania. 2) Wynagrodzenie łączne w 6 i 12 miesiącu od rozpoczęcia subsydiowania.

2. DANE ADMINISTRACYJNE

Zbiory danych administracyjnych charakteryzują się zazwyczaj szerokim zasięgiem oraz dużą liczebnością próby. W wielu przypadkach możliwe jest łączenie różnych zbiorów danych⁴. Wiarygodność danych administracyjnych jest zazwyczaj wyższa niż tych z badań ankietowych. Często zdarza się jednak, że są niekompletne. W ewaluacjach polityki rynku pracy najczęstszymi zbiorami danych administracyjnych są te dotyczące ubezpieczeń społecznych, pochodzące z ewidencji osób bezrobotnych, dane publicznych służb zatrudnienia (PSZ) – włączając w to szczegółowe dane dotyczące programów rynku pracy i ich uczestników – oraz bazy pracodawców. Ponieważ zbiory danych administracyjnych zawierają dane na poziomie indywidualnym lub na poziomie firm, dostęp do nich jest zwykle ściśle uregulowany i ograniczony ze względu na przepisy dotyczące ochrony danych osobowych. Sposób gromadzenia i przechowywania danych administracyjnych oraz uwarunkowania prawne związane z ich pozyskiwaniem i obsługą różnią się w poszczególnych krajach. Oznacza to, że ewaluator może napotkać trudności związane z długim procesem uzyskiwania zgody na dostęp do niezbędnych danych. Przed skorzystaniem z danych administracyjnych należy zadać sobie kilka pytań⁵:

- **Czy istnieje interesująca nas baza danych administracyjnych na szczeblu krajowym?**

Na zakres ewaluacji wpływ ma to, czy dane administracyjne są gromadzone i przetwarzane na szczeblu krajowym czy regionalnym. Na **Węgrzech**, w **Hiszpanii** i w **Polsce** rejestry osób bezrobotnych są zintegrowane w krajowej bazie danych. **Włochy** prowadzą krajową ewidencję administracyjną zatrudnienia o nazwie SISCO.

- **Czy możliwe jest łączenie różnych zbiorów danych, aby wzbogacić je o nowe zmienne?**

Informacje o danej osobie zwykle są gromadzone w różnych bazach danych prowadzonych przez różne instytucje, na przykład zakłady ubezpieczeń społecznych, urzędy pracy, urzędy skarbowe, organizatorów programów rynku pracy. Łączenie różnych zbiorów może zwiększyć zakres dostępnych informacji. Na przykład na **Węgrzech** bazy PSZ można połączyć z bazami instytucji przynajmniej świadczenia społeczne, które zawierają informacje o zatrudnieniu i wynagrodzeniu wszystkich zarejestrowanych w PSZ osób bezrobotnych. To umożliwia korzystanie z większej liczby zmiennych wynikowych (takich, jak jakość zatrudnienia), a także bogatszej puli obserwowalnych cech (takich, jak historia zatrudnienia). W **Hiszpanii** FEDEA ma dostęp do dwóch baz danych administracyjnych, które można połączyć ze sobą: zbioru dotyczącego ubezpieczeń społecznych (CSWH) oraz rejestrów PSZ dotyczących bezrobocia i umów zawieranych każdego miesiąca z pracownikami w Hiszpanii. W wielu przypadkach łączenie różnych zbiorów danych jest jednak ściśle uregulowane i ograniczone lub nawet niedozwolone ze względu na przepisy dotyczące ochrony danych osobowych. Na przykład w **Polsce** nie można połączyć danych o osobach bezrobotnych zarejestrowanych w powiatowych urzędach pracy z danymi Zakładu Ubezpieczeń Społecznych o zatrudnieniu i wynagrodzeniach. Wyjątek stanowią tu przypadki precyzyjnie określone przez przepisy prawne, np. dotyczące sprawozdawczości dla Komisji Europejskiej.

4 Zob. Caliendo et al. (2017) na temat wartości dodanych dla ewaluacji polityk rynku pracy wynikających z posiadania dostępu do danych.

5 Aby uzyskać wskazówki krok po kroku w zakresie wykorzystywania danych administracyjnych w ewaluacjach kontrfaktycznych programów Europejskiego Funduszu Społecznego, zob. także KE (2020).

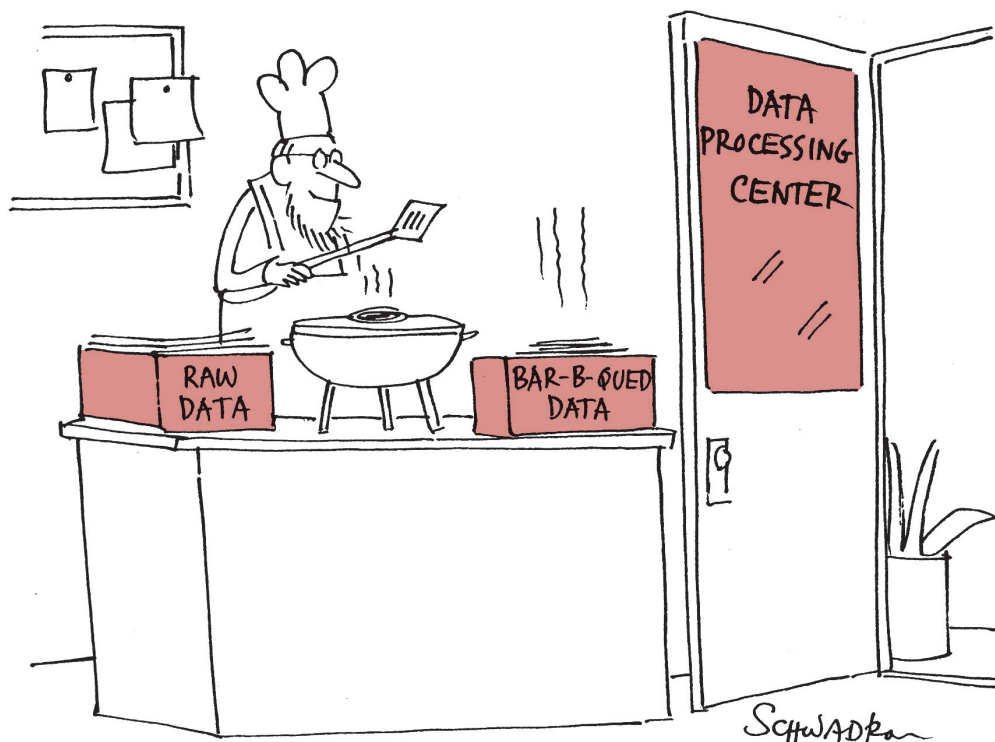
- **Jak wygląda procedura udostępniania danych?**

Nie ma unijnych regulacji dotyczących dostępu do danych, a wymogi administracyjne w tym zakresie różnią się w poszczególnych krajach. Na przykład w Hiszpanii bazę danych o historii zatrudnienia (CSWH) można otrzymać po wystosowaniu prośby do hiszpańskiego zakładu ubezpieczeń społecznych. Dane PSZ nie są dostępne dla badaczy, ale FEDEA uzyskała dostęp do nich na podstawie umowy o współpracy. W Polsce proces ten nie jest sformalizowany i jest w dużej mierze uznaniowy⁶. Brakuje procedur, które precyzowałyby, w jaki sposób można uzyskać dostęp do danych z ZUS oraz danych na temat osób bezrobotnych zarejestrowanych w powiatowych urzędach pracy.

- **Jakie są uwarunkowania instytucjonalne i administracyjne związane z wykorzystywaniem danych?**

Badaczom zdecydowanie zaleca się nawiązanie kontaktu z administratorami danych, aby wyjaśnić uwarunkowania instytucjonalne, omówić celowość badania, sposoby gromadzenia i przetwarzania danych. Wiarygodność i jakość danych mogą różnić się nawet w obrębie jednej bazy, zależnie od tego, czy są one wykorzystywane w ramach procedury sprawozdawczej, ewaluacyjnej czy administracyjnej. Dogłębna wiedza o mechanizmach gromadzenia, przechowywania i przetwarzania danych ma zasadnicze znaczenie podczas przygotowywania strategii ewaluacji.

W niektórych przypadkach, zwłaszcza gdy gromadzenie danych i ich walidacja należą do kompetencji jednej organizacji lub kilku współpracujących ze sobą, ewaluator może mieć wpływ na sposób gromadzenia danych i ich zakres. Dzięki temu w większym stopniu przyczynią się one do osiągnięcia określonego celu. Stała wymiana informacji pomiędzy ewaluatorem a administratorem danych jest szczególnie istotna przy prowadzeniu długofalowych lub regularnych badań ewaluacyjnych.



© www.CartoonStock.com

6 Polska obecnie buduje zintegrowaną platformę analityczną, która będzie działać jako centralny system, ułatwiający analizowanie danych administracyjnych z różnych źródeł.

- **Czy dane są kompletne i wysokiej jakości?**

Ponieważ administracyjne zbiory danych zwykle obejmują kompletny zbiór obserwacji – lub losową próbę ze wszystkich obserwacji – reprezentatywność danych jest zazwyczaj zapewniona. Mogą jednak zawierać poważne błędy i luki. Dotyczy to zwłaszcza tych zmiennych, których wypełnienie jest opcjonalne. Badacz musi zawsze ocenić, jaka jest skala problemu, które zmienne zawierają błędy lub braki i w razie potrzeby je skorygować. Kompletność i jakość bazy danych można sprawdzić na przykład poprzez porównanie jej ze zintegrowanymi danymi pozyskanymi z innych źródeł.

- **Czy dla grupy eksperymentalnej i grupy kontrolnej dostępne są te same dane?**

Częstym problemem przy ewaluacji polityk z wykorzystaniem danych administracyjnych jest to, że zbiory zawierają inne informacje (mają inną strukturę, inne zmienne) na temat uczestników programu i inne na temat potencjalnej grupy kontrolnej. Może to utrudniać zdefiniowanie wspólnych zmiennych wynikowych i kontrolnych. W wielu przypadkach ewidencja bezrobotnych zawiera jedynie ograniczone informacje na temat statusu zatrudnienia osób, które zostały wykreślone z rejestru lub zakończyły udział w programie rynku pracy.

Na **Węgrzech** dane PSZ zawierają informacje na temat statusu zatrudnienia wyłącznie uczestników programów rynku pracy i tylko przez 180 dni po ich zakończeniu. Brakuje jednak danych na temat statusu zatrudnienia osób bezrobotnych, które nie uczestniczyły w programach rynku pracy. Problem ten można rozwiązać, wykorzystując jako zmienną wynikową status w rejestrze bezrobotnych, który określany jest dla wszystkich osób bezrobotnych. Badacz może dzięki temu sprawdzić, czy dana osoba nadal figuruje w rejestrze, i założyć, że jeśli nie, to znaczy, że jest zatrudniona. Należy jednak pamiętać, że poza ewidencją pozostają nie tylko osoby, które znalazły zatrudnienie, ale także te, które nie pracują i nie poszukują zatrudnienia (bierne zawodowo). W ewaluacji węgierskiej problem ten rozwiązano, wykorzystując dane dotyczące zatrudnienia z połączonej bazy o ubezpieczeniach społecznych. Podobne problemy występują w Polsce. Status zatrudnienia można określić jedynie w przybliżeniu na podstawie informacji o tym, czy dana osoba figuruje w rejestrze osób bezrobotnych i czy jednocześnie nie bierze udziału w żadnym programie rynku pracy, lub na podstawie ograniczonych informacji na temat przyczyny wyrejestrowania z ewidencji bezrobotnych.

- **Czy dane umożliwiają identyfikację uczestników programu czy tylko osób spełniających kryteria kwalifikacji?**

W wielu przypadkach baza danych zawiera informacje, które umożliwiają identyfikację osób kwalifikujących się do programu, ale nie jego uczestników. W takiej sytuacji badacz może oszacować efekt programu wyłącznie w odniesieniu do kwalifikującej się subpopulacji, ale nie może ocenić jego wpływu na uczestników. W terminologii ewaluacyjnej: badacz nie jest w stanie zidentyfikować średniego efektu netto (ATE) lub średniego efektu netto na jednostki poddane oddziaływaniu (ATT), a jedynie zamiar oddziaływania (ITT).

Jest to mniejszy problem, jeżeli program jest uniwersalny (np. bon lub quasi-bon, do którego uprawnione są wszystkie młode osoby bezrobotne lub który był obowiązkowy dla niektórych grup). Rozróżnienie to może mieć jednak większe znaczenie, jeżeli w programie może wziąć udział tylko część osób bezrobotnych lub gdy kryteria kwalifikacji do programu nie są sprecyzowane.

Na przykład: jeżeli szacowany wpływ programu na całą kwalifikującą się populację jest niski, badacz nie ma pewności, czy za niedostateczne wyniki odpowiada mała liczba osób biorących w nim udział czy jego słaby efekt. To rozróżnienie może być ważne, ponieważ wnioski dotyczące dalszego postępowania będą inne w przypadku niskiego wskaźnika udziału, a inne w przypadku słabego wpływu na uczestników programu⁷. Pierwszy może wymagać zintensyfikowania działań na rzecz zwiększenia zasięgu, a drugi zmian w sposobie realizacji programu.

Zbiory PSZ zazwyczaj zawierają szczegółowe informacje dotyczące programów rynku pracy i umożliwiają identyfikowanie ich uczestników⁸. Na **Węgrzech** baza danych PSZ zawiera szczegółowe informacje na temat rodzaju programu oraz daty jego rozpoczęcia i zakończenia. W przypadku programu zatrudnienia subsydiowanego, który był analizowany w ramach ewaluacji węgierskiej – baza danych nie zawiera informacji na temat kwoty subsydiowania i wynagrodzeń, ale dane te są dostępne w powiązanej bazie danych na temat ubezpieczeń społecznych.

W **Hiszpanii** obie bazy danych umożliwiają identyfikację osób pracujących w oparciu o umowę o staż. W obu bazach danych wykorzystanych do ewaluacji hiszpańskiej – ewidencji historii zatrudnienia (Continuous Sample of Work Histories, CSWH) i rejestrach bezrobotnych PSZ – można bez problemu zidentyfikować osoby zatrudnione w oparciu o umowę o staż, a także określić początek i koniec ich umowy. Co więcej, zawierają one informacje o wszystkich pozostałych umowach, innych niż umowa o staż, dzięki czemu możemy wykorzystać je jako zmienne kontrolne. We **Włoszech** natomiast baza SISCO nie zawiera informacji o uczestnikach programu, w związku z czym model ewaluacji określa jedynie oddziaływanie na kwalifikującą się populację. W **Polsce** rejestr osób bezrobotnych zawiera szczegółowe informacje na temat programów polityki rynku pracy, łącznie z datą rozpoczęcia i zakończenia uczestnictwa w danym programie.

- **Na ile szczegółowy jest dostępny zbiór obserwowalnych cech?**

Pula dostępnych obserwowalnych cech, które mogą służyć jako zmienne kontrolne, ma wpływ zarówno na strategię identyfikacji, jak i analizę heterogeniczności wśród różnych grup. Dlatego, przed planowaniem strategii ewaluacji należy zastanowić się, które zmienne są dostępne lub możliwe do uzyskania. W rejestrach administracyjnych zwykle znajdują się dane na temat płci, wieku, miejsca zamieszkania. Jednakże, poza tym podstawowymi informacjami, zmienne w administracyjnych bazach danych są zróżnicowane.

Poziom wykształcenia to zazwyczaj kluczowa zmienna, która ma ogromny wpływ na rezultaty w zakresie zatrudnienia, a tym samym jest to zwykle jedna z najważniejszych zmiennych w ewaluacji polityk rynku pracy. Jest ona dostępna w **węgierskich i polskich** danych PSZ oraz **włoskim** SISCO. W **Hiszpanii** natomiast dane o wykształceniu pochodzą ze spisu ludności i są aktualizowane tylko raz na dziesięć lat.

Dodatkowe przydatne zmienne to informacje dotyczące: gospodarstwa domowego (na przykład w **Hiszpanii** o liczbie osób współzamieszkujących), różnych świadczeń (na przykład o zasiłkach macierzyńskich na **Węgrzech**), obecności małego dziecka w gospodarstwie domowym (w **Polsce**).

⁷ Problem ten można zredukować poprzez wykorzystanie informacji zewnętrznych, które pomogą określić przybliżony wskaźnik udziału.

⁸ Jednakże, liczbę osób spełniających kryteria kwalifikacji do programu należy jednak zazwyczaj oceniać w przybliżeniu, na podstawie podstawowych cech osób bezrobotnych.

Co do ewidencji zatrudnienia, SISCO we **Włoszech** zawiera informacje na temat branży i zawodu, ale tej ostatniej brakuje na przykład w **hiszpańskiej** bazie danych dotyczących ubezpieczeń społecznych.

Oprócz obserwowalnych cech osób fizycznych, umów i firm, warto uwzględnić także zmienne makroekonomiczne. Na przykład w polskiej ewaluacji uwzględniono lokalną stopę bezrobocia i lokalny poziom wynagrodzeń.

Ponadto, wskazane jest posiadanie informacji na temat uczestników sprzed wprowadzenia ewaluowanego programu lub sprzed jego istotnej modyfikacji, dzięki czemu można porównać je do sytuacji po wprowadzeniu lub modyfikacji programu, stosując metodę różnicy w różnicach. W innych przypadkach zmienne pokazujące historię zatrudnienia mogą służyć jako kontrolne. Na **Węgrzech** pełna historia wcześniejszego zatrudnienia wraz z informacjami z bazy danych o ubezpieczeniach społecznych umożliwiają skonstruowanie historii zatrudnienia danej osoby. Może to wzbogacić naszą wiedzę na temat jej doświadczenia zawodowego i motywacji.

3. PROCES EWALUACJI

3.1. Wybór programu poddanego ewaluacji

Programy rynku pracy składają się często z wielu elementów. Ewaluator musi zdecydować, który z nich wybrać do ewaluacji. Badanie może obejmować cały pakiet interwencji lub koncentrować się na jednej z nich. Wybór ten bywa ograniczony zasadami programu i sposobem realizacji. Na przykład: jeżeli elementy programu są ze sobą połączone, ponieważ w ten sposób najlepiej odpowiadają na potrzeby osób bezrobotnych, ewaluacja oddzielnego elementu może nie być możliwa (chyba że dostępne są obszerne dane na temat osób bezrobotnych). Na przykład, w przypadku ewaluacji **węgierskiej** – po 90-dniowym programie zatrudnienia subsydiowanego możliwe jest dalsze zatrudnienie u tego samego pracodawcy. Przez maksymalnie 8 miesięcy może on liczyć na dalsze częściowe subsydiowanie zatrudnienia, ponieważ zasady programu pozwalają na takie połączenie w określonych przypadkach. Pominięcie tego aspektu w analizie spowodowałoby poważne przeszacowanie wskaźnika zatrudnienia liczonego po 6 miesiącach od ukończenia programu. W podmiotach, które mają dostęp do tego dodatkowego subsydiowania, wskaźnik zatrudnienia będzie bowiem zbliżony do 100%. Z kolei programy rynku pracy oferowane przez powiatowe urzędy pracy w **Polsce** zawierają najczęściej tylko jedną formę wsparcia, np. szkolenie, staż lub dotację na założenie działalności gospodarczej. Bardzo rzadko zdarza się łączenie różnych elementów (jakkolwiek usługi doradztwa zawodowego są dostępne dla wszystkich).

Wybór programu ma wpływ na wybór grupy eksperymentalnej i potencjalnych grup kontrolnych. Jeżeli dane nie pozwalają na identyfikację uczestników biorących udział w danym programie, a jedynie umożliwiają obserwację grupy spełniającej kryteria kwalifikacji do programu, ewaluacja konkretnego elementu programu również nie jest możliwa.

Kluczowe jest przeanalizowanie całego programu oraz wszystkich jego elementów z osobna, aby ocenić zakres potencjalnej ewaluacji. Warto odpowiedzieć sobie na następujące pytania: Jaki problem starano się rozwiązać za pomocą tego programu i skąd jego inicjatorzy wiedzieli, że było to problemem? Czy program był obowiązkowy czy dobrowolny? Jakie kryteria kwalifikacji zostały przyjęte? Jak kształtuje się zakres geograficzny i czasowy programu? Co wiemy o grupach docelowych, kolejności przyznawania wsparcia, skali programu? Na jakie inne kwestie należy zwrócić uwagę, aby dostrzec mocne i słabe punkty programu? Jak wyglądają typowe kombinacje różnych programów i jakie niosą one konsekwencje dla ewaluacji i interpretacji wyników?

Na przykład program objęty ewaluacją we **Włoszech** zachęca pracodawców do zatrudniania osób młodych poprzez obniżanie kwot składek na ubezpieczenia społeczne oraz zastąpienie umowy na czas nieokreślony umową z gradacją ochrony – co skutkuje obniżeniem kosztów zwolnienia. Z pierwszego rozwiązania można było skorzystać na zasadzie dobrowolności w określonym czasie, natomiast drugie spowodowało zmianę przepisu prawnego i obowiązywało w odniesieniu do wszystkich nowych umów na czas nieokreślony, zawieranych po wprowadzeniu go. Programy weszły w życie w zbliżonym czasie, w związku z czym ewaluacja ocenia połączonego efekt ich obu.

3.2. Wybór zmiennych wynikowych

Punktem wyjścia przy wyborze zmiennej wynikowej i metody ewaluacji jest ustalenie celu programu i logiki interwencji. Jakie są oczekiwane efekty? Jak długo powinny trwać? Czy występują efekty niezamierzone? Po uzyskaniu jasnej odpowiedzi na te pytania pojawiają się kolejne:

- Jaki rezultat nas interesuje: status zatrudnienia, zatrudnienie w tej samej firmie, rodzaj umowy, poziom wynagrodzenia itp.?
- Jaki horyzont czasowy będzie analizowany: krótko- czy długoterminowy (więcej niż jeden rok po programie)? W idealnych warunkach badacz może analizować rezultaty w różnych momentach i oceniać, jak zmieniają się w czasie. Podczas ewaluacji programów zatrudnienia subsydiowanego musi uwzględnić fakt, że mogą być one połączone z obowiązkiem późniejszego zatrudnienia (bez subsydiowania) przez określony czas. Badacz powinien zatem analizować rezultat po całym okresie zatrudnienia.

Koncepcja może zostać poważnie ograniczona ze względu na dostęp do danych. W wielu przypadkach nie można uzyskać informacji na temat długoterminowych rezultatów, a dane PSZ zazwyczaj nie zawierają informacji na temat wynagrodzeń.

Celem programów zatrudnienia subsydiowanego jest zwiększenie prawdopodobieństwa znalezienia (dobrej) pracy. Najbardziej istotnymi rezultatami są zatem: prawdopodobieństwo bycia zatrudnionym przez określony czas po zakończeniu programu oraz – na przykład – poziom wynagrodzenia. Na przykład – w przypadku ewaluacji **węgierskiej** – 90-dniowy program zatrudnienia subsydiowanego ma umożliwić osobom młodym zdobycie doświadczenia zawodowego i jednocześnie zmniejszyć ryzyko, które podejmuje pracodawca, zatrudniając pracownika bez doświadczenia. Założenie jest takie, że udział w programie zwiększy prawdopodobieństwo zatrudnienia po jego zakończeniu, a najlepiej, jeśli dodatkowo przyczyni się do wzrostu wynagrodzenia uczestników. Głównymi zmiennymi wynikowymi są zatem: status zatrudnienia po pół roku, roku, dwóch latach itp. od zakończenia programu oraz wysokość wynagrodzenia w tym czasie. Brak dostępu do danych ogranicza jednak możliwość długofalowego badania efektów; w tym konkretnym przypadku rezultaty długoterminowe nie są dostępne, ponieważ program został wprowadzony w 2015 r., a dostępną próbą obejmuje tylko 3 lata: 2015–2017.

Dane powiatowych urzędów pracy w **Polsce** ograniczają wybór zmiennych wynikowych. Na ich podstawie można zaobserwować jedynie, czy dana osoba powróciła do ewidencji bezrobotnych po x miesiącach. Nieobecne w rejestrze są jednak nie tylko osoby pracujące, ale także nieaktywne zawodowo. Ponadto w Polsce udział w programie rynku pracy jest równoznaczny z wykreśleniem danej osoby z rejestru na czas uczestnictwa w nim. Te osoby zatem również nie figurują w ewidencji bezrobotnych. Aby ominąć ten problem, rezultat można zawęzić do osób nieuwzględnionych w ewidencji i nieuczestniczących w żadnym programie rynku pracy⁹.

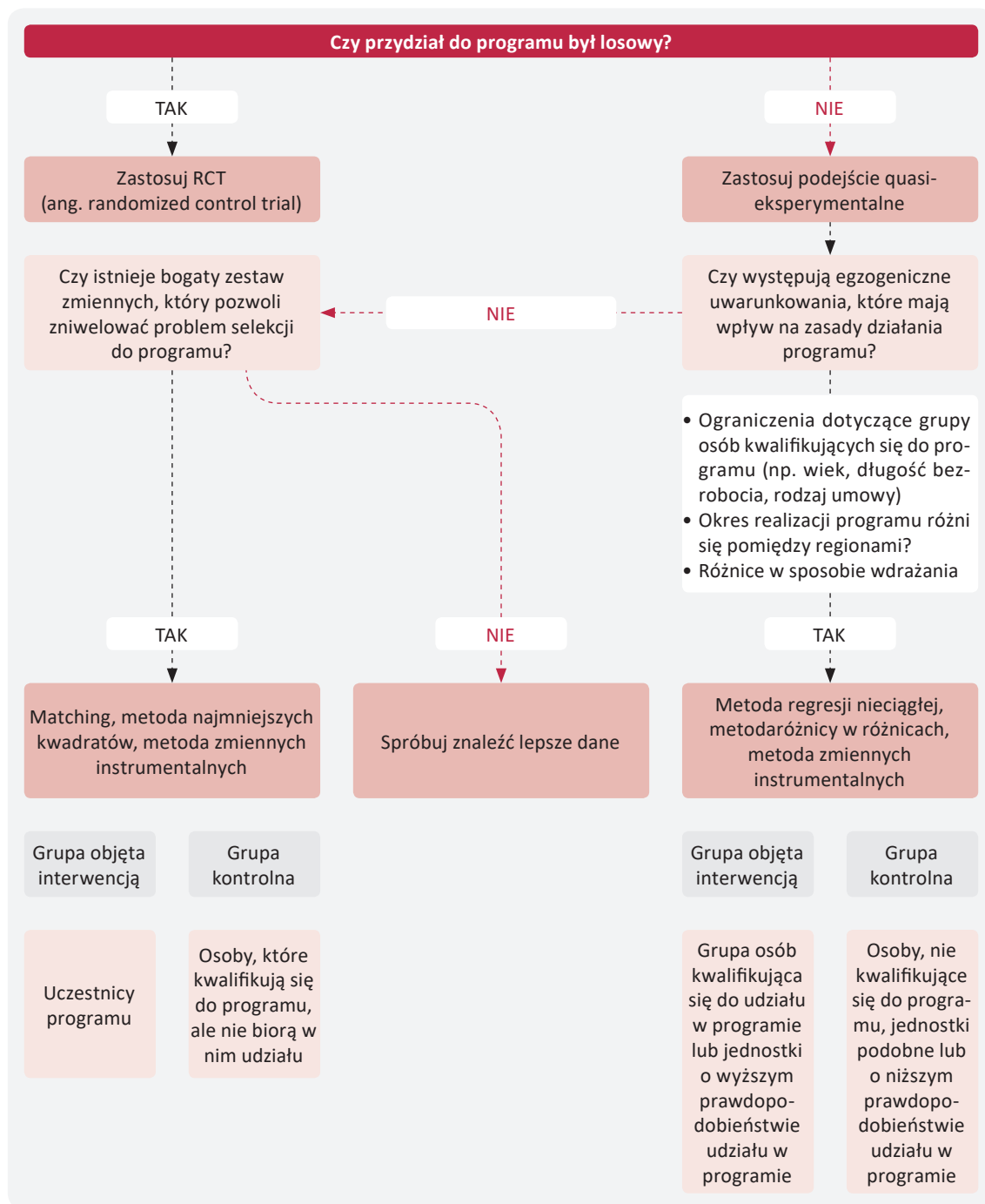
Celem hiszpańskiej umowy o staż i włoskich instrumentów zachęty do zatrudnienia jest wzrost stabilności zatrudnienia. Z tego względu zmienne wynikowe wskazują, czy sytuacja zawodowa młodej osoby ustabilizowała się po udziale w programie. Główną zmienną wynikową ewaluacji **hiszpańskiej** jest prawdopodobieństwo pozostania w tej samej firmie po zakończeniu stażu. Drugą jest prawdopodobieństwo pozostania w tej samej firmie na umowie na czas nieokreślony. Trzecią jest prawdopodobieństwo znalezienia po stażu stałej pracy w innej firmie, co również pokazuje stabilność rynku pracy jako pośredni efekt stażu. Główną zmienną wynikową **ewaluacji włoskiej** jest udział nowych umów na czas nieokreślony w 2015 r. wśród wszystkich umów zarejestrowanych w tym roku.

⁹ Horyzont czasowy rezultatu objętego ewaluacją jest również ważny. Na przykład subsydiowanie dla przedsiębiorców w Polsce wymaga kontynuowania działalności przez co najmniej rok pod groźbą konieczności zwrotu subsydiowania, a zatem jego skuteczność można ocenić dopiero po zakończeniu okresu zatrudnienia subsydiowanego i obligatoryjnego zatrudnienia bez subsydiowania.

3.3. Strategia identyfikacji: wybór metody ewaluacji kontrfaktycznej¹⁰

Nie da się obserwować funkcjonowania tej samej osoby w dwóch różnych sytuacjach: gdy uczestniczy w programie i gdy w nim nie uczestniczy. To główny problem, który pojawia się podczas pomiaru efektu przyczynowego każdego programu. Badacz musi wnioskować na temat sytuacji kontrfaktycznej na podstawie zmiennych, które może zaobserwować.

Rys. 3-1. Schemat logiki ewaluacji kontrfaktycznej



¹⁰ Podstawy najważniejszych koncepcji i metod są opisane w Glosariuszu.

W idealnej sytuacji ewaluacja polityk rynku pracy powinna bazować na podejściu eksperymentalnym, w którym osoby spełniające kryteria kwalifikacji do programu są losowo przydzielane do grupy eksperymentalnej i kontrolnej. Jeżeli tak nie jest, przydział nie jest losowy, a uczestnicy mogą się różnić od osób nieobjętych programem pod względem różnych cech – obserwowalnych i nieobserwowalnych. To może być źródłem dwóch podstawowych problemów. Po pierwsze, perspektywy zatrudnienia i zarobki uczestników i osób nieuczestniczących w programie mogłyby się różnić nawet bez jego wprowadzenia (różnica bazowa). Po drugie: program mógłby mieć inny wpływ na uczestników programu niż na osoby, które nie zostały nim objęte (heterogeniczny efekt oddziaływania). Ponadto trzecim źródłem problemów może być sytuacja, gdy samo wprowadzenie programu wpływa zarówno na wyniki uczestników programu, jak i osób, które w nim nie uczestniczyły, na przykład poprzez efekt naśladownictwa (naruszenie zasady Stable Unit Treatment Values Assumption, SUTVA). Jeśli nie można skorzystać z podejścia eksperymentalnego, badacz musi zastosować podejście quasi-eksperymentalne, żeby niwelować skutki nielosowego przydziału do programu (zob. rys. 3–1).

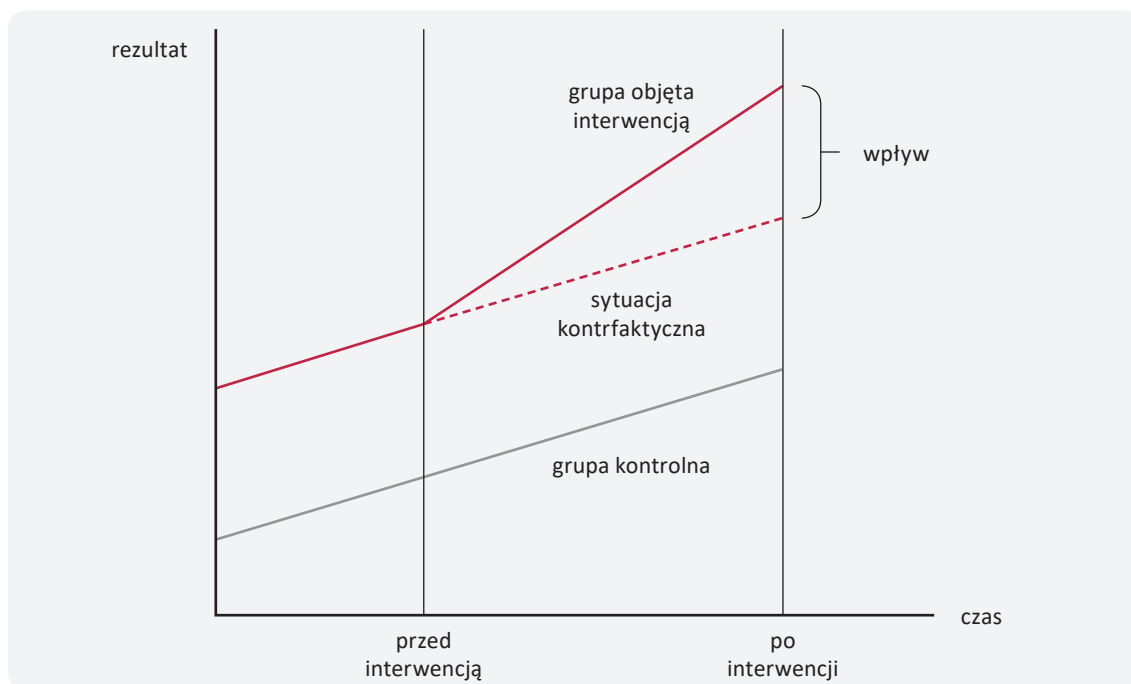
Warunki quasi-eksperymentalne można stworzyć na dwa sposoby.

Jedną opcją jest znalezienie **egzogenicznych uwarunkowań umożliwiających zakwalifikowanie się do programu**. Celem tego ćwiczenia jest wyodrębnienie osób, które np. ze względu na zasady programu: (i) nie spełniają kryteriów kwalifikacji i są przez to wyłączone z programu, (ii) kwalifikują się do programu i w nim uczestniczą lub (iii) kwalifikują się do programu, ale w nim nie uczestniczą. Najczęstszym uwarunkowaniem egzogenicznym są:

- *kryteria kwalifikacji (np. wiek, okres bez zatrudnienia lub bez umowy długoterminowej)*

We **Włoszech** kryteria kwalifikacji do programu spełniają tylko te osoby, które przez 6 miesięcy przed przystąpieniem do niego nie były zatrudnione na podstawie umowy na czas nieokreślony. Celem ewaluacji było oszacowanie wpływu programu na udział nowych umów o pracę na czas nieokreślony we wszystkich umowach zarejestrowanych w 2015 r. Wykorzystano metodę różnicy w różnicach, przy pomocy której porównano umowy zarejestrowane w rejestrze zatrudnienia (SISCO) między 1 stycznia a 31 grudnia 2015 r. z tymi, które zarejestrowane zostały w poprzednich latach. Zdefiniowano dwie grupy – eksperymentalną, w której znalazły się osoby spełniające kryteria kwalifikacji programu, tj. takie, które zostały zatrudnione w latach 2014–2015, a w ciągu 6 miesięcy poprzedzających to zatrudnienie nie zakończyły pracy w oparciu o umowę na czas nieokreślony. Grupa kontrolna obejmowała osoby zatrudnione w latach 2014–2015 r., które w ciągu 6 miesięcy poprzedzających to zatrudnienie zakończyły pracę w oparciu o umowę na czas nieokreślony, a tym samym nie kwalifikowały się do programu (ilustracja metody różnicy w różnicach – zob. rys. 3–2).

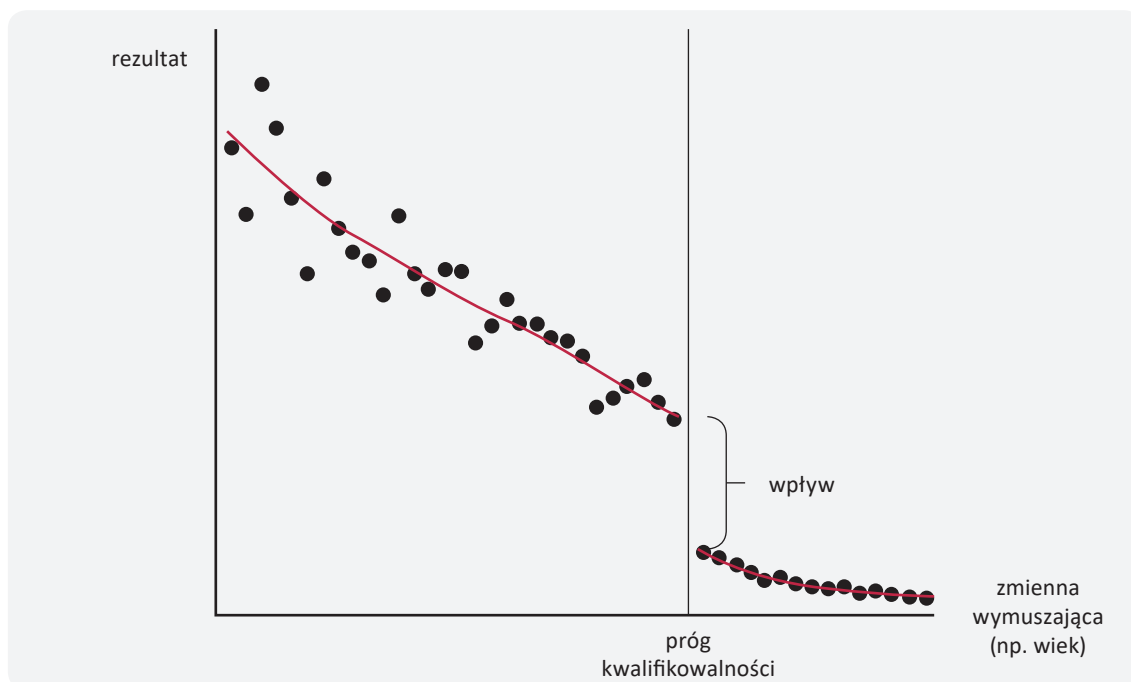
Rys. 3-2. Metoda różnicy w różnicach (DiD)



Ponieważ baza danych nie dostarcza bezpośrednich informacji na temat subsydiowania i umożliwia zidentyfikowanie tylko uczestników, którzy kwalifikują się do programu, ewaluacja mierzy efekt programu tylko dla tej grupy (zamiar oddziaływania, efekt ITT).

Ponieważ „Gwarancje dla młodzieży” są skierowane do młodych osób, większość programów posiada ograniczenia wiekowe. Prostim rozwiązaniem jest wykorzystanie tego kryterium, tak jak zrobiono to w jednej z **polskich** ewaluacji. 12-miesięczny program zatrudnienia subsydiowanego był przeznaczony wyłącznie dla osób bezrobotnych w wieku do 30 lat. Pozwoliło to na wykorzystanie kryterium wieku do zastosowania modelu regresji nieciągłej (dalej: RDD) (zob. rys. 3–2). Istnieją też inne aspekty instytucjonalne, które różnicują interwencje dla osób poniżej i powyżej 30 r.ż. Po pierwsze: bezrobotni w wieku poniżej 30 lat mają dostęp do większej liczby form wsparcia (m.in. bonów szkoleniowych, bonów na zasiedlenie czy bonów stażowych) i otrzymują je na innych zasadach (np. staż może trwać do 12 miesięcy w przypadku młodszych osób i do 6 miesięcy w przypadku starszych). Po drugie, również finansowanie jest związane z kryterium wieku. Programy ze środków UE skierowane do młodszych osób bezrobotnych są finansowane z budżetu programu operacyjnego zarządzanego centralnie, a te dla starszych z budżetu programów regionalnych zarządzanych na szczeblu wojewódzkim. W efekcie między programami mogą pojawiać się znaczące różnice – zarówno w zakresie dostępności, jaki i zasad realizacji. Strategia identyfikacyjna opiera się na założeniu, że przed wprowadzeniem programu zatrudnienia subsydiowanego, już istniały wszystkie pozostałe różnice między grupą eksperymentalną i kontrolną. Można więc połączyć model regresji nieciągłej z metodą różnicy w różnicach. Rozwiązanie to ma jednak jedną wadę – wyniki opierają się tylko na porównaniu grup, których członkowie są w wieku zbliżonym do wyznaczonej granicy wieku. Wymaga to dużej ostrożności podczas generalizacji wniosków na inne grupy wieku.

Rys. 3-3. Ilustracja: model regresji nieciągłej (RDD)

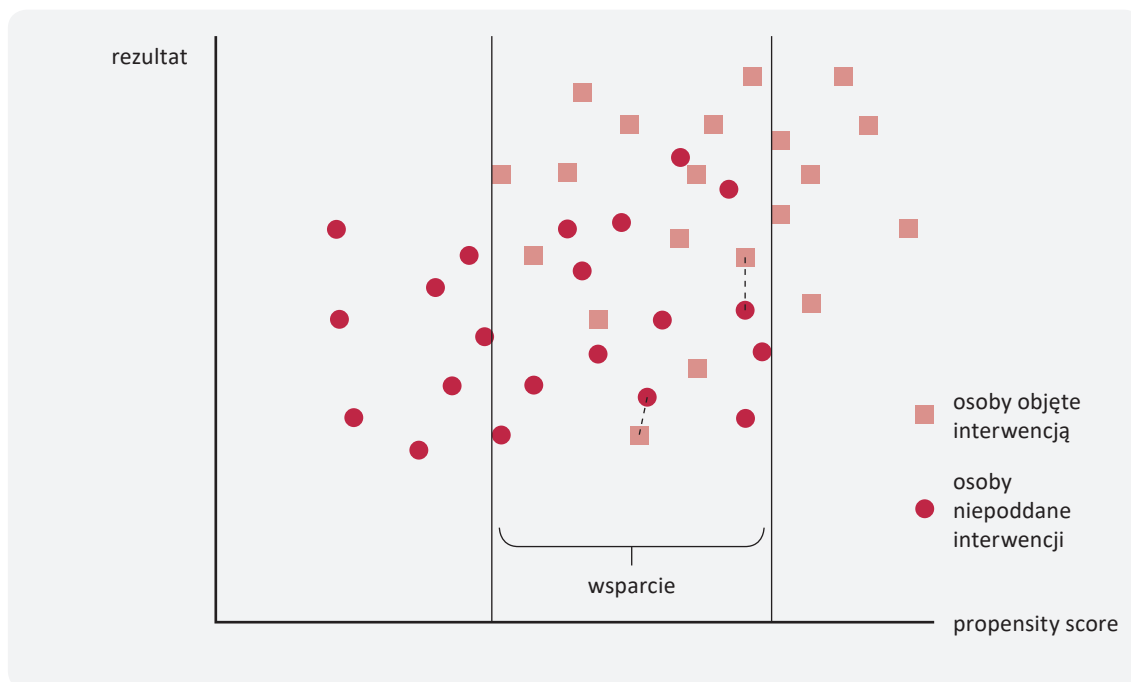


- *Różnice regionalne dotyczące intensywności działań, daty rozpoczęcia programu, wartości docelowych itp.*

Na **Węgrzech** program „Gwarancje dla młodzieży” w regionie Węgier środkowych rozpoczął się w październiku 2015 r., 9 miesięcy później niż w pozostałych regionach. Powodem opóźnienia był fakt, że ta część Węgier nie była objęta celem Konwergencji Funduszu Strukturalnego UE. W konsekwencji występowały pewne różnice w zakresie finansowania programu. Opóźnienie wynikało z przyczyn administracyjnych i może być uważane za egzogeniczne. Program realizowany w regionie Węgier środkowych miał identyczne warunki kwalifikacji i zasady, jak programy w innych regionach. Rozpoczęcie programu w późniejszym terminie umożliwiło zastosowanie metody różnicy w różnicach. Grupa eksperymentalna składała się z osób bezrobotnych mieszkających w regionach, w których program rozpoczął się wcześniej. Grupa kontrolna składa się z bezrobotnych, mieszkających w regionie Węgier środkowych, w którym program rozpoczął się z opóźnieniem. Grupę kontrolną i eksperymentalną porównano w okresie 9 miesięcy przed i po dacie rozpoczęcia programu w regionach objętych celem konwergencji, tj. w styczniu 2015 r. Oczwistym ograniczeniem tej strategii jest krótki okres obserwacji.

Alternatywną strategią identyfikacji, która łagodzi problem krótkiego okresu obserwacji, może być metoda zmiennych instrumentalnych. Może się ona opierać na specyficznych warunkach implementacji polityk wdrażanych w określonych regionach lub przez niektóre PSZ, które odbiegają od ogólnie ustalonych zasad implementacji (Maestas Mueller, 2013). Zidentyfikowanie takich odstępstw od ustalonych reguł może posłużyć jako zmienna instrumentalna.

Rys. 3-4: Ilustracja: propensity score matching



Główną strategią przy korygowaniu problemu selekcji jest szeroki zestaw zmiennych obejmujący zarówno dla uczestników programu, jak i osoby kwalifikujące się do niego, ale w nim nieuczestniczące. Następnym krokiem jest utworzenie grupy osób nieuczestniczących, które możliwie najbardziej przypominają osoby uczestniczące w programie. W tym przypadku strategia identyfikacyjna oparta jest na założeniu, że obserwujemy wszystkie czynniki, które wpływają na prawdopodobieństwo udziału w programie. Innymi słowy zakładamy, że kontrolujemy wszystkie istotne różnice w zmiennych obserwowalnych pomiędzy grupą kontrolną a eksperymentalną. Niemniej jednak problemem może być wpływ zmiennych nieobserwowalnych, np. motywacji i zdolności, na osiągnięte rezultaty. Przykładowo: bardziej zmotywowana młoda osoba ma większą szansę zgłosić się do programu, ale prawdopodobnie jej perspektywy zawodowe są i tak lepsze, nawet bez programu. Podobny argument może dotyczyć innych umiejętności i zdolności, które są trudne do zidentyfikowania na podstawie zmiennej dotyczącej poziomu wykształcenia. Badacz może starać się znaleźć zmienne, które wskazują na nieobserwowalne zdolności i motywację, takie jak wyższe wyniki w testach kompetencyjnych, historia zatrudnienia lub udział w programach szkoleniowych. Co do zasady, te nieobserwowalne czynniki wpływają na przeszacowanie efektu i uzyskany w ten sposób wynik powinien być traktowany jako mieszczący się w górnym przedziale.

Na **Węgrzech** 90-dniowy program zatrudnienia subsydiowanego jest porównywany również do innych programów: szkoleń i robót publicznych. Założenie jest takie, że problemy związane z cechami nieobserwowalnymi powinny być mniejsze, jeżeli grupa kontrolna składa się z uczestników innych programów zamiast z osób nieuczestniczących w żadnym programie. Ewaluacja w tym podejściu pokazuje relatywną efektywność danego programu w porównaniu do innych. Na Węgrzech nie można ewaluować programu, porównując osoby biorące w nim udział z tymi, które nie brały udziału w żadnym programie, ponieważ PSZ starają się włączyć każdą osobę zarejestrowaną jako bezrobotna do jakiegoś programu. Z tego wynika, że pula młodych osób, które nie

uczestniczyły w żadnym programie, ale figurują w rejestrze, prawdopodobnie nie stanowi trafnej grupy kontrolnej, ponieważ są to osoby albo o najniższej motywacji, albo o najwyższych szansach na ponowne zatrudnienie. Spodziewano się, że nawet uczestnicy programu robót publicznych będą się znacznie różnić od uczestników innych programów pod względem perspektyw na rynku pracy i poziomu motywacji. Te nieobserwowalne czynniki zostały uwzględnione w analizie dzięki wykorzystaniu danych o wykształceniu i historii zatrudnienia. W drugiej polskiej ewaluacji wykorzystano tę samą strategię do oceny relatywnej efektywności stażu, bonu stażowego, szkolenia, bonu szkoleniowego, prac interwencyjnych i robót publicznych.

W **hiszpańskiej** ewaluacji umów o staż porównano rezultaty uczestników programu i osób, które kwalifikowały się do programu, ale nie zostały nim objęte. Grupa eksperymentalna składa się z osób w wieku poniżej 30 lat, które podpisały umowę o staż, a grupa kontrolna z młodych osób bezrobotnych, które mogły podpisać umowę o staż (spełniały kryteria kwalifikacji), ale podpisały zwykłe umowy na czas określony. Dane z rejestrów ubezpieczeń społecznych pozwoliły na zrekonstruowanie historii zatrudnienia, na podstawie takich zmiennych, jak staż pracy, liczba wcześniejszych okresów zatrudnienia, liczba wcześniejszych pracodawców, okres bezrobocia. Zmienne te pomogły zidentyfikować istotne różnice między grupą kontrolną a eksperymentalną. Próba została ograniczona jedynie do osób, które dopiero wchodzi na rynek pracy. Wynika to z tego, że celem umowy o staż jest pomoc w zdobyciu pierwszego doświadczenia zawodowego. Aby ustalić, czy efektywność stażu jest uzależniona od wcześniejszej sytuacji na rynku pracy, porównano osoby, które podpisały umowę o staż i nie miały doświadczenia zawodowego, z osobami, które podpisały umowę o staż i posiadały doświadczenie zawodowe.

We **Włoszech** w ramach ewaluacji szacowano prawdopodobieństwo zatrudnienia na umowę na czas nieokreślony pomiędzy pracownikami kwalifikującymi się do programu a pracownikami niespełniającymi kryteriów kwalifikacji. Wykorzystano metodę różnicy w różnicach, zakładając, że program nie wpływa na rezultat w grupie kontrolnej. Innymi słowy, program (zachęty do zatrudnienia w oparciu o umowę na czas nieokreślony) mający na celu zwiększenie prawdopodobieństwa zatrudnienia pracowników, którzy mogliby z niego skorzystać (spełniają kryteria kwalifikacji) nie ma wpływu na prawdopodobieństwo zatrudnienia pracowników, którzy nie spełniają kryteriów kwalifikacji. Statystyki opisowe pokazują, że założenie to się nie zostało spełnione, ponieważ w 2015 r. zaobserwowano nagły spadek zatrudnienia wśród osób niekwalifikujących się do programu. Wynikać to mogło ze zmiany nastawienia niektórych pracodawców, którzy szacując koszty i korzyści związane z wprowadzeniem programu, zaczęli preferować pracowników, którzy spełniają kryteria kwalifikacji do programu, choć mają mniejsze doświadczenie zawodowe, kosztem osób, które mają bardziej stabilną historię zatrudnienia, ale nie kwalifikują się do programu¹¹. Innymi słowy, samo wprowadzenie programu mogło mieć wpływ na grupę kontrolną. Zostało to skorygowane przy zastosowaniu dwuetapowej procedury. Na pierwszym etapie oszacowano teoretyczne wartości zmiennych wynikowych – wartości, które zostałyby osiągnięte bez wprowadzenia programu. Następnie zastąpiono wartości zmiennych wynikowych w grupie kontrolnej wartościami teoretycznymi. Na drugim etapie, bazując na skorygowanych danych, oszacowano efektywność programu dla różnych grup wieku, wykorzystując model najmniejszych kwadratów. Aby wyodrębnić wpływ programu na osoby młode, efekt oszacowano, wykorzystując interakcje zmiennych definiujących grupę wieku ze zmienną wskazującą, czy dana osoba otrzymała wsparcie.

11 Jak wyjaśniono w następnej części, ta kwestia staje się jeszcze bardziej złożona ze względu na to, że zmiana kosztu alternatywnego była nierówna w poszczególnych grupach wiekowych.

3.4. Heterogeniczność efektów

Ważnym aspektem ewaluacji programu jest analiza, jak wpływa on na różne grupy. Mogą być one wyróżniane ze względu na wiek, płeć, miejsce zamieszkania, uwarunkowania makroekonomiczne, rok realizacji programu czy jego wersję. Zrozumienie heterogeniczności efektów pomaga też zrozumieć i ocenić:

- trafność zewnętrzną wyników i porównywalność z innymi wynikami ewaluacji,
- efektywność kosztów i korzyści programu,
- słabe strony programu i możliwości jego poprawy,
- oddziaływanie ze względu na płeć.

Wyniki programu mogą być różne w zależności od poziomu wykształcenia jego uczestników. Decydenci prawdopodobnie z zadowoleniem przyjmą informację, że subsydia do zatrudnienia mają większy wpływ na osoby z niższym wykształceniem, ponieważ – przypuszczalnie – ta grupa osiąga gorsze rezultaty na rynku pracy. Jeśli ewaluacja wykazała, że subsydia mają pozytywny wpływ na osoby z wykształceniem średnim lub niższym, ale nie na absolwentów szkół wyższych, program subsydiów powinien być skierowany do tej pierwszej grupy. Ponadto, ponieważ koszty interwencji przeznaczonej dla osób dobrze wykształconych, pracujących na lepiej płatnym stanowisku, są przypuszczalnie wyższe, ukierunkowanie działań na osoby z niższym wykształceniem może poprawić efektywność kosztową programu. Na **Węgrzech** program 90-dniowego zatrudnienia subsydiowanego, w porównaniu do programu robót publicznych, silniej oddziaływał na uczestników z **wykształceniem** podstawowym niż na osoby z wykształceniem średnim lub wyższym. Było to zgodne z oczekiwaniami, ponieważ młode osoby z niskim wykształceniem mają mniejsze szanse na zatrudnienie. Jednocześnie uczestnicy programu zatrudnienia subsydiowanego są średnio lepiej wykształceni niż uczestnicy programu robót publicznych. Pokazuje to, że do programu subsydiów kwalifikowane są osoby bezrobotne o największym potencjale zatrudnienia. Kierowanie do niego w pierwszej kolejności osób bezrobotnych z niższym wykształceniem mogłoby zwiększyć średni efekt oddziaływania tego programu i jednocześnie zmniejszyć efekt zbędnej straty społecznej.

Ewaluacja może również pokazać, że czas trwania programu ma wpływ na jego efektywność. Taki wniosek może pomóc określić optymalną długość podobnego programu w kolejnej edycji.

Analiza **hiszpańska** sprawdza wpływ **płci** uczestników na efektywność programu. Wniosek: bezrobocie po stażu jest mniej prawdopodobne w przypadku kobiet niż mężczyzn. W innych analizowanych kwestiach nie stwierdzono jednak statystycznych różnic między mężczyznami a kobietami.

Ewaluacja **włoska** wykazała istotne różnice w efektywności programu w zależności od grupy **wieku**. Metoda różnic w różnicach pokazała, że program zwiększył udział umów o pracę na czas nieokreślony wśród wszystkich umów zarejestrowanych w 2015 r. Udział ten zmniejszał się wraz z wiekiem uczestników.

Ewaluacja w Polsce wykazała, że szkolenia i bony szkoleniowe sprawdzają się stosunkowo dobrze w przypadku mężczyzn, ale nie w przypadku kobiet. Roboty publiczne są mniej efektywne w przypadku osób z niższym wykształceniem i osób mieszkających w regionach o wysokiej stopie bezrobocia. Bony stażowe są bardziej skuteczne niż standardowy staż niezależnie od płci, poziomu wykształcenia czy miejsca zamieszkania osób bezrobotnych.

4. KWESTIE ISTOTNE DLA DECYDENTÓW

Dotychczas omówiliśmy zagadnienia umożliwiające badaczom oszacowanie efektu przyczynowego interwencji. Istnieją jednak dodatkowe zagadnienia, które są szczególnie interesujące dla decydentów. W tej części ogólnie je omówimy oraz pokażemy, jakie zbiory danych mogą być pomocne podczas analizy tych zagadnień.



„Moje pytanie brzmi: czy to co robimy ma jakikolwiek wpływ?”

4.1. Interpretacja wyników

Bardzo ważne jest jasne określenie, w jaki sposób interpretować wyniki ewaluacji. Często odnoszą się one do uczestników programu lub jakiejś podgrupy uczestników. Prezentując wyniki decydentom, należy podkreślić, że przy pewnych założeniach, można rozciągnąć wyniki na szerszą populację. Badacze powinni rozważyć również zgromadzenie danych na temat większej grupy, aby umożliwić generalizację. Należy również mieć na uwadze, że ewaluacje, które oceniają zamiar oddziaływania (ITT) lub średni efekt oddziaływania (ATE), czyli efekt oddziaływania na wszystkie osoby spełniające kryteria kwalifikacji, a nie tylko uczestników danego programu – mogą być bardziej użyteczne dla decydentów.

4.2. Powiązanie wyników z zasadami programu i praktyką jego wdrażania

Wychodząc poza ilościową ocenę efektów programu, ewaluacja powinna prowadzić do wniosków na temat czynników mających wpływ na wyniki. Na przykład: jeżeli z ewaluacji wynika, że

program nie wykazuje żadnego istotnego efektu w kontekście zatrudnienia, ważna jest wiedza o tym, czy niepowodzenie jest spowodowane złymi założeniami programu (nieodpowiednia grupa docelowa, źle zdefiniowane kryteria kwalifikacji, niewystarczające zasoby finansowe i kadrowe), czy problemami podczas realizacji (niewystarczające rozpowszechnienie informacji o programie, arbitralny wybór uczestników programu itp.).

4.3. Trafność zewnętrzna, porównywalność z innymi badaniami i zbiorczymi statystykami

Odbiorcy badań ewaluacyjnych, zainteresowani tworzeniem polityk publicznych, mogą chcieć porównywać efektywność programów poddanych analizie z innymi programami. Mogą chcieć dowiedzieć się, między innymi, jaka jest skuteczność podobnych programów dla młodych osób w innych regionach/krajach, jaka jest efektywność innych programów skierowanych do tej samej grupy docelowej czy jaka byłaby efektywność ewaluowanego programu wśród innych grup docelowych. Żeby odpowiedzieć na te pytania, badacze muszą być świadomi kilku kwestii. Po pierwsze, muszą odpowiedzieć sobie na pytanie, w jaki sposób zaprezentować wyniki, aby można było porównać je z innymi badaniami. Po drugie, muszą zwrócić uwagę na ograniczenia związane z interpretacją wyników i wytłumaczyć, w jaki sposób mogą być generalizowane w innych warunkach, aby uniknąć prostych ekstrapolacji. Po trzecie: badacze muszą także wyraźnie udokumentować istotne zasady programów objętych ewaluacją czy uwarunkowania ekonomiczne, w jakich były wdrażane, aby inni interesariusze mogli zorientować się, w jaki sposób czynniki te mogły wpłynąć na efektywność programów. Na przykład: wszystkie programy objęte ewaluacją na Węgrzech, w Włoszech i w Polsce zostały uruchomione, gdy sytuacja gospodarcza się poprawiała, a w Hiszpanii odbywały się one zarówno w okresie wzrostu gospodarczego, jak i podczas recesji.

Interpretując i prezentując wyniki, powinno się mieć na uwadze, że wskaźniki mogą mieć różne definicje, a co za tym idzie wartości liczbowe mogą różnić się w zależności od kontekstu. Na przykład: analiza może koncentrować się na zarejestrowanych osobach bezrobotnych, więc liczba osób zarejestrowanych jako bezrobotne może znacznie się różnić od liczby osób bezrobotnych, określanej w oparciu o definicję bezrobocia, stosowaną np. w badaniu Aktywności Ekonomicznej Ludności. Interpretacja powinna uwzględniać te różnice i dokładnie je wyjaśniać.

4.4. Efekt wypierania i zbędna strata społeczna

Aby uzyskać pełniejszy obraz efektywności programów, należy podkreślić dwie kwestie. Po pierwsze: efekt wypierania ma miejsce wtedy, gdy dane działanie pomaga w zatrudnieniu młodych osób, ale zmniejsza szanse tych, które mają podobne umiejętności, ale nie mają dostępu do tego działania. Po drugie: program generuje zbędne straty społeczne, jeżeli dane działanie zwiększa szanse na zatrudnienie, ale znacząca część wsparcia nie jest potrzebna, ponieważ pracodawcy i tak zatrudniliby te konkretne osoby. Dostarczenie dowodów na potwierdzenie tych wniosków nie jest jednak łatwe i wymaga posiadania dodatkowych informacji na temat innych pracowników w tych samych i konkurencyjnych firmach – w idealnej sytuacji danych z dopasowaniem pracowników do firm, które ich zatrudniają.

4.5. Analiza kosztów i korzyści oraz efektywności kosztowej

W omawianej dotychczas metodzie ewaluacji skupiliśmy się na analizie efektywności programu w odniesieniu do sytuacji uczestników na rynku pracy po jego zakończeniu. Ewaluacja powinna jednak uwzględniać również koszty programu. Jest to informacja istotna dla decydentów.

Istnieją dwa rodzaje analiz, za pomocą których można oszacować koszty i korzyści programu. Analiza efektywności kosztowej porównuje koszty programu z rezultatami wyrażonymi w jednostkach fizycznych, na przykład: program stworzył jeden pełny etat za x euro. W analizie kosztów i korzyści staramy się ocenić, czy całkowite korzyści przeważają nad całkowitymi kosztami. Naturalnie ta druga analiza stanowi większe wyzwanie, ponieważ, aby oszacować korzyści, trzeba mieć więcej informacji.

Analiza efektywności kosztowej może być sporym wyzwaniem, jeżeli baza danych administracyjnych nie zawiera informacji o kosztach programu, co często się zdarza. Badacz może korzystać z danych zbiorczych, pochodzących z innych źródeł, aby oszacować jednostkowe koszty uczestnictwa w programie. Alternatywnie, w przypadku takich programów, jak zatrudnienie subsydiowane, znając zasady programu i posiadając informacje o wynagrodzeniach pochodzące z powiązanej bazy danych na temat ubezpieczeń społecznych, można obliczyć szacunkowe koszty programu. W każdym przypadku badacze powinni planować pozyskanie tego rodzaju danych już na początku ewaluacji.

BIBLIOGRAFIA

Abadie, Alberto and Cattaneo, Matias D. (2018). Econometric Methods for Program Evaluation (August 2018). Annual Review of Economics, Vol. 10, pp. 465-503, 2018
<http://dx.doi.org/10.1146/annurev-economics-080217-053402>

Angrist J.D., Pischke J.S. (2015). Mastering Metrics: The Path from Cause to Effect, Princeton University Press, Princeton.
<https://doi.org/10.1111/j.1475-4932.2011.00742.x>

Angrist, Joshua and Pischke, Jörn-Steffen (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Caliendo M., Mahlstedt R., Mitnik O. (2017). Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labor market policies. Labour Economics, 46, 14-25.
<https://doi.org/10.1016/j.labeco.2017.02.001>

Card D., Krueger A.B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania, American Economic Review, 84, n. 4, pp. 772–793.
<http://davidcard.berkeley.edu/papers/njmin-aer.pdf>

European Commission (2013). Evalsed Sourcebook, Methods and Techniques
https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf

European Commission (2020). How to use administrative data for European Social Funds counterfactual impact evaluations, Directorate-General for Employment Social Affairs and Inclusion, Directorate G.
<https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8303>

Gertler, Paul J.; Martinez, Sebastian; Premand, Patrick; Rawlings, Laura B.; Vermeersch, Christel M. J. (2016). Impact evaluation in practice (English). Washington, D.C. : World Bank Group.
<http://documents.worldbank.org/curated/en/698441474029568469/Impact-evaluation-in-practice>

Imbens, Guido, and Thomas Lemieux (2008). Regression Discontinuity Designs: A Guide to Practice, Journal of Econometrics, 142, 615-635.
<https://doi.org/10.1016/j.jeconom.2007.05.001>

Leeuw, Frans & Vaessen, Jos (2009). Impact Evaluations and Development: NONIE Guidance on Impact Evaluation. Independent Evaluation Group, World Bank.
<https://www.oecd.org/development/evaluation/dcdndep/47466906.pdf>

Morris, Stephen & Tödting-Schönhofer, Herta & Wiseman, Michael (2013). Design and Commissioning of Counterfactual Impact Evaluations - A Practical Guidance for ESF Managing Authorities. Publisher: European Commission, Directorate-General for Employment, Social Affairs and Inclusion. DOI:10.2767/94454

UNDP (2009). Handbook on Planning Monitoring and Evaluating for Development Result.
<http://web.undp.org/evaluation/handbook/documents/english/pme-handbook.pdf>

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
<https://doi.org/10.1093/biomet/70.1.41>

ZAŁĄCZNIK 1: SKRÓCONY OPIS EWALUACJI W CZTERECH KRAJACH

WŁOCHY



Program:

- 1) Obniżenie kosztów zwolnienia w przypadku umów na czas nieokreślony (umowa z gradacją ochrony).
- 2) Niższe składki na ubezpieczenia społeczne dla nowych umów na czas nieokreślony i w przypadku zmiany umowy na czas określony na umowę na czas nieokreślony.

Ewaluacja wspólna obu programów.

Dane: na temat umów zarejestrowanych z bazy danych ministerstwa pracy i polityki społecznej (SISCO), pozyskane z informacji wysyłanej przez pracodawców.

Rezultat: liczba nowych umów na czas nieokreślony w stosunku do całkowitej liczby umów zarejestrowanych w 2015 r.

Strategia identyfikacyjna: metoda różnicy w różnicach.

Grupa spełniająca kryteria kwalifikacji:

- Osoby zatrudnione na umowę na czas nieokreślony w latach 2014–2015.
- Wszyscy pracodawcy bez określenia konkretnej branży lub regionu geograficznego, w tym stowarzyszenia i przedsiębiorstwa publiczne, ale z wyłączeniem administracji publicznej i rolnictwa.
- Pracodawcy bez zaległości w składkach i z uregulowaną sytuacją prawną.

Grupa niespełniająca kryteriów kwalifikacji:

- Osoby zatrudnione w 2015 r., które w okresie 6 miesięcy przed rozpoczęciem nowej umowy na czas nieokreślony, zostały zwolnione z pracy, w której pracowały na umowę na czas określony.
- Osoby zatrudnione w 2015 r., które w okresie 3 miesięcy poprzedzających reformę (01.10.2014 – 31.12.2014) nie zostały zwolnione przez tego samego pracodawcę, który ich zatrudnił lub przez innego pracodawcę powiązanego z tym poprzednim.

Wykorzystano interakcje zmiennych definiujących grupę wieku ze zmienną wskazującą, czy dana osoba otrzymała wsparcie, w celu odseparowania wpływu programu na osoby młode.

WĘGRY



Program: ewaluacja węgierska obejmuje program 90-dniowego zatrudnienia subsydiowanego, stworzonego dla osób w wieku do 25 lat. Subsidia w wysokości 100% kosztów wynagrodzenia przyznawane są na 90 dni bez zobowiązania do dalszego zatrudnienia. Program rozpoczął się w 2015 r. w ramach programu „Gwarancje dla młodzieży”.

Dane: ewaluacja oparta jest na dwóch bazach danych administracyjnych. Dane PSZ zawierają informacje o wszystkich zarejestrowanych osobach bezrobotnych i uczestnikach programu. Dane PSZ są powiązane ze zbiorem danych administracyjnych instytucji przyznającej świadczenia społeczne, który zawiera informacje o statusie zatrudnienia i historii zatrudnienia około połowy populacji.

Rezultaty:

głównym rezultatem jest status zatrudnienia po 6 i 12 miesiącach od zakończenia programu oraz poziom wynagrodzenia.

Strategia identyfikacyjna: ewaluacja stosuje dwie różne strategie identyfikacyjne. W pierwszej porównywane są rezultaty uczestników programu zatrudnienia subsydiowanego i uczestników innych programów (robót publicznych i szkoleń), przy zastosowaniu techniki propensity score matching.

Druga ocena opiera się na metodzie różnicy w różnicach i wykorzystuje fakt rozpoczęcia programu w regionie Węgier środkowych o 9 miesięcy później niż w innych częściach kraju z powodów administracyjnych.

POLSKA (EWALUACJA NR 1)



Program: program zatrudnienia subsydiowanego dla osób bezrobotnych w wieku do 30 lat. Funkcjonował w latach 2016–2018. Subsydia po stronie pracodawcy do wysokości wynagrodzenia minimalnego plus ubezpieczenia społeczne były wypłacane przez 12 miesięcy, a pracodawcy byli zobowiązani do przedłużenia zatrudnienia na kolejne 12 miesięcy po zakończeniu subsydiowania.

Osoby spełniające kryteria kwalifikacji: osoby zarejestrowane jako bezrobotne w wieku do 30 lat z ustalonym I lub II profilem (III profil nie spełniał kryteriów kwalifikacji). Ponadto do programu nie kwalifikowali się pracodawcy, którzy redukowali zatrudnienie ze swojej inicjatywy w okresie ostatnich 6 miesięcy.

Dane: dane z rejestru powiatowych urzędów pracy z pełną historią rejestracji osób bezrobotnych, umożliwiające identyfikację uczestników programu, niepowiązane z danymi dotyczącymi ubezpieczeń społecznych.

Próba: osoby, które zostały bezrobotne między 2015 a 2016 rokiem, z wyłączeniem osób, które były bezrobotne lub uczestniczyły w aktywnej polityce rynku pracy w okresie ostatnich 6 miesięcy przed rejestracją – aby przeanalizować nowe rejestracje, niebędące częścią dłuższego okresu bezrobocia. Aby móc zaobserwować rezultat po 36 miesiącach od rejestracji, ograniczono próbę do osób, które zarejestrowały się jako bezrobotne do kwietnia 2016 r.

Rezultaty: analizujemy dwa rodzaje rezultatów: wykreślenie z rejestru po 12–36 miesiącach od rejestracji i łączna liczba dni poza rejestrem po 12–36 miesiącach od rejestracji.

Strategia identyfikacyjna: RDD z wykorzystaniem progu wieku w połączeniu z DiD.

POLSKA (EWALUACJA NR 2)



Program: aktywne polityki rynku pracy oferowane młodym osobom: staż, szkolenie, prace interwencyjne, roboty publiczne, bon stażowy, bon szkoleniowy.

Osoby spełniające kryteria kwalifikacji: osoby w wieku do 30 lat, zarejestrowane jako bezrobotne. Dane: dane z rejestru powiatowych urzędów pracy z pełną historią rejestracji osób bezrobotnych, umożliwiające identyfikację uczestników danego programu, niepowiązane z danymi dotyczącymi ubezpieczeń społecznych.

Próba: ograniczono próbę do ukończonych aktywnych polityk rynku pracy, które rozpoczęły się w latach 2015–2016 i trwały przez co najmniej jeden dzień. Włączono tylko uczestników, którzy mieli od 18 do 29 lat w momencie rozpoczęcia aktywnej polityki rynku pracy, i wyłączono osoby, które w przeszłości odbywały karę pozbawienia wolności. Analizowano 6 rodzajów interwencji: staż, szkolenie, prace interwencyjne, roboty publiczne, bon stażowy, bon szkoleniowy. Całość próby to 319 610 obserwacji.

Rezultaty: wykreślenie z ewidencji bez objęcia aktywną polityką rynku pracy w 12–36 miesięcy od rejestracji.

Strategia identyfikacyjna: porównanie uczestników programu w parach, z zastosowaniem techniki PSM.

SPAIN



Program: umowa o staż. Program został wdrożony w Hiszpanii w 1998 r. Według ustawy jego głównym celem jest wzrost stabilności zatrudnienia, rozwinięcie umiejętności i kwalifikacji zawodowych zgodnych z poziomem i kierunkiem wykształcenia oraz stworzenie instrumentów zachęcających firmy, aby zatrudniały młodych pracowników po zakończeniu stażu.

Osoby spełniające kryteria kwalifikacji: ograniczenie kryteriów kwalifikacji dotyczy wykształcenia i wieku: wymagane było uzyskanie licencjatu lub wykształcenia zawodowego i wiek poniżej 30 lat lub uzyskanie dyplomu na mniej niż 5 lat przed rozpoczęciem udziału w programie.

Dane: dane o historii zatrudnienia (Continuous Sample of Work Histories – CSWH). Baza składa się z rejestrów ubezpieczeń społecznych dla 5% populacji Hiszpanii. Zawiera pełną historię rynku pracy dla reprezentatywnej próby pracowników w Hiszpanii.

Próba: osoby, które zostały bezrobotne od 2002 r. w wieku poniżej 30 lat, z wyższym wykształceniem, tj. potencjalni stażyści.

Rezultaty: analizowano dwa rodzaje rezultatów: prawdopodobieństwo pozostania w firmie po umowie stażu oraz prawdopodobieństwo podpisania umowy stałej po stażu, w szczególności w tej samej firmie.

Strategia identyfikacyjna: do celów pierwszej analizy opracowano wielomianową regresję logistyczną, w której zmienna zależna przyjmuje trzy różne wartości, to jest: przejście do innej firmy, pozostanie w firmie oraz utrata pracy. W drugim kroku, przy zastosowaniu modelu logitowego, oceniano oddzielnie prawdopodobieństwo podpisania umowy stałej (w porównaniu do umowy czasowej) w zależności od tego, czy uczestnik pozostał w firmie, czy przeszedł do innej firmy.

ZAŁĄCZNIK 2: WYKAZ SKRÓTÓW I POJĘĆ

ALMP: aktywne polityki rynku pracy (active labour market policy).

ATE, ATT: średni efekt oddziaływania (Average Treatment Effect) i średni efekt oddziaływania na jednostki poddane oddziaływaniu (Average Treatment Effect on Treated). Metody te polegają na porównaniu sytuacji kontrfaktycznej, która nie zaistniała w przypadku uczestników programu (tzn. sytuacji, w której nie byłyby one poddane oddziaływaniu), z sytuacją osób z grupy kontrolnej. Metody te wskazują, jaka jest różnica między średnimi wynikami osiąganymi przez osoby poddane oddziaływaniu i osoby przypisane do grupy kontrolnej. Istnieją dwie główne koncepcje średniego efektu oddziaływania. ATE pokazuje oczekiwanie średniej różnicy oddziaływania w populacji w parze potencjalnych rezultatów, z uśrednieniem w całej populacji będącej przedmiotem zainteresowania. Działanie jest istotne, jeżeli cała populacja jest objęta rozważaną polityką.

$$ATE = E(Y_i(1) - Y_i(0))$$

Gdzie $Y_i(1)$ jest rezultatem jednostki i poddanej oddziaływaniu, a $Y_i(0)$ jest rezultatem jednostki i niepoddanej oddziaływaniu.

ATT pokazuje średni efekt oddziaływania na subpopulację jednostek poddanych oddziaływaniu:

$$ATE = E(Y_i(1) - Y_i(0) | D_i = 1)$$

CIE: ewaluacja efektu kontrfaktycznego (Counterfactual Impact Evaluation).

CO, COB: rejestr obowiązkowych zgłoszeń, prowadzony przez Ministerstwo Pracy i Polityki Społecznej we Włoszech (Comunicazioni Obbligatorie *on-line*).

CSWH: dane o historii zatrudnienia (Continuous Sample of Work History). Próba stanowiąca 5% hiszpańskich rejestrów ubezpieczenia społecznego.

Diff in Diffs, DiD: metoda różnicy w różnicach (Difference in Differences). Quasi-eksperymentalna metoda oceny, w której efekt programu jest obliczany poprzez porównanie grupy eksperymentalnej i grupy kontrolnej przed i po oddziaływaniu. Efekt oddziaływania to średnia zmiana w czasie w zakresie zmiennej wynikowej dla grupy eksperymentalnej minus średnia zmiana dla grupy kontrolnej. Metoda DiD opiera się na założeniu, że przy braku oddziaływania rezultat grupy eksperymentalnej byłby zgodny z rezultatem jednostek niepoddanych oddziaływaniu, tj. rezultaty dla obu grup byłyby równoległe (aby uzyskać bardziej szczegółowe informacje, zob. np. Card i Krueger, 1994, Angrist i Pischke, 2008).

ISCED: Międzynarodowa Standardowa Klasyfikacja Edukacji (International Standard Classification of Education), https://ec.europa.eu/education/international-standard-classification-of-education-isced_it

INAPP: Instytut analiz polityk publicznych (Istituto Nazionale per l'Analisi delle Politiche Pubbliche), działający przy ministerstwie pracy i polityki społecznej Włoch.

ITT: zamiar poddania oddziaływaniami (Intention To Treat). Pokazuje oddziaływanie polityki na populację spełniającą kryteria kwalifikacji. Innymi słowy, pokazuje nam efekt przyczynowy oferty interwencji. Jeżeli nie wszystkie osoby należące do populacji spełniającej kryteria kwalifikacji zostaną poddane interwencji, ITT będzie różnić się od średniego efektu oddziaływania.

NACE: Statystyczna Klasyfikacja Działalności Gospodarczej we Wspólnocie Europejskiej (Statistical Classification of Economic Activities in the European Community), wer. 2 (2008), wer. 2 (NACE wer. 2) https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2&StrLanguageCode=IT&IntPcKey=&StrLayoutCode=HIERARCHIC

OLS: metoda najmniejszych kwadratów (Ordinary Least Squares).

PSZ: publiczne służby zatrudnienia.

PSM: Propensity score matching – metoda porównująca zmienną wynikową dla osób poddanych oddziaływaniami ze zmienną wynikową dopasowanych osób w grupie kontrolnej, których cechy obserwowalne są podobne do cech jednostek eksperymentalnych. W metodzie kontrfaktycznej ewaluacji jednostki eksperymentalne są dopasowywane do jednostek kontrolnych z podobnym wynikiem ewaluacji, tj. prawdopodobieństwem poddania oddziaływaniami z uwagi na zbiór obserwowalnych zmiennych. Metoda ta polega na założeniu, że w zależności od obserwowalnej charakterystyki wybór do grupy oddziaływania jest losowy (aby uzyskać więcej informacji, zob. np. Rosenbaum i Rubin, 1983; Angrist i Pischke, 2008).

RDD: model regresji nieciągłej (Regression Discontinuity Design). Quasi-eksperymentalna metoda, która umożliwia oszacowanie efektu oddziaływania programu, gdy kwalifikacja lub prawdopodobieństwo uczestniczenia w programie zależy od określonych cech obserwowalnych, na przykład wieku (zmienna przypisująca lub bieżąca). Efekt programu mierzy się poprzez porównanie rezultatów osób znajdujących się poniżej i powyżej progu kwalifikacji i w jego bliskim sąsiedztwie. RDD służy do obliczenia lokalnego średniego efektu oddziaływania poprzez dopasowanie (lokalnego) wielomianu po obu stronach progu i polega na założeniu, że wybór między grupą eksperymentalną a grupą kontrolną jest losowy. RDD może być rozmyty lub ostry (aby uzyskać więcej szczegółowych informacji, zob. np. Imbens i Lemieux, 2008; Angrist i Pischke, 2008).

SISCO: Sistema Statistico delle Comunicazioni Obbligatorie on-line, obowiązkowe zgłoszenia zbierane przez ministerstwo pracy i polityki społecznej (MPPS) we Włoszech

SUTVA: Stable Unit Treatment Value Assumption, założenie przyjmowane zazwyczaj we wnioskowaniu przyczynowym: oddziaływanie na jedną jednostkę nie ma wpływu na rezultat innej jednostki (brak interferencji między jednostkami).

Brak zmiennych zakłócających: odnosi się do przypadku, w którym kontrola różnic w określonym zbiorze zmiennych towarzyszących przed zastosowaniem oddziaływania eliminuje odchylenia porównań między jednostkami eksperymentalnymi a kontrolnymi. Dzięki temu w zależności od tej kontroli przypisanie oddziaływania można uznać za randomizowane. Znane również jako założenie pomijalności lub warunkowej niezależności.

© BUDAPEST INSTITUTE FOR POLICY ANALYSIS 2020

Adres: Dohány utca 84, Budapest, 1074 Hungary

info@budapestinstitute.eu

www.budapestinstitute.eu



Ta publikacja jest objęta [licencją CC BY 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

