

Ocena wpływu i prognozowanie efektów w badaniach ewaluacyjnych

**redakcja
Agnieszka Haber
Rafał Trzeciński**

**Ocena wpływu i prognozowanie efektów w badaniach ewaluacyjnych
pod redakcją Agnieszki Haber i Rafała Trzcíńskiego**

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

Tekst Daniela Fujiwary *Analiza kosztów i korzyści społecznych na przykładzie programów wspierania zatrudnienia* jest tłumaczeniem fragmentów Department for Work and Pensions Working Paper „Department for Work and Pensions Social Cost-Benefit Analysis framework: Methodologies for estimating and incorporating the wider social and economic impacts of work in Cost- Benefit Analysis of employment programmes” (2010)

Tekst Howarda White’a *Ocena oddziaływania oparta na teorii: zasady i praktyka* jest tłumaczeniem 3ie Working Paper „Theory-Based Impact Evaluation: Principles and Practice” (2009)

Redakcja tekstu D. Fujiwary: Maciej Szałaj

Redakcja techniczna: Zuzanna Popis

Grafika na okładce: Marcin May

© Copyright by Polska Agencja Rozwoju Przedsiębiorczości

ISBN: 978-83-7633-177-5

Wydanie pierwsze

Nakład: 1000 egz.

Przygotowanie do druku, druk i oprawa:

Agencja Reklamowo-Wydawnicza A. Grzegorzcyk

Spis treści

Wprowadzenie

<i>Bożena Lublińska-Kasprzak</i>	5
<i>Agnieszka Haber, Rafał Trzciński</i>	7

Stanisław Bienias

Metaewaluacja jako narzędzie służące projektowaniu przyszłych interwencji publicznych. Przykłady zastosowania w ramach Polityki Spójności	9
--	---

Alberto Martini

Wprowadzenie do podstawowych pytań stawianych w ewaluacji oddziaływania interwencji: Czy interwencja coś zmieniła? A jeśli tak, to dla kogo?	23
---	----

Daniel Fujiwara

Analiza kosztów i korzyści społecznych na przykładzie programów wspierania zatrudnienia	31
--	----

Howard White

Ocena oddziaływania oparta na teorii: zasady i praktyka	65
--	----

Paweł Penszko

Przykład ewaluacji efektów Aktywnych Polityk Rynku Pracy	81
---	----

Piotr Rosik, Elżbieta Opalka

Szacowanie efektów projektów drogowych. Przykłady zastosowania modelu ustalania i prognozowania efektów z wykorzystaniem wskaźnika dostępności transportowej	93
---	----

Szanowni Państwo,

Mamy przyjemność zaprezentować Państwu ósmy tom serii wydawniczej PARP poświęconej ewaluacji. Naszym zamierzeniem jest, aby tematyka prezentowanych Państwu tekstów odpowiadała na najbardziej aktualne wyzwania związane z realizacją polityk europejskich i krajowych oraz na oczekiwania wobec badań ewaluacyjnych, jakie z tych wyzwań wynikają.

Dobiega końca perspektywa finansowa 2007-2013. Jednocześnie trwają coraz bardziej zaawansowane prace nad kształtem Polityki Spójności w kolejnym okresie programowania. Kluczowe zadanie ewaluacji stanowi zatem trafne zdiagnozowanie osiągniętych efektów oraz rzetelna prognoza efektów oczekiwanych. Są to zagadnienia, wokół których skoncentrowana jest problematyka niniejszego tomu.

Wiodącymi motywami zamieszczonych w nim tekstów jest pomiar wpływu (*impact evaluation*) i ewaluacja oparta na teorii programu (*theory-based evaluation*) oraz metody prognostyczne, z uwzględnieniem metaewaluacji jako sposobu na wzmocnienie użyteczności dostępnych danych.

Mam nadzieję, że przekazywana Państwu publikacja wnieśli wkład w debatę środowisk zaangażowanych w prowadzenie ewaluacji oraz zarządzanie interwencjami publicznymi. Będziemy niezmiernie usatysfakcjonowani, jeśli odnajdą w niej Państwo idee, które będą mogli wykorzystać w zlecanych lub wykonywanych przez siebie badaniach. Serdecznie zapraszam do lektury.

Bożena Lublińska-Kasprzak
Prezes Polskiej Agencji Rozwoju Przedsiębiorczości

Wprowadzenie

Trafne diagnozowanie i prognoza efektów polityk publicznych stanowią kluczowe zadania, jakie stoją przed badaniami ewaluacyjnymi. Wiedzy tej nie jesteśmy bowiem w stanie uzyskać z innych źródeł, którymi posługujemy się w naszej codziennej praktyce. Potrzeba badań pełniących taką rolę jest szczególnie silna podczas podsumowania zakończonych i programowania nowych interwencji. Na takim właśnie etapie znajdujemy się obecnie.

Systematyczna ocena efektów wdrożonych programów jest niezbędnym warunkiem dla rzetelnego wyznaczenia celów i odpowiedniego doboru instrumentów. Zagadnieniom tym poświęcony został niniejszy zbiór artykułów obejmując trzy obszary: metody i podejścia ustalania efektów programów publicznych (tzw. *impact evaluation*), wyjaśnianie efektów programów publicznych oraz przewidywanie efektów planowanych interwencji publicznych.

Zagadnienie rygorystycznego, opartego na analizie stanów kontrfaktycznych, pomiaru efektów zostało podjęte w artykułach Alberto Martiniego oraz Pawła Penszko. Pomiar efektów stanowi punkt wyjścia do oceny interwencji i odpowiedzi na kluczowe pytania, takie jak „czy zrealizowana interwencja rzeczywiście miała wpływ na zmianę obserwowanej rzeczywistości społeczno-gospodarczej” oraz „czy obserwowane usprawnienia to wynik naszych działań”. Jednocześnie etap ten nie stanowi ewaluacji w pełnym tego słowa znaczeniu, bowiem nie daje odpowiedzi na pytania dotyczące przyczyn powodzenia lub niepowodzenia interwencji. Kolejnym wyzwaniem jest więc ustalenie dlaczego uzyskano takie, a nie inne efekty. Kwestia ta stanowi przedmiot drugiej części publikacji.

Howard White oraz Daniel Fujiwara wskazują, że użyteczność ewaluacji w dużej mierze zależy od tego, na ile pozwalają one wskazać mechanizmy odpowiadające za osiągnięcie zakładanych celów. Kluczowa jest odpowiedź na pytania, dlaczego udało się bądź nie udało osiągnąć określone efekty, co należałoby zrobić, aby osiągnąć lepsze efekty przy tych samych lub mniejszych nakładach oraz jakie grupy beneficjentów w sposób szczególny skorzystały na udziale w ocenianych programach publicznych. Uzyskanie wiedzy w tym zakresie pozwala lepiej projektować przyszłe interwencje publiczne bazując na faktycznych doświadczeniach, co de facto jest jednym z ważniejszych zadań stawianych przed ewaluacją (eliminacja działań w rzeczywistości nieskutecznych, bardziej precyzyjne projektowanie interwencji i ich adresowanie). W literaturze przedmiotu zagadnienie to rozwijane jest szerzej w ramach tzw. ewaluacji opartej na teorii programu (*theory based impact evaluation*). Problem ten równocześnie dotyka szerszego zagadnienia, jakim jest wykorzystanie ewaluacji na potrzeby tworzenia przyszłych programów wsparcia.

Trzecie kluczowe zagadnienie dotyczące przewidywania efektów planowanych interwencji publicznych, zostało poruszone w artykułach Stanisława Bieniasa oraz Piotra Rosika i Elżbiety Opałki. Wykorzystanie wniosków z dotychczasowych ewaluacji oraz przewidywanie efektów przyszłych programów publicznych stanowi wciąż wyzwanie w praktyce realizowanych ewaluacji. Szczególnie trudne jest przewidywanie, jak duże efekty (w rozumieniu przyczynowym) będą osiągnięte po realizacji interwencji w danym kształcie

(przy określonych zasobach i założonych działaniach). Jeszcze trudniej przewidywać efekty interwencji, które nie miały swoich bliskich odpowiedników w historii. Postęp w tym obszarze na gruncie nauk społecznych jest cały czas ograniczony. Rozwój podejść badawczych oraz przykłady dobrych praktyk mogą z tego powodu stanowić ważny wkład w rozwój dyscypliny i tym samym realizację efektywnych polityk publicznych.

Mamy nadzieję, że zaproponowany wybór tekstów spotka się z Państwa zainteresowaniem i stanowić będzie inspirację dla Państwa praktyki zawodowej.

Agnieszka Haber
Rafał Trzciński

Metaewaluacja jako narzędzie służące projektowaniu przyszłych interwencji publicznych. Przykłady zastosowania w ramach Polityki Spójności

1. Wprowadzenie

Terminu metaewaluacja używamy najczęściej w kontekście ewaluacji samego procesu ewaluacji. Tak krótko zresztą definiuje ją M. Scriven (1991) jako ewaluację ewaluacji. Metaewaluacja ogniskuje się na ogół na ocenie mocnych i słabych stron badania, dotyczy samego raportu bądź też całego procesu, jednej lub więcej ewaluacji.

Rola metaewaluacji w literaturze światowej systematycznie rośnie. Istnieje wiele typologii tego rodzaju badań. Przykładowo, Cook i Gruder (1978) opracowali typologię bazującą na trzech wymiarach i siedmiu typach wynikających z ich nałożenia się. W ramach pierwszego wymiaru rozpatruje się, czy metaewaluacja jest realizowana post factum, czy w trakcie przygotowywania właściwego badania. Wymiar drugi dotyczy zakresu danych poddawanych metaewaluacji (czy są to dane pierwotne zebrane w samej metaewaluacji, czy też powtórnej analizie poddawane są dane z wcześniejszego badania). Ostatni aspekt dotyczy tego, czy zakresem metaewaluacji objęte jest jedno, czy więcej badań dotyczących zdefiniowanego tematu. Jeszcze inne typologie przedstawia np. Bustelo (2002), czy Pierre (1982), który ilustruje przykładami teorie Cooka i Grudera.

O specyficznym typie metaewaluacji możemy mówić też wówczas, gdy poddajemy analizie same wyniki badań ewaluacyjnych. Jak pokazuje praktyka, badania takie często polegają (poza oceną rzetelności otrzymanych wniosków) na podsumowaniu wyników prac badawczych i wyciągnięciu strategicznych rekomendacji. Metaewaluacje mogą poddawać analizie wyniki całej grupy wcześniejszych ewaluacji dotyczących określonego sektora, tematu, czy problemu. Jak wskazuje Olejniczak (pozycja w druku) metaewaluacje, mające na celu syntezę wcześniejszych wniosków i rekomendacji, stają się coraz częstszą praktyką badawczą.

W Polsce, w praktyce funduszy strukturalnych, często badania, które bazują w swojej analizie na wcześniejszych ewaluacjach, określane są mianem metaanalizy. Termin ten nie jest jednak do końca poprawnie stosowany. Metaanaliza nie jest bowiem prostym podsumowaniem wcześniejszych wyników prac badawczych. W ramach metaanalizy poddajemy usystematyzowanej analizie statystycznej wyniki wcześniej prowadzonych badań ilościowych. W przypadku podsumowań badań ewaluacyjnych, gdzie wykorzystywane są nie tylko badania ilościowe, ale i jakościowe i nie są one później poddawane rygorystycznej analizie statystycznej, używanie terminu metaanalizy wydaje się być nadużyciem (zob. Skórska, 2010).

Równocześnie należy mieć na względzie, że podsumowanie wyników różnych, punktowych badań ewaluacyjnych staje się pilną potrzebą w ramach realizacji Polityki Spójności. Polityka ta jest jedną z najczęściej ewaluowanych polityk w Europie (w samej Polsce zrealizowano już przeszło pół tysiąca badań). Mogą one stanowić dobrą bazę faktograficzną, dającą możliwość wyciągania wniosków dla prowadzenia

Polityki Spójności i urzeczywistnienia zasady *evidence-based policy*. Problemem jest jednak zebranie, a następnie podsumowanie tak szerokiego materiału dowodowego, który w swej pierwotnej postaci jest mało użyteczny dla decydentów.

Konieczne jest więc dokonywanie przeglądu zrealizowanych ewaluacji, identyfikowanie najistotniejszych badań (w kontekście istniejących potrzeb decyzyjnych), analizowanie ich rzetelności, a następnie syntetyzowanie wyników i wyciąganie (na podstawie badań cząstkowych) strategicznych wniosków. W niniejszym artykule termin metaewaluacja będzie używany w tym właśnie znaczeniu.

Metaewaluacja, mająca na celu podsumowanie istniejących badań ewaluacyjnych, zaczyna być powoli narzędziem wykorzystywanym również w praktyce ewaluacji Polityki Spójności. Na przestrzeni ostatnich dwóch lat można zidentyfikować co najmniej kilka takich przedsięwzięć. W ramach niniejszego artykułu skoncentrowano się na bardziej szczegółowym przedstawieniu dwóch z nich. Pierwsze to *Ewaluacja ex post perspektywy finansowej 2000-2006* przygotowana przez Dyрекcję Generalną ds. Polityki Regionalnej Komisji Europejskiej (DG Regio). Było to największe pod względem skali i nakładów finansowych przedsięwzięcie ewaluacyjne w historii Polityki Spójności obejmujące swym zasięgiem wszystkie Państwa Członkowskie. Drugim omówionym badaniem jest *Ewaluacja ex post Narodowego Planu Rozwoju na lata 2004-2006*. Ewaluacja zrealizowana została na poziomie krajowym stanowiąc kamień milowy w rozwoju ewaluacji Polityki Spójności w Polsce. Badania te zostały wybrane również ze względu na skalę oddziaływania i rolę jaką odgrywają w projektowaniu przyszłej Polityki Spójności. W artykule przedstawiono powyższe badania wskazując: powody uruchomienia ewaluacji, sposób jej realizacji, wykorzystanie wyników, mocne i słabe strony badania oraz wnioski dla przyszłych prac badawczych.

W końcowej części artykułu przedstawione zostały najważniejsze wnioski dotyczące perspektywy wykorzystania metaewaluacji w przyszłości m.in. kolejnych interwencji publicznych.

2. Metaewaluacja na poziomie europejskim – ewaluacja ex post perspektywy 2000-2006 realizowana przez Dyрекcję Generalną ds. Polityki Regionalnej Komisji Europejskiej (KE, 2010a)

Głównym celem badania było podsumowanie efektów realizacji Polityki Spójności (w zakresie Europejskiego Funduszu Rozwoju Regionalnego – EFRR) w okresie 2000-2006, w tym:

- zbadanie, w jaki sposób wykorzystane zostało wsparcie udzielone w ramach EFRR w regionach państw EU15 i EU10 (stare i nowe państwa członkowskie);
- ocena efektów osiągniętych dzięki wdrażaniu EFRR w tych regionach oraz stopnia, w jakim efekty te pozwoliły na przybliżenie się do osiągnięcia celów Polityki Spójności;
- przedstawienie wniosków na przyszłość umożliwiających zarówno właściwe zaprojektowanie, jak również późniejszą implementację Polityki Spójności, zwłaszcza po roku 2014.

Realizacja badania zajęła ponad trzy lata (marzec 2007 – połowa 2010), a całkowity budżet sięgnął blisko 5 mln Euro (Gaffey, 2010).

2.1. Uzasadnienie dla realizacji badania

Obowiązek przeprowadzenia ewaluacji ex post ciąży, zgodnie z zapisami Rozporządzenia Ogólnego (1260), na Komisji Europejskiej. Wg art. 43 Komisja powinna w tym zakresie współpracować z Państwami Członkowskimi. Należy jednak wziąć pod uwagę, że poza oczywistym wymogiem prawnym, istotny był kontekst przeprowadzenia badania.

Realizacja ewaluacji ex post zbiegła się bowiem z rozpoczęciem debaty na temat przyszłości Polityki Spójności po roku 2013. Komisja Europejska znalazła się pod olbrzymią presją, z jednej strony instytucji europejskich (Parlamentu, Trybunału Obrachunkowego...), z drugiej płatników netto (państw członkowskich). Presja ta wywołana została licznymi głosami krytyki, zarzucającymi nieefektywne i nieskuteczne wykorzystanie pomocy strukturalnej przez beneficjentów Polityki Spójności, a w skali europejskiej ograniczoną widocznością jej efektów. Ewaluacja ex post miała więc za zadanie dostarczyć kontrargumentów, dowodów na skuteczność prowadzonych działań, jak również wskazać obszary potencjalnych usprawnień. Realizacja ewaluacji ex post zbiegła się również w czasie z głębokim kryzysem gospodarczym, którego skutki objawiają się w szczególności szukaniem oszczędności budżetowych.

2.2. Sposób realizacji

W przeciwieństwie do wcześniej prowadzonych ewaluacji ex post (okresu 1989-1993 i 1994-1999) Komisja Europejska zdecydowała się istotnie zmienić podejście badawcze. W ramach perspektywy 1994-1999 starano się bowiem uchwycić w sposób systematyczny (dwa raporty dla każdego celu oraz raporty cząstkowe dla poszczególnych Państw Członkowskich) pełne efekty Polityki Spójności. Niestety podejście to uniemożliwiało pogłębioną analizę poszczególnych obszarów wsparcia (brak możliwości dostosowania metodologii do specyficznych typów interwencji), a sama agregacja efektów – z poszczególnych państw na poziom Komisji Europejskiej – okazała się mocno utrudniona. Utrudnienia te wynikały między innymi ze słabości istniejących w Państwach Członkowskich systemów monitoringu rzeczowego, w tym w szczególności z braku jednolitego systemu wskaźników kluczowych. Ponadto, stosunkowo ograniczona liczba wcześniej przeprowadzonych badań dotyczących efektów realizacji Polityki Spójności nie pozwalała na dokonanie takich podsumowań w ramach pojedynczej ewaluacji zleconej z poziomu Komisji Europejskiej.

Realizując ewaluację ex post dla okresu 2000-2006 podjęto decyzję o odejściu od podejścia „systemowego” na rzecz skoncentrowania się na kluczowych obszarach tematycznych wsparcia. Taka organizacja pozwoliła na dostosowanie metodologii do specyficznych obszarów interwencji, jak również na uzyskanie pogłębionej wiedzy. Podejście to jednocześnie przekreślało możliwość uzyskania pełnego i wyczerpującego obrazu efektów realizacji Polityki Spójności (co, jak wskazywały wcześniejsze doświadczenia i tak miało małe szanse powodzenia).

Prace podzielono na jedenaście „pakietów tematycznych”, na bazie których została przeprowadzona metaewaluacja. Pakiety te zostały krótko scharakteryzowane w poniższej tabeli.

PAKIET TEMATYCZNY	CEL I ZAKRES PRAC	METODOLOGIA
Koordinacja i synteza	Pierwszy etap – analiza ogólnej sytuacji społeczno-gospodarczej oraz zgromadzenie informacji na temat wykorzystania środków finansowych. Drugi etap – pomoc w koordynacji prac nad pozostałymi pakietami tematycznymi. Trzeci etap – przeprowadzenie metaewaluacji i syntezy wszystkich pakietów tematycznych oraz opracowanie raportu końcowego.	Analiza danych zastanych, przegląd literatury, metaewaluacja.
Zebranie danych	Opracowanie jednej spójnej bazy danych gromadzącej informacje ze wszystkich raportów rocznych z programów operacyjnych.	Analiza danych zastanych, IDI.
Modelowanie makroekonomiczne	Ocena wpływu realizacji Polityki Spójności na gospodarkę państw członkowskich.	Wykorzystano dwa modele makroekonomiczne: HERMIN i QUEST (model klasy DSGE).

Zmiany strukturalne i globalizacja	Ocena wpływu realizacji Polityki Spójności na zmiany strukturalne w wybranych regionach Celu 2.	Przegląd literatury, analiza danych zastanych, studia przypadków (na poziomie 12 regionów, w tym: wywiady indywidualne, ankiety, mini studia przypadków), warsztat podsumowujący.
Transport	Określenie wpływu wdrażania EFRR na rozwój europejskiego systemu transportu.	Przegląd literatury, analiza danych zastanych, studia przypadków, IDI.
Środowisko i zmiany klimatyczne	Określenie wpływu realizacji interwencji Polityki Spójności na wdrażanie dyrektyw środowiskowych KE. Ocena wpływu zrealizowanych projektów na rozwój społeczno-gospodarczy oraz oszacowanie stopnia, w jakim prowadzona interwencja przyczyniła się do przeciwdziałania zmianom klimatycznym.	Przegląd literatury, analiza danych zastanych, IDI, studia przypadków.
Przedsiębiorczość i innowacyjność	Podsumowanie efektów 30 programów operacyjnych dedykowanych wsparciu przedsiębiorstw oraz określenie wpływu wsparcia bezpośredniego przedsiębiorstw.	Przegląd dokumentacji dot. 30 programów operacyjnych dedykowanych wsparciu przedsiębiorstw, analiza danych zastanych, IDI. Dla określenia wpływu wsparcia bezpośredniego przedsiębiorstw wykorzystano: <i>propensity score matching</i> , <i>difference in difference</i> oraz <i>Heckman selection model</i> .
Równe traktowanie i zmiany demograficzne	Określenie wpływu realizacji interwencji współfinansowanych w ramach EFRR na kwestie równouprawnienia i dostosowania się do zachodzących zmian demograficznych.	Przegląd literatury, analiza danych zastanych, studia przypadków (na poziomie regionów), IDI, mini studia przypadków.
Transport modelowanie	Ocena efektów ekonomicznych projektów transportowych współfinansowanych z EFRR.	Nie udało się w pełni wykorzystać modelu TRANSTOOLS.
Rozwój obszarów wiejskich	Określenie wpływu EFRR na rozwój obszarów wiejskich.	Przegląd literatury, analiza danych zastanych, IDI, studia przypadków (poziom regionalny), mini studia przypadków.
Efektywność dużych projektów	Analiza ponoszonych kosztów oraz harmonogramów realizacji dużych inwestycji infrastrukturalnych (tzw. <i>major projects</i>). Celem szczegółowym badania było stworzenie (udało się to zrobić w ograniczonym zakresie) katalogu kosztów, który mógłby być wykorzystany w planowaniu przyszłych projektów.	Przegląd literatury, analiza dokumentacji dużych projektów, analizy statystyczne.
System zarządzania i wdrażania	Kompleksowa ocena systemu realizacji Polityki Spójności w państwach członkowskich UE, a na tej podstawie wypracowanie horyzontalnych wniosków i rekomendacji.	Przegląd literatury, analiza danych zastanych, mapping, studia przypadków, analiza porównawcza, mini studia przypadków, panel ekspertów.

Źródło: Opracowanie własne na podstawie raportów dostępnych na stronie: http://ec.europa.eu/regional_policy/information/evaluations/archives_2000_2006_en.cfm#1

2.3. Wykorzystanie wyników

Warunkiem koniecznym dla wykorzystania wyników ewaluacji ex post było właściwe rozpowszechnienie jej wyników. Strategia wykorzystania wyników była wbudowana w proces ewaluacji od samego początku, między innymi dzięki włączeniu w proces realizacji każdego z pakietów tematycznych grupy sterującej złożonej z przedstawicieli różnych Dyrekcji Generalnych, potencjalnie zainteresowanych wynikami badania. Takie podejście partycypacyjne umożliwiło wytworzenie „zapotrzebowania na wyniki”, jak również dostosowania wyników do potrzeb interesariuszy procesu. Intencjonalnym rozwiązaniem było również zaangażowanie wybitnych ekspertów, akademików, jako niezależnych recenzentów każdego z pakietów

tematycznych. Osoby te często wcześniej postrzegane były jako istotni krytycy Polityki Spójności. Poza podniesieniem jakości samych badań, umożliwiło to włączenie tej bardzo ważnej grupy, mającej duży wpływ na opinie środowiska naukowego i eksperckiego, w proces udoskonalania Polityki Spójności.

Ponieważ każdy z pakietów tematycznych był dyskutowany osobno, przeprowadzono około 45 spotkań grup sterujących i około 35 spotkań eksperckich (Gaffey, 2010). Dzięki temu sam raport końcowy – wynik metaewaluacji – oraz wnioski i rekomendacje z niego płynące były postrzegane jako bardzo wiarygodne. Bazowały bowiem na solidnych i dogłębnie przedyskutowanych wynikach wcześniejszych prac badawczych.

Wyniki tych prac, przedstawione w metaewaluacji, znalazły swoje odzwierciedlenie w przygotowanym V Raporcie Kohezyjnym. Wnioski z tego Raportu dotyczące:

- potrzeby silniejszego ukierunkowania na osiągnięte rezultaty,
- systematycznej dyskusji na temat postępu w zakresie osiągnięcia zakładanych efektów,
- koncentracji środków na ograniczonej ilości obszarów wsparcia,
- szerszego stosowania nowych i eksperymentalnych interwencji,
- lepszego systemu wskaźników oraz jego wykorzystywania w procesie zarządzania Polityką Spójności,
- rozwiązań w zakresie ewaluacji (planowanie procesu, szersze wykorzystanie zaawansowanych metod badawczych, obowiązek przeprowadzanie ewaluacji podsumowujących przez państwa członkowskie),
- implementacji wymiaru terytorialnego Polityki Spójności,
- wypracowania odpowiedniego podejścia do wsparcia obszarów miejskich,
- umożliwienia stosowania podejścia lokalnego,
- utrzymania szerokiego zakresu wsparcia obejmującego wszystkie państwa członkowskie (a nie ograniczania wsparcia wyłącznie dla najbiedniejszych),

miały swoje silne uzasadnienie w zebranych materiale dowodowym (Gaffey, 2010). Większość z nich znalazła też odzwierciedlenie w ostatnio opublikowanych projektach rozporządzeń na przyszły okres programowania.

2.4. Mocne i słabe strony badania oraz wnioski dla przyszłych prac badawczych

Metaewaluacja, jako narzędzie służące z jednej strony podsumowaniu efektów, z drugiej wypracowaniu wniosków na przyszłość, okazała się skutecznym narzędziem służącym przygotowaniu strategicznych wniosków i rekomendacji w ramach *Ewaluacji ex post EFRR w okresie 2000-2006*. Kluczowe czynniki, jakie się do tego przyczyniły to:

- Zaplanowanie całego procesu jako jednej i spójnej całości. Wnioski z poszczególnych pakietów tematycznych wzajemnie się uzupełniały, a wyniki wcześniejszych prac były wykorzystywane przez kolejne pakiety tematyczne.
- Zapewnienia wysokiej jakości poszczególnych badań tematycznych (szczegółowe i skoncentrowane pytania badawcze, możliwość wykorzystania dostosowanych do obszaru tematycznego metod badawczych (w większym stopniu niż miało to miejsce w poprzednich badaniach ex post), punktowe zastosowanie badań z wykorzystaniem grup porównawczych, stała współpraca ze zlecającymi, wnikliwa analiza dostarczanych produktów, zaangażowanie ekspertów zapewniających dodatkowy *quality check*).
- Wypracowanie „zaufania” kluczowych odbiorców badania poprzez włączenie ich w cały proces na wszystkich jego etapach. Pozwoliło to, już na etapie wykorzystania wyników metaewaluacji, skupić się na dyskusji nad wnioskami i rekomendacjami, a nie nad dyskusją na temat rzetelności dowodów stojących u podstawy wnioskowania.

Pewną słabością raportu podsumowującego było to, że ewaluator zewnętrzny nie w pełni wykorzystał zgromadzony materiał dowodowy wypracowany w ramach poszczególnych pakietów tematycznych. Inny problem pojawił się na etapie prac nad pakietami tematycznymi, mianowicie bardzo słaby system monitorowania (zarówno na szczeblu państw członkowskich, jak i samej KE) i brak możliwości bazowania na porównywalnych danych. Powodowało to konieczność zbierania danych monitoringowych w samym procesie ewaluacji (utracony czas i poniesione koszty, które mogły zostać wykorzystane w fazie analizy i oceny). Ostatnia uwaga dotyczy ograniczonego wykorzystania badań ewaluacyjnych zrealizowanych na poziomie Państw Członkowskich. Wykonawcy poszczególnych pakietów tematycznych słabo identyfikowali i podchodzili z pewną nieufnością do badań realizowanych przez krajowe instytucje. Wynikało to również z braku gromadzenia (w tamtym czasie) badań krajowych na poziomie Komisji, w ramach jednej spójnej bazy.

Reasumując, oparcie finalnej metaewaluacji na szeregu badań częściowych zorientowanych tematycznie okazało się być słusznym wyborem.

Co jednak powinniśmy zmienić by w przyszłości ewaluacja ex post okresu 2007-2013 była jeszcze lepszej jakości (poza eliminowanymi obecnie mankamentami)? Po pierwsze, jakość finalnego badania metaewaluacyjnego zależy w największym stopniu od jakości badań ewaluacyjnych, na których jest oparta. Konieczne jest więc podniesienie jakości prowadzonych badań, w szczególności szersze niż dotychczas wykorzystanie metod kontrfaktycznych, które stanowią najlepszy materiał dla późniejszych metaewaluacji. Badania wykorzystujące grupy porównawcze powinny jednak być uzupełniane o rzetelne badania oparte na teorii. Badania te powinny dogłębnie analizować przyczyny zaistnienia efektów (dlaczego dana interwencja działa?) oraz analizować kontekst wdrażania i wskazywać możliwości wykorzystania dobrych praktyk w ramach przyszłych interwencji.

3. Metaewaluacja na poziomie krajowym – ewaluacja ex post perspektywy 2004-2006 (MRR, 2010b) realizowana przez Krajową Jednostkę Oceny

Głównym celem realizacji ewaluacji ex post Narodowego Planu Rozwoju na lata 2004-2006 była odpowiedź na pytanie „Czy interwencje zrealizowane w ramach NPR 2004–2006 rzeczywiście przyczyniły się do podniesienia konkurencyjności gospodarki, zdolnej do długofalowego i harmonijnego rozwoju, zarówno na poziomie krajowym, jak i regionalnym?” Badanie zostało przeprowadzone w latach 2009-2010 obejmując swym zakresem wszystkie fundusze Polityki Spójności.

3.1. Uzasadnienie dla realizacji badania

Mimo że Komisja Europejska nie nakłada obowiązku realizacji ewaluacji ex post na Państwa Członkowskie, w Polsce podjęto decyzję o realizacji kompleksowego badania efektów pierwszej perspektywy finansowej. Pierwszym istotnym argumentem za uruchomieniem procesu ewaluacji ex post był wymóg przeprowadzania takiego badania przez ustawę o Narodowym Planie Rozwoju. Drugi argument miał natomiast naturę poznawczą. W roku 2004 Polska stała się jednym z największych beneficjentów wsparcia w ramach europejskiej Polityki Spójności. NPR 2004-2006 był (zarówno co do skali środków finansowych, jakości opracowania samego dokumentu strategicznego, jak i sposobu realizacji) bezprecedensowy w historii Polski i stanowił największą ustrukturyzowaną interwencję publiczną wdrażaną konsekwentnie przez kolejne koalicje parlamentarno-rządowe. Stanowił jednocześnie swoiste preludium dla możliwości wykorzystania wielokrotnie większych środków w ramach Narodowej Strategii Spójności na lata 2007-2013 (MRR, 2010b). Z tego względu analiza efektów, wypracowanie wniosków i rekomendacji dla realizacji Polityki Spójności w przyszłości, jak również bardziej ogólnych wniosków dla prowadzenia polityki rozwoju w Polsce, miało swój głęboko uzasadniony cel.

3.2. Sposób realizacji

Realizacja ewaluacji ex post Narodowego Planu Rozwoju 2004–2006 przebiegała etapowo. Pierwszym etapem było zlecenie tematycznych badań cząstkowych. Przyjęto więc podobny scenariusz, jak w przypadku ewaluacji dokonanej przez Komisję Europejską. Ewaluacje tematyczne, poza komponentem służącym badaniu nowych zagadnień i wypełnieniu istniejących luk informacyjnych, posiadały rozbudowany komponent metaewaluacyjny. Celem metaewaluacji w ramach tematycznych badań cząstkowych było podsumowanie wiedzy i wyciągnięcie wniosków z przeprowadzonych do tej pory blisko 400 badań zrealizowanych przez instytucje usytuowane na różnych poziomach wdrażania Polityki Spójności. Badania tematyczne zlecono w trzech wymiarach: horyzontalnym, sektorowym i terytorialnym. W tabeli poniżej przedstawiony został zakres badań cząstkowych:

WYMIAR	NAWA BADANIA CZĄSTKOWEGO	CEL/ZAKRES BADANIA	METODOLOGIA BADANIA
HORYZONTALNY	Ocena korzyści uzyskiwanych przez Państwa UE-15 w wyniku realizacji Polityki Spójności w Polsce.	Przedstawienie korzyści jakie uzyskują kraje UE-15 w wyniku realizacji Polityki Spójności w Polsce – głównie w postaci dodatkowego eksportu do naszego kraju.	Przegląd literatury, analiza danych zastanych, CAWI, IDI, Model DSGE, analiza dokumentacji projektowej, studia przypadków.
HORYZONTALNY	Wpływ Polityki Spójności na poziom i jakość zatrudnienia w Polsce.	Określenie wpływu realizacji NPR 2004-2006 na poziom i jakość zatrudnienia w Polsce z uwzględnieniem wymagań gospodarki opartej na wiedzy.	Przegląd literatury, metaewaluacja wcześniejszych wyników badań ewaluacyjnych, metaewaluacja wyników modelowania makroekonomicznego, analiza danych zastanych, IDI, CAPI, panel ekspercki, CATI, studia przypadków, <i>propensity score matching</i> (w zakresie skuteczności szkoleń dla bezrobotnych).
HORYZONTALNY	Komplementarność i synergia interwencji realizowanych w ramach EFS i EFRR w perspektywie finansowej 2004-2006.	Ocena realizacji zasady komplementarności pomiędzy EFS i EFRR, wewnątrz funduszy oraz – na poziomie projektów, czy zapewnienie komplementarności prowadziło do zaistnienia pozytywnych efektów synergii.	Przegląd literatury, analiza danych zastanych, IDI, FGI, CAPI, analiza dokumentacji projektowej, CATI, studia przypadków, analiza porównawcza (<i>matching</i>).
HORYZONTALNY	Ocena systemu realizacji Polityki Spójności w Polsce w ramach perspektywy 2004-2006.	Kompleksowa ocena systemu realizacji Polityki Spójności w Polsce w ramach perspektywy 2004-2006 z wyodrębnieniem funkcji wdrożeniowej, strategicznej i refleksyjnej.	Przegląd literatury, analiza danych zastanych, IDI, CAWI, panel ekspertów, analiza systemowa.
HORYZONTALNY	Wpływ makroekonomicznej realizacji Narodowego Planu Rozwoju.	Ocena oddziaływania makroekonomicznego zarówno na poziomie krajowym, jak i regionalnym.	Oceny zostały opracowane za pomocą dwóch różnych modeli makroekonomicznych (CGE i DSGE). Modele te zostały opracowane przez różne, niezależne grupy ekspertów.
SEKTOROWY	Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach Polityki Spójności na wzrost konkurencyjności regionów.	Ocena wpływu inwestycji infrastruktury transportowej na wzrost konkurencyjności regionów, w tym ocena osiągnięcia celów strategicznych, spójności i komplementarności powstałej infrastruktury, uzyskanych efektów społeczno-ekonomicznych.	Przegląd literatury, analiza danych zastanych, IDI, CATI, wywiad kwestionariuszowe, panele ekspertów, studia przypadków, model międzygałęziowej dostępności transportowej, analizy przestrzenne.

SEKTOROWY	Ocena efektów inwestycji środowiskowych finansowanych w ramach NPR 2004-2006.	Ocena efektów inwestycji środowiskowych, w tym: stopnia osiągnięcia celów krajowych i zobowiązań akcesyjnych, oszacowania efektów ekonomicznych, jak również ich trwałości i użyteczności. Ponadto podjęto próbę oceny wpływu programu na przeciwdziałanie zmianom klimatycznym.	Przegląd literatury, analiza danych zastanych, IDI, CAWI, studia przypadków, analizy przestrzenne, analizy kosztów i korzyści, macierze powiązań.
SEKTOROWY	Ocena wpływu Polityki Spójności na wzrost konkurencyjności i innowacyjności polskich przedsiębiorstw i gospodarki.	Ocena wpływu Polityki Spójności na wzrost konkurencyjności i innowacyjności polskich przedsiębiorstw i gospodarki.	Przegląd literatury, metaewaluacja wcześniejszych wyników badań ewaluacyjnych, analiza danych zastanych, IDI, CAWI, studia przypadków, wielowymiarowa analiza statystyczna, <i>propensity score matching</i> (w ramach komponentu dot. oceny bezpośredniego wsparcia przedsiębiorstw), model makroekonomiczny DSGE.
SEKTOROWY	Ocena wpływu realizacji interwencji współfinansowanych z EFS na podniesienie jakości zasobów ludzkich.	Ocena wpływu realizacji interwencji współfinansowanych z EFS na podniesienie jakości zasobów ludzkich w Polsce.	Przegląd literatury, metaewaluacja wcześniejszych wyników badań ewaluacyjnych, analiza danych zastanych, IDI, CATI, panel ekspertów, analiza historii zdarzeń, <i>propensity score matching</i> (dla oceny trwałość powstałych wsparciu EFS przedsiębiorstw).
TERYTORIALNY	Ocena ex post wpływu realizacji Polityki Spójności perspektywy 2004-2006 na zwiększenie możliwości rozwoju społeczno-gospodarczego regionów Polski Wschodniej.	Ocena realizacji Polityki Spójności w perspektywie 2004-2006 w regionach Polski Wschodniej pod kątem zwiększenia ich możliwości i dynamiki rozwoju społeczno-gospodarczego.	Analiza danych zastanych, IDI, CATI, PAPI, FGI, panel ekspertów, benchmarking, studia przypadków.
TERYTORIALNY	Ocena wpływu Polityki Spójności na rozwój polskich miast.	Ocena wpływu Polityki Spójności na rozwój polskich miast i ich relacje z otoczeniem regionalnym. W ramach badania dokonano oceny wpływu Polityki Spójności na rozwój dużych miast pod kątem zwiększania ich konkurencyjności przy zachowaniu spójności przestrzennej i społecznej oraz wpływu Polityki Spójności na rozprzestrzenianie się procesów rozwojowych w otoczeniu regionalnym dużych miast.	Przegląd literatury, analiza danych zastanych, IDI, CAWI, FGI, studia przypadków.
TERYTORIALNY	Efekty transgranicznej współpracy polskich regionów w okresie 2004-2006.	Pogłębienie wiedzy nt. skuteczności i efektywności programów współpracy terytorialnej w aspekcie transgranicznym, w szczególności ocena, w jaki sposób wpłynęły one na spójność społeczną, gospodarczą i terytorialną, współpracę przygraniczną i integrację społeczności po obu stronach granicy.	Analiza danych zastanych, IDI, CATI, CAWI, FGI, studia przypadków, analiza wycinków prasowych, analiza statystyczna, analiza wielokryteriowa.

Źródło: Opracowanie własne na podstawie raportów dostępnych na stronie: www.ewaluacja.gov.pl/ewaluacja_ex_post_npr/

Kolejnym etapem procesu było przeprowadzenie metaewaluacji wyników badań cząstkowych. Raport został przygotowany przez pracowników Ministerstwa Rozwoju Regionalnego: Krajowej Jednostki Oceny oraz Jednostki Ewaluacyjnej w ramach Instytucji Zarządzającej EFS.

Głównym celem metaewaluacji była odpowiedź na pytanie „Czy interwencje zrealizowane w ramach NPR 2004–2006 rzeczywiście przyczyniły się do podniesienia konkurencyjności gospodarki, zdolnej do długofalowego i harmonijnego rozwoju, zarówno na poziomie krajowym, jak i regionalnym?”.

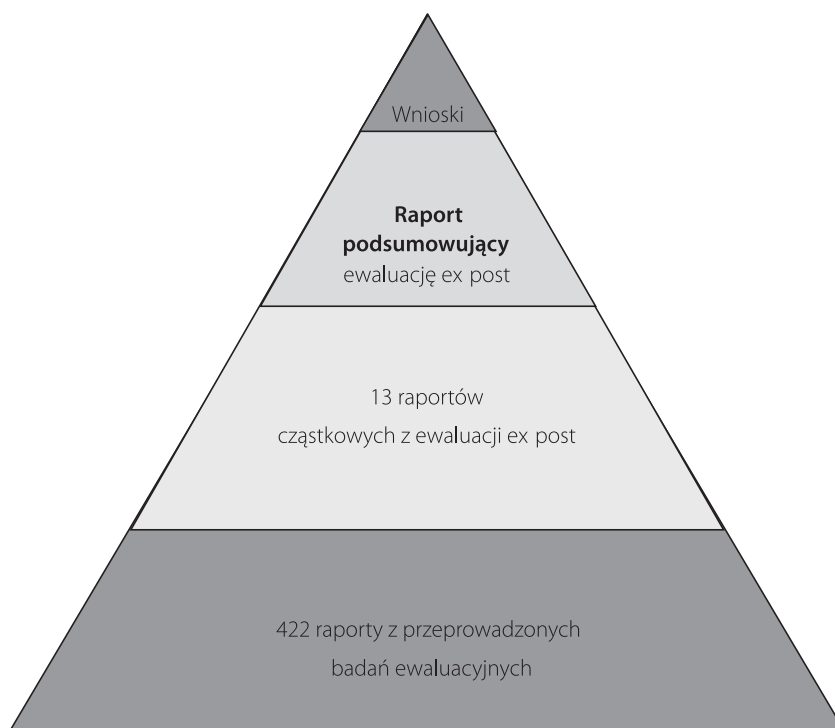
Celem dodatkowym było sformułowanie (w oparciu o wyniki szczegółowe badania) wniosków i rekomendacji dla:

- realizacji interwencji w ramach Narodowych Strategicznych Ram Odniesienia oraz Programów Operacyjnych w latach 2007-2013,
- optymalnego ukierunkowania środków krajowej rezerwy wykonania,
- kształtu i systemu realizacji Polityki Spójności po roku 2013,
- krajowych prorozwojowych polityk publicznych prowadzonych w obszarach wsparcia Narodowego Planu Rozwoju oraz Narodowych Strategicznych Ram Odniesienia.

Ostatnim etapem badania było przygotowanie (przy udziale Instytucji Zarządzających) krótkiego dokumentu: *Wnioski z realizacji Polityki Spójności w świetle wyników prac analityczno-ewaluacyjnych okresu 2004–2006*. Celem niniejszego dokumentu było przedstawienie zaleceń dla prowadzenia polityki rozwoju w Polsce w oparciu o analizę i ewaluację efektów Polityki Spójności okresu 2004-2006.

Całościowy model realizacji ewaluacji ex post Narodowego Planu Rozwoju na lata 2004-2006 przedstawiony został na poniższym rysunku.

Ryc. 1. Model realizacji ewaluacji ex post Narodowego Planu Rozwoju na lata 2004-2006



Źródło: P. Żuber, *Ewaluacja ex post Narodowego Planu Rozwoju 2004-2006: efekty, wnioski, rekomendacje*, VI Konferencja Ewaluacyjna 13 grudnia 2010, www.konferencja-ewaluacja.pl/2010/

3.3. Wykorzystanie wyników

Strategia wykorzystania wyników ewaluacji rozpoczęła się praktycznie tuż po podjęciu decyzji o jej realizacji. Pierwszym krokiem było bowiem włączenie maksymalnie szerokiego grona interesariuszy w przygotowanie koncepcji badania. Ogólną koncepcję skonsultowano i przedyskutowano ze wszystkimi Instytucjami Zarządzającymi, powołując jednocześnie grupy sterujące dla realizacji poszczególnych ewaluacji cząstkowych. Tym samym wzmocniono „oczekiwania na wyniki” i „współwłasność” całego procesu. Istotnie podniosło to też jakość opisów przedmiotu zamówienia, jak i późniejszych produktów końcowych.

W ramach każdego badania były zorganizowane co najmniej trzy spotkania. Pierwsze służyło dyskusji nad założeniami metodologicznymi. Drugie służyło zgłoszeniu uwag do projektu raportu. Sam raport był prezentowany na większym spotkaniu z udziałem szerszej grupy potencjalnie zainteresowanych odbiorców, w tym decydentów. Raporty końcowe zostały wydrukowane i rozdawane na licznych spotkaniach i konferencjach. Dodatkowo przygotowano i rozpowszechniono krótkie, kilkustronicowe broszurki informacyjne prezentujące najważniejsze wyniki i wnioski z ewaluacji. Ponadto utworzona została specjalna strona internetowa, na której publikowano na bieżąco wszystkie produkty procesu.¹

Raport podsumowujący był przedmiotem pogłębionej dyskusji panelowej w trakcie VI Konferencji Ewaluacyjnej², a następnie uzupełniony o kluczowe wnioski i rekomendacje.

Ostatnim i prawdopodobnie najważniejszym etapem całego procesu było przygotowanie krótkiego, kilkunastostronicowego dokumentu, którego odbiorcami byli decydenci – *Wnioski z realizacji Polityki Spójności w świetle wyników prac analityczno-ewaluacyjnych*. Dokument ten poddany był pogłębionej dyskusji z udziałem wszystkich dyrektorów instytucji zarządzających. Dzięki temu udało się na tym ostatnim etapie zaznajomić kluczowe (z punktu widzenia realizacji zasady *evidence based policy*) osoby z wynikami prac badawczych. Jednocześnie, dzięki dyskusji udało się wspólnie wypracować wysokiej jakości rekomendacje dla prowadzenia Polityki Spójności w przyszłości, z którymi identyfikuje się większość decydentów. Dokument został następnie przedstawiony na forum Komitetu Koordynacyjnego NSRO (najważniejsze forum decyzyjne w zakresie realizacji Polityki Spójności) oraz przyjęty przez Komitet Koordynacyjny ds. Polityki Rozwoju (najważniejsze forum decyzyjne w zakresie prowadzenia krajowej polityki rozwoju).

Wnioski i rekomendacje z ewaluacji ex post stanowią jedno z ważniejszych źródeł informacji przy opracowywanych obecnie założeniach systemu programowania Polityki Spójności w perspektywie 2014-2020. Dodatkowo wyniki te zostały wykorzystane przy opracowywaniu zintegrowanych strategii rozwoju przygotowywanych przez poszczególne ministerstwa. Wnioski z procesu ewaluacji pomagają również na bieżąco kształtować polskie stanowisko negocjacyjne. Wyniki badań są jednocześnie istotnym argumentem za kontynuowaniem tej polityki w przyszłości (bardzo ważną rolę w tym zakresie odegrała jedna z ewaluacji cząstkowych (Instytut Badań Strukturalnych, 2010) szacująca korzyści, jakie w wyniku realizacji Polityki Spójności w Polsce uzyskują płatnicy netto).

3.4. Mocne i słabe strony badania oraz wnioski dla przyszłych prac badawczych

Krajowa Jednostka Oceny przystępując do projektowania ewaluacji ex post była o tyle w komfortowej sytuacji, że mogła czerpać doświadczenia z realizacji badania na poziomie Komisji Europejskiej. Analizując mocne strony ewaluacji ex post realizowanej w Polsce można wskazać ponadto:

- przeprowadzenie końcowej metaewaluacji wewnętrznie przez Krajową Jednostkę Oceny. Dzięki temu możliwe było dostosowanie wniosków i rekomendacji do aktualnych potrzeb informacyjnych.

¹ www.ewaluacja.gov.pl/zakladka/Ewaluacja_ex_post_NPR_2004-2006/

² www.konferencja-ewaluacja.pl/2010/

Dodatkowo sam język raportu był dostosowany do głównych odbiorców, którymi byli pracownicy administracji;

- przygotowanie we współpracy z decydentami (na bazie raportu końcowego), kilkunastostronicowego dokumentu z samymi wnioskami i rekomendacjami z realizacji Polityki Spójności w perspektywie 2004-2006. Zapewniło to wysoki stopień identyfikacji decydentów z wypracowanym (i dostosowanym w formie do tychże odbiorców) dokumentem;
- istotną, mocną stroną badania realizowanego w Polsce była kompleksowa ocena Polityki Spójności: EFS, EFRR i tam, gdzie to możliwe Funduszu Spójności. Dało to znacznie lepszy i pełniejszy obraz efektów, a także szerszy katalog wniosków i rekomendacji.

Słabą stroną badania realizowanego w Polsce był ograniczony dostęp do danych (w szczególności dotyczących efektów rzeczowych – pokłosie słabości systemu monitorowania / brak systemu IT). W ramach badania nie wyodrębniono też osobnego „pakietu”, który zebrałby i udostępnił materiał wszystkim pozostałym ewaluatorom. W związku z tym badacze zbyt dużo czasu spędzili na „pogoni” za informacjami, zamiast na ich analizie.

Generalnie słabiej też wypadły badania realizowane w wymiarze terytorialnym i dotyczące określonych obszarów (np. rozwoju Polski Wschodniej). Biorąc pod uwagę ilość czynników oddziałujących na zmiany społeczno-gospodarcze na obszarze terytorialnym, którego dotyczyło badanie, wyizolowanie czystego wpływu Polityki Spójności okazywało się zbyt trudne. Środki UE miały ograniczony wpływ (z uwagi na swoją relatywnie niewielką skalę) na ogólne zmiany w wymiarze terytorialnym. Jednocześnie metody zaproponowane przez ewaluatorów zewnętrznych okazały się zbyt mało wrażliwe na uchwycenie tego wpływu. Znacznie lepiej wypadły badania realizowane sektorowo, ale starające się umieścić efekty i ich analizę w przestrzeni. O sukcesie decydowała tutaj możliwość precyzyjnego uchwycenia efektów (metodologia badania dostosowana do typu interwencji), ich właściwa analiza w przestrzeni okazywała się już łatwiejszym zadaniem.

Na koniec warto wskazać współlistnienie dwóch procesów: zakończenie realizacji pierwszej perspektywy finansowej UE i towarzyszącej jej oceny efektów, z równoległe toczącymi się pracami nad uporządkowaniem krajowych dokumentów strategicznych. Dzięki temu możliwe było wykorzystanie wyników ewaluacji ex post nie tylko dla projektowni przyszłej Polityki Spójności, lecz również – do pewnego stopnia – wykorzystanie wniosków do strategicznego zaprogramowania krajowej polityki rozwojowej.

Wnioski wynikające z metaewaluacji przygotowanej w Polsce w ramach procesu ewaluacji ex post są zbieżne z tymi, które wynikają z badania KE. Po pierwsze sama metaanaliza w ramach raportu końcowego, spełniła pokładane w niej oczekiwania. Po drugie jakość badania (metaewaluacji) uzależniona jest w największym stopniu od jakości ewaluacji cząstkowych. Po trzecie wskazane wydaje się planowanie realizacji ostatniej fazy (samej metaewaluacji) siłami wewnętrznymi. Kompetentni pracownicy administracji gwarantują odpowiednią jakość metaanalizy, a równocześnie istotnie podnoszą potencjalną użyteczność wyników (dostosowując treść raportu do potrzeb decydentów). Bardzo dobrym rozwiązaniem okazało się przygotowanie bardzo krótkiego dokumentu – treścią i formą – dostosowanego do potrzeb decydentów.

4. Wnioski. Perspektywa wykorzystania metaewaluacji w przyszłości

Dwa powyższe przykłady jednoznacznie dowodzą skuteczności metaewaluacji opartej na szeregu badań tematycznych w wypracowywaniu wniosków dla przyszłych interwencji publicznych. W przypadku polityk kompleksowych, zintegrowanych, do których zaliczyć należy Politykę Spójności, jest to bardzo dobry sposób pozwalający na przeprowadzenie pogłębionej i rzetelnej analizy, a jednocześnie umożliwiającą wypracowanie bardziej generalnych wniosków co do efektów i kierunków zmian całej polityki. Takie strategiczne wnioski mogą zostać wyciągnięte dopiero po przeprowadzeniu analizy wyników szeregu badań

szczegółowych, które się wzajemnie uzupełniają lub weryfikują. Ponadto wnioski z badania metaewaluacyjnego są znacznie silniejsze. Zmniejszamy bowiem ryzyko, bazując na szeregu ewaluacji, a nie na pojedynczych wynikach badania, gdzie znacznie łatwiej o błąd oceny.

Podstawą przygotowania wysokiej jakości metaewaluacji są wysokiej jakości badania pierwotne³. W tym celu, myśląc już o kolejnej perspektywie finansowej i jej ewaluacji, musimy odpowiedzieć sobie na kilka pytań: co, w jaki sposób i kiedy będziemy chcieli badać?

Odpowiedź na te pytania istotna jest zwłaszcza w systemach hierarchicznych (a takim jest system wdrażania i ewaluacji Polityki Spójności). Wszystkie zaangażowane podmioty muszą mieć jasność, co jest celem realizowanych interwencji. Najważniejsze jest więc skoncentrowane programowanie, wąsko i precyzyjnie adresujące wsparcie. Absolutnie kluczowe jest w tym zakresie precyzyjne zdefiniowanie pożądanych rezultatów. Jasność i precyzyjność celów możemy wzmocnić poprzez dobre zdefiniowanie zestawu wskaźników rezultatu dla kwantyfikacji pożądanych efektów. Wówczas będziemy też mogli odwoływać się do tych definicji na etapie projektowania badań ewaluacyjnych.

Następnym etapem jest określenie metodologii badania wpływu interwencji na zmianę tych wskaźników. Tam, gdzie to wskazane, na szerszą skalę powinny być stosowane metody wykorzystujące grupy porównawcze. Uzyskane w ten sposób wyniki badań są bardzo mocne i łatwo poddają się dalszym analizom (w tym, wspomnianym we wprowadzeniu, rygorystycznym metaanalizom). W tym celu, już od początku uruchomienia programu, konieczne jest zapewnienie gromadzenia odpowiednich informacji dotyczących zarówno beneficjentów, jak i potencjalnych beneficjentów, którzy nie zostali objęci daną interwencją. Nie należy oczywiście zapominać o *theory based evaluation*, która dostarcza niezwykle istotnych informacji, dlaczego dana interwencja działa, w jakim kontekście uzyskano zmierzone efekty, jakie niezamierzone efekty udało nam się osiągnąć i wreszcie, jakie są pola dalszego wykorzystania wyników. Konieczne jest następnie przyjęcie i stosowanie przez wszystkie instytucje wypracowanego podejścia do badania wpływu.

Ostatni etap to określenie właściwego harmonogramu realizacji badań na wszystkich poziomach wdrażania. Harmonogramu, który z jednej strony umożliwi rzetelny pomiar pojawiających się efektów, z drugiej będzie użyteczny dla decydentów w kontekście planowanych potrzeb informacyjnych. Harmonogram ten będzie musiał spełnić jeszcze jedną istotną rolę – koordynującą prace w ramach różnych programów operacyjnych, tak, by pozwolić w określonym momencie na dokonanie metaewaluacji/metaanaliz wyników i wyciągnięcia strategicznych wniosków.

Rolę instrumentu koordynacji tych wszystkich prac powinny pełnić, obligatoryjne w przyszłym okresie programowania, plany ewaluacji. Zanim jednak przystąpimy do opracowania planów, będziemy musieli precyzyjnie odpowiedzieć sobie na pytania, co, jak i kiedy będziemy chcieli badać. Odpowiedź na te pytania będzie musiała być gotowa zanim jeszcze przystąpimy do wdrażania kolejnej perspektywy. Tylko w ten sposób będziemy mogli dokonywać rzetelnej, użytecznej ewaluacji, czerpiąc wiedzę z całego procesu i tworząc profesjonalną bazę wiedzy. Bazę wiedzy wykorzystywaną następnie w analizach polityk publicznych, urzeczywistniając tym samym zasadę *evidence-based policy*. Jak pokazują dotychczasowe doświadczenia, metaewaluacja (skoncentrowana na analizie efektów) może odegrać kluczową rolę w tym zakresie.

Stanisław Bienias – absolwent Szkoły Głównej Handlowej w Warszawie oraz podyplomowych studiów z zakresu ewaluacji projektów na Uniwersytecie Warszawskim. Od siedmiu lat pracuje w Departamencie Koordynacji Polityki Strukturalnej Ministerstwa Rozwoju Regionalnego, od 2006 roku na stanowisku naczelnika Wydziału Ewaluacji (Krajowa Jednostka Oceny). Odpowiedzialny za realizację procesu ewaluacji na po-

³ Odwieczny problem *garbage in, garbage out*. Jeżeli bazujemy w swoich analizach na „śmieciowych danych” otrzymujemy „śmieciowe wyniki”

ziomie Narodowego Planu Rozwoju na lata 2004-2006 i Narodowych Strategicznych Ram Odniesienia na lata 2007-2013 oraz za koordynację procesu ewaluacji na poziomie programów operacyjnych realizowanych w latach 2004-2006 i 2007-2013. Reprezentuje Polskę w spotkaniach „Evaluation Network” w ramach Dyrekcji Generalnej do Spraw Regionalnych w Komisji Europejskiej, współpracuje z OECD i innymi instytucjami krajowymi i zagranicznymi. Prowadzi wykłady na studiach podyplomowych z zakresu ewaluacji, monitoringu i wskaźników. Współautor wielu publikacji w obszarze ewaluacji i funduszy strukturalnych.

Bibliografia

- Bustelo, M. (2002). *Metaevaluation as a tool of the improvement and development of the evaluation function in public administrations*.
- Cook, T.D., Gruder, C.L. (1978). *Meta-evaluation research*. Evaluation Quarterly.
- EGO s.c, 2010, *Ocena systemu realizacji Polityki Spójności w Polsce w ramach perspektywy 2004-2006*, EGO s.c na zlecenie MRR.
- EUROREG, 2010, *Ocena wpływu Polityki Spójności na rozwój miast polskich*; Centrum Europejskich Studiów Regionalnych i Lokalnych EUROREG Uniwersytetu Warszawskiego na zlecenie MRR.
- FUNDEKO S.C., 2010, *Ocena efektów inwestycji środowiskowych finansowanych w ramach NPR 2004-2006*, FUNDEKO S.C na zlecenie MRR.
- Gaffey, V., *Future of Cohesion Policy – building on evidence*, prezentacja w trakcie konferencji ewaluacyjnej w Warszawie, 13 grudnia 2010, www.konferencja-ewaluacja.pl/2010/
- IBC Group Central Europe Holding, Białostocka Fundacja Kształcenia Kadr, 2010, *Ocena ex post wpływu realizacji Polityki Spójności perspektywy 2004-2006 na zwiększenie możliwości rozwoju społeczno-gospodarczego regionów Polski Wschodniej*; Konsorcjum w składzie: IBC Group Central Europe Holding S.A., Białostocka Fundacja Kształcenia Kadr; na zlecenie MRR.
- Instytut Badań Strukturalnych, Reytech, 2010, *Ocena wpływu Polityki Spójności na wzrost konkurencyjności i innowacyjności polskich przedsiębiorstw i gospodarki*, Instytut Badań Strukturalnych & Reytech Sp. z o.o.; na zlecenie MRR.
- Instytut Badań Strukturalnych, 2010, *Ocena korzyści uzyskiwanych przez Państwa UE-15 w wyniku realizacji Polityki Spójności w Polsce – Aktualizacja 2010*; Instytut Badań Strukturalnych na zlecenie MRR.
- Instytut Badań Strukturalnych, Reytech, ASM, 2010, *Ocena wpływu realizacji interwencji współfinansowanych z EFS na podniesienie jakości zasobów ludzkich*; Instytut Badań Strukturalnych & Reytech Sp. z o.o., ASM Sp. z o.o.; na zlecenie MRR.
- Komisja Europejska, 2008, *Work package 2: Data feasibility study*; ADE s.a.
- Komisja Europejska, 2009(a), *Work Package 10: Efficiency: Unit costs of major projects*; RGL Forensics, Frontier Economics, Faber Maunsell AECOM.
- Komisja Europejska, 2009(b), *Work Package 11: Management and implementation systems for cohesion policy*; EPRC, METIS, University of Strathclyde Glasgow.
- Komisja Europejska, 2009(c), *Work package 3: A Model-based Analysis of the Impact of Cohesion Policy Expenditure 2000-06: Simulations with the QUEST III endogenous R&D mode*.
- Komisja Europejska, 2009(d), *Work package 3: Analysis of EU Cohesion Policy 2000-2006 using the CSHM: Aggregate impacts and inter-country comparison*; John Bradley EMDS, Gerhard Untiedt GEFRA.
- Komisja Europejska, 2009(e), *Work package 3: Analysis of EU Cohesion Policy 2000-2006 using the CSHM: Sectoral impacts*; John Bradley EMDS, Gerhard Untiedt GEFRA;
- Komisja Europejska, 2009(f), *Work Package 5b Environment and Climate Change*; ADE s.a, AGROTEC, BIO Intelligence Service, Orbicon.
- Komisja Europejska, 2009(g), *Work Package 6a: Enterprise and innovation*; Mikkel Holm-Pedersen, Jeremy Millard, Kristian Pedersen Policy and Business Analysis.
- Komisja Europejska, 2009(h), *Work package 7: Effectiveness of the Cohesion Policy: Gender Equality and Demographic Change*; IRS, CSIL.
- Komisja Europejska, 2009(i), *Work package 8: Note on the suitability of the TRANSTOOLS Model for the performance of ex-post evaluations of local transport infrastructure projects*, KE.
- Komisja Europejska, 2009(j), *Work Package 9: Rural Development*; METIS.
- Komisja Europejska, 2010(a), *Ex Post Evaluation of Cohesion Policy Programmes 2000-2006 financed by the European Regional Development Fund in Objective 1 and 2 Regions*; APPLICA, ISMERI EUROPA, WIIW The Vienna Institute for International Economic Studies.

- Komisja Europejska, 2010(b), *Work Package 4 "Structural Change and Globalisation"*; CSIL, Centre for Industrial Studies, Joanneum Research, Technopolis Group.
- Komisja Europejska, 2010(c), *Work Package 5a Transport*; Steer Davies Gleav.
- Komisja Europejska, 2010(d), *Work Package 6b Ex Post Evaluation Of Cohesion Policy Programmes Financed By The ERDF Enterprise Support*; RAMBOLL.
- Komisja Europejska, 2010(e), *Work Package 6c: Enterprise Support – an exploratory study using counterfactual methods on available data from German*; GEFRA , IAB.
- MRR, 2010(a), *Wnioski z realizacji Polityki Spójności w świetle wyników prac analityczno-ewaluacyjnych okresu 2004–2006*.
- MRR, 2010(b), *Ewaluacja ex post Narodowego Planu Rozwoju 2004–2006*.
- Olejniczak K. *Ewaluacja i monitoring (w:) System zarządzania rozwojem Polski*, red. Górniak, J., Mazur, S. MRR (w druku).
- PAG Uniconsult, 2010, *Wpływ Polityki Spójności na poziom i jakość zatrudnienia w Polsce*; PAG Uniconsult na zlecenie MRR.
- PAN Instytut Geografii i Przestrzennego Zagospodarowania, 2010, *Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach Polityki Spójności na wzrost konkurencyjności regionów*; Polska Akademia Nauk Instytut Geografii i Przestrzennego Zagospodarowania Im. S. Leszczyńskiego na zlecenie MRR.
- Pierre, R.G., 1982, *Follow Through: A case study in meta-evaluation research. Educational Evaluation and Policy Analysis*.
- PSDB grupa WYG, 2010, *Komplementarność i synergia interwencji realizowanych w ramach Europejskiego Funduszu Społecznego i Europejskiego Funduszu Rozwoju Regionalnego w perspektywie finansowej 2004-2006*; PSDB grupa WYG na zlecenie MRR.
- PSDB, CRSG, 2010, *Efekty transgranicznej współpracy polskich regionów w okresie 2004 -2006*; konsorcjum PSDB Sp z o.o. oraz CRSG Sp. z o.o.; na zlecenie MRR; 2010.
- Scriven, M., 1991, *Evaluation thesaurus (4th ed.)*. Newbury Park, CA: Sage.
- Skórska P., Strzebińska A., *Metaanaliza – wyzwania i nowe obszary zastosowania*, Centrum Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego, prezentacja w trakcie XIV Ogólnopolskiego Zjazdu Socjologicznego w Krakowie 2010 r.
- Żuber P., *Ewaluacja ex post Narodowego Planu Rozwoju 2004-2006: efekty, wnioski, rekomendacje*, VI Konferencja Ewaluacyjna 13 grudnia 2010, www.konferencja-ewaluacja.pl/2010/

Wprowadzenie do podstawowych pytań stawianych w ewaluacji oddziaływania interwencji: *Czy interwencja coś zmieniła? A jeśli tak, to dla kogo?*

Wprowadzenie

W okresie coraz większych cięć budżetowych i wzrastającej kontroli publicznej, ewaluacja zyskuje na znaczeniu dla skuteczniejszej i wydajniejszej alokacji środków. Obecnie w ewaluacji polityk, programów i projektów – tj. *ocenie zmierzającej do ich udoskonalenia*, stosuje się wiele podejść i narzędzi analitycznych. Dostępność wielu zróżnicowanych podejść jest korzystna, gdyż mogą one dostarczyć odpowiedzi na szereg różnych pytań. Niektóre z tych pytań dotyczą procesu realizacji polityk, programów i projektów, inne – zgodności ze standardami i postawionymi celami. W innych przypadkach celem ewaluacji jest po prostu monitorowanie wykorzystania funduszy i tego, w jaki sposób przekładają się na powstawanie produktów – priorytetem jest tu zapewnienie transparentności w realizacji polityk publicznych.

Dlaczego pytania dotyczące oddziaływania są ważne?

Kluczowe pytanie ewaluacyjne dotyczy ustalenia, na ile realizowane programy przynoszą pożądane efekty. Innymi słowy, czy interwencja zmieniła coś w poziomie życia jednostek, do których była skierowana. Dla kogo dane rozwiązanie miało (i będzie miało) korzystny wpływ i jakie mechanizmy się do tego przyczyniły?

Ta informacja, mimo że jest dość „suchą” miarą powodzenia, ma zasadnicze znaczenie podczas szacowania efektywności kosztowej ocenianego działania. Gdy N beneficjentów otrzymuje wsparcie, ale korzyści z interwencji czerpie tak naprawdę jedynie M jednostek, stosunek kosztów do korzyści powinien być rozumiany jako wynik działania C/M . Może się okazać, że jest to znacznie większa liczba (tj. mniej zadowalająca) niż wynik działania C/N , który odzwierciedla po prostu sposób rozdzielenia dostępnych środków. Jeśli okaże się, że $M=0$, powinno to rodzić pewne wątpliwości co do przydatności interwencji.

Oddziaływanie interwencji może być zróżnicowane w zależności od typu beneficjentów. Stwierdzenie to jest kluczowe z punktu widzenia lepszego ukierunkowania interwencji. Decydenci polityczni najczęściej podejmują decyzje w oparciu o rozważania dotyczące intensywności wsparcia (*Do kogo powinny być skierowane konkretnego rodzaju zachęty? Na jakiego rodzaju firmach powinniśmy skupić naszą uwagę?*) niż jego zakresu (*Czy program powinien zostać utrzymany, czy może powinniśmy z niego całkowicie zrezygnować?*). Naszym zdaniem znaczenie ewaluacji oddziaływania interwencji dla polityki zawsze zależy od tego, na ile umożliwia ona oszacowanie różnych skutków interwencji w odniesieniu do zastosowanych narzędzi oraz charakterystyk odbiorców. I tak na przykład, ukierunkowywanie interwencji poprzez ograniczanie kwalifikowalności przedsiębiorstw (ze względu na rozmiar czy sektor działalności) oraz określenie zakresu wsparcia, stanowią główne instrumenty projektowania wsparcia dla przedsiębiorstw.

Niniejszy tekst omawia metody pozwalające określić, czy dobrze zdefiniowana interwencja przynosi pożądaną skutki w interesującym nas zakresie. Istotą problemu jest „przypisanie” obserwowanej zmiany do konkretnej interwencji. *Czy zmiana wynika z wprowadzenia danej polityki czy wystąpiłaby i tak bez niej?* Udzielenie odpowiedzi na te pytania nie jest tak proste, jak mogłoby się wydawać. Znaleźnię wiarygodnego przybliżenia tego, co wydarzyłoby się, gdyby nie nastąpiła interwencja (tj. w „sytuacji kontrfaktycznej”) i porównania go z tym, co faktycznie się wydarzyło, jest największym wyzwaniem z punktu widzenia pomiaru efektów działań. Różnica stanowi przybliżony efekt lub oddziaływanie interwencji pod kątem konkretnego interesującego nas rezultatu (może to być np. wartość PKB na jednego mieszkańca, nakłady na działalność B+R, wskaźniki dotyczące mieszkalnictwa lub poziom zatrudnienia).

Uporządkowanie języka wykorzystywanego przy ocenie oddziaływania interwencji

Koncepcja oddziaływania cieszy się dużym zainteresowaniem, często jednak rozumiana jest nieprawidłowo. Dwoma najczęściej popełnianymi (lecz możliwymi do uniknięcia) błędami są: stosowanie niepotrzebnego rozróżnienia na efekty (*effects*) i oddziaływanie (*impacts*) oraz dość mglistej dekompozycji efektu brutto mierzonego za pomocą wskaźników oddziaływania (*impact indicators*).

Efekty i oddziaływanie: rozróżnienie bez znaczenia

Naszym zdaniem pierwszy błąd polega na tym, że pojęciom „oddziaływania” i „efektów” często przypisuje się różne znaczenie (nierzadko stosowane wymiennie). Zgodnie z jedną z najbardziej popularnych definicji, „efekty” oznaczają natychmiastowe rezultaty dla bezpośrednich beneficjentów, a „oddziaływanie” to efekty osiągnięte w dłuższej perspektywie czasowej i/lub mające zastosowanie do szerszej, zwykle bliżej nieokreślonej populacji.

Gdy zastosujemy podejście kontrfaktyczne, nie ma tak naprawdę znaczącej różnicy między tymi dwoma pojęciami. By podkreślić ich równoważność, mówimy o efektach/oddziaływaniu „przyczynowym”, czyli o różnicy pomiędzy rezultatem interwencji i rezultatem, który nastąpiłby w przypadku braku interwencji. To, czy przyjmujemy krótszą czy dłuższą perspektywę czasową oraz to, czy odniesiemy się do beneficjentów bezpośrednich lub pośrednich ma istotne znaczenie dla sposobu przeprowadzenia ewaluacji, ale rozróżnienie na efekty i oddziaływanie samo w sobie nie jest szczególnie przydatne.

Od oddziaływania „brutto” poprzez efekt bezwładności (*deadweight*) do oddziaływania „netto”

We wszystkich dokumentach powtarza się rozbitcie „oddziaływania brutto” (lub „efektów brutto”) na kategorie „efektu bezwładności” (efekt *deadweight*), „efektu substytucyjnego”, „efektu przeniesienia” i „efektu netto”. Stosowane rozbitcie nigdy nie znajduje odzwierciedlenia w wiarygodnych oszacowaniach poszczególnych pozycji. Wprawdzie na pierwszy rzut oka taka kategoryzacja może zdawać się atrakcyjna, kryje ona jednak w sobie brak rygoru analitycznego i jasnego sprecyzowania celu.

Oba elementy dotyczące podmiotów niebędących beneficjentami – „substytucja” i „przeniesienie” mogą znaleźć zastosowanie na gruncie teoretycznym, są one jednak bardzo trudne do oszacowania z tego prostego powodu, że podmioty, które padły „ofiarami” efektów substytucyjnego i przeniesienia nie są nam znane. Możemy zakładać ich istnienie, ale nie możemy zmierzyć ich statusu (często chodzi o status zatrudnienia), nie wspominając już o zmierzeniu tego, jak wpłynęła na niego dana interwencja. Ten „niewielki” szczegół jest regularnie ignorowany przez zwolenników dekompozycji efektu „brutto”.

Należy również zauważyć, że efekt bezwładności także nie może być w prosty sposób wydzielony z obserwowanej zmiany. Nim zostanie on odseparowany, konieczne jest bowiem jego oszacowanie. Oka-

zuje się, że pojęcie bezwładności to szczególnie przypadek sytuacji kontrfaktycznej, a zatem jest ona równie trudna do zidentyfikowania i oszacowania. Niestety, pojęcie bezwładności w literaturze dotyczącej ewaluacji traktuje się powierzchownie: koncepcja, według której bezwładność można po prostu „zmierzyć” wzmocniła przekonanie, że „efekty netto” również można obliczyć poprzez odseparowanie obserwowanych wartości. Jest to stwierdzenie nieprawdziwe: efekty nie są możliwe do zaobserwowania. Możemy wnioskować na ich temat na podstawie odpowiednich danych i przyjętych założeń. Zatem, jeśli sytuacja kontrfaktyczna nie jest możliwa do zaobserwowania, nie możemy również zaobserwować efektu bezwładności.

Pomimo swoich słabych stron i braku zastosowań empirycznych, rozłożenie „efektu brutto” na czynniki stanowi wciąż, niestety, jedyny sposób konceptualizacji ewaluacji oddziaływania w europejskiej literaturze dotyczącej ewaluacji – szczególnie tej na temat funduszy strukturalnych.

Kolejnym przykładem niespójności między teorią i stosowanymi metodami jest często przywoływane stosowanie (i nadużywanie) wskaźników oddziaływania. Czysto opisowe narzędzie, jakim jest kwantyfikujący dane zjawisko wskaźnik, po dodaniu do niego słowa „oddziaływanie”, w sposób niemal magiczny staje się narzędziem umożliwiającym ustalenie przyczynowości. Jak słusznie zauważono w Raporcie Barca: *Wskaźniki rezultatów i ewaluacja oddziaływania odpowiadają na dwa zupełnie różne problemy. Wskaźniki mają służyć skupieniu uwagi decydentów politycznych i opinii publicznej na celach oraz bieżącemu monitorowaniu realizacji celów w trakcie wdrażania interwencji. Z kolei ewaluacja to narzędzie, dzięki któremu dowiadujemy się, czy dana interwencja odniosła skutek w danym obszarze. Wbrew dotychczasowej praktyce tych dwóch funkcji nie należy mylić.*

Wskaźniki „oddziaływania” często stosowane są do opisywania problemów społecznych/gospodarczych, na które dany program ma mieć wpływ. Wskaźniki te mogą być porównywane dla różnych programów lub też w różnych okresach dla tego samego programu, aby pokazać, czy w odniesieniu do danego problemu następuje poprawa czy pogorszenie sytuacji. Istnieje potrzeba wyznaczenia punktu bazowego i wyraźnych celów oraz ich okresowego monitorowania. Nie istnieje jednak potrzeba wnioskowania przyczynowego.

Główne podejścia do ewaluacji oddziaływania

W celu ustalenia sytuacji kontrfaktycznej i oszacowania oddziaływania programu, stosowane są trzy podstawowe podejścia: (i) wykorzystanie jako sytuacji kontrfaktycznej rezultatu odnotowanego dla beneficjentów zanim przeprowadzona została interwencja (porównanie „przed i po”); (ii) pytanie beneficjentów o to, co postrzegają jako sytuację kontrfaktyczną lub po prostu o efekt interwencji (badanie z udziałem beneficjentów); (iii) wykorzystanie jako sytuacji kontrfaktycznej rezultatów zaobserwowanych dla innych zbliżonych podmiotów niebędących beneficjentami, lub takich których „upodobniono” do beneficjentów (analiza grupy porównawczej). Wprawdzie trzecie podejście ze swoimi licznymi wariantami jest stosowane najczęściej, każde z nich ma swoje mocne i słabe strony.

Z drugiej jednak strony, ponieważ informacje dotyczące przyczynowości są istotne dla decydentów politycznych (którzy rozważają, czy program powinien być kopiowany/utrzymany/zmodyfikowany czy zakończony), abyśmy mogli zrozumieć „co działa”, musimy opierać się na solidnych szacunkach przyczynowych, a nie na prostych związkach między naszymi działaniami i rezultatami. I tak na przykład, czy różnica pomiędzy nakładami na B+R pomiędzy firmami, które otrzymały dofinansowanie i tymi, które go nie otrzymały świadczy o skutkach otrzymanej dotacji, czy po prostu pozwala nam dowiedzieć się, w jaki sposób firmy dokonują autoselekcji, podejmując decyzję o złożeniu wniosku o dofinansowanie? Sednem ewaluacji oddziaływania jest interpretacja różnic. Wyzwaniem dla ewaluatora jest uniknięcie przyczynowej interpretacji różnic wynikających z czynników innych niż sama interwencja. To właśnie jest istotą kontrfaktycznej ewaluacji oddziaływania.

Porównania „przed i po” w odniesieniu do beneficjentów

Przyjrzymy się pierwszemu z trzech podstawowych podejść, a mianowicie porównaniom typu „przed i po”. Gdy ten sam podmiot jest obserwowany zarówno przed, jak i po interwencji, podstawowym problemem ewaluacji jest to, że obserwowana zmiana może wynikać zarówno z samej interwencji, jak i z innych zmian zachodzących w tym samym okresie. Gdy nie dysponujemy dodatkowymi informacjami, tych dwóch elementów nie da się rozdzielić.

Zjawiska społeczne i gospodarcze, na które ukierunkowane są polityki publiczne, ewoluują w czasie niezależnie od wdrażanych polityk. Trzymając się wcześniejszego przykładu, porównalibyśmy średnie nakłady na działalność B+R ponoszone w roku poprzedzającym złożenie przez firmy wniosków o dofinansowanie i te ponoszone rok po zakończeniu realizacji projektu (by uniknąć uwzględnienia dotacji jako części nakładów). Czy otrzymanie dofinansowania ma długotrwały wpływ na skłonność przedsiębiorstw do inwestowania w działalność badawczo-rozwojową (B+R)? Czy firmy, których nakłady na działalność B+R wzrosły, są tymi samymi firmami, które wnioskowały o dofinansowanie?

Wiarygodność szacunków dotyczących oddziaływania interwencji na podstawie porównań „przed i po” jest bardzo niska. Niejednokrotnie niewiele można zrobić poza przyjęciem założenia czasowej stabilności, tj. braku zmian i naturalnej ewolucji. Różnice „przed i po” nie pokazują prawdziwych efektów interwencji, chyba że założymy pełną stabilność — tj. brak „naturalnej dynamiki”.

Pytanie beneficjentów o ich zdanie na temat oddziaływania interwencji

Dość popularną metodą jest po prostu pytanie danej grupy beneficjentów o to, jak postrzegają oddziaływanie interwencji. Po uśrednieniu odpowiedzi można dalej wnioskować na temat oszacowanego oddziaływania interwencji. W naszym przykładzie przyjęcie tego podejścia wiązałoby się z zapytaniem próby beneficjentów o ich wydatki na działalność B+R w roku poprzedzającym dotację i rok po otrzymaniu dotacji i o to, czy dofinansowanie miało wpływ na ich ewentualny wzrost. Alternatywnie, można byłoby zadać pytanie z „kontraktycznym zabarwieniem”: „Czy nakłady na działalność badawczo-rozwojową uległyby zmianie, nawet gdyby Państwa przedsiębiorstwo nie otrzymało dotacji?”

Zarówno słabe, jak i mocne strony tego podejścia są widoczne na pierwszy rzut oka. Wśród zalet można wymienić konieczność przeprowadzenia tylko jednego wywiadu wśród beneficjentów, dla których administratorzy programu zazwyczaj posiadają dane identyfikacyjne. Do wad można zaliczyć fakt, że podejście takie opiera się na wrażeniach osób pytanym, nakłada przy tym na nich znaczne obciążenie poznawcze oraz wiąże się z niebezpieczeństwem uzyskania odpowiedzi nieobiektywnych, gdy beneficjenci uznają, że w ich interesie leży udzielenie „pozytywnych” odpowiedzi.

Porównanie beneficjentów i podmiotów niekorzystających ze wsparcia ocenianej interwencji

Najbardziej typową strategią szacowania efektu przyczynowego interwencji jest wykorzystanie faktu, że niektóre „podmioty” (osoby, rodziny, firmy i miejsca) zostały objęte interwencją, a inne nie – zgodnie z określonym procesem *autoselekcji*.

I tak na przykład, kwalifikujące się do wsparcia, przedsiębiorstwa mogą lecz nie muszą wnioskować o pomoc publiczną na finansowanie inwestycji, bezrobotni mogą lecz nie muszą uczestniczyć w programie przekwalifikowania pracowników po zamknięciu zakładu, a dzielnice miejskie mogą lecz nie muszą otrzymywać dofinansowania do projektów renowacji miasta. Mimo że nie można zlekceważyć istnienia polityk uniwersalnych, w przypadku Polityki Spójności są one stosunkowo rzadkie. W większości przypadków możliwe jest zidentyfikowanie podmiotów, które nie zostały objęte daną polityką, programem, projektem. W celu uproszczenia, rozważamy jedynie przypadek dwubiegunowy, w którym podmioty zostały lub nie zostały objęte działaniem danej interwencji.

Rezultaty obserwowane wśród beneficjentów można porównać do tych obserwowanych wśród podmiotów niebędących beneficjentami. Różnica ta jednak sama w sobie nie obrazuje faktycznego oddziaływania interwencji. Niekoniecznie musi być interpretowana w sensie przyczynowym. Interpretacja przyczynowa zależy od charakteru procesu, który powoduje, że niektóre podmioty są objęte interwencją, a inne nie. Sytuacją idealną jest, gdy proces selekcji jest niezależny od efektów uzyskiwanych przez poszczególne podmioty. Można to zagwarantować tylko w przypadku, gdy selekcja beneficjentów ma charakter losowy. Gdy selekcja nie ma charakteru losowego, obserwowaną różnicę zawsze można postrzegać jako sumę dwóch elementów: faktycznego efektu polityki oraz różnicy wynikającej z samego procesu selekcji. Oba z tych elementów są hipotetyczne i żadnego z nich nie możemy zaobserwować.

Przykładowo, w przypadku dotacji dla przedsiębiorstw na inwestycje w działalność B+R na różnice w wynikach firm, które otrzymały dofinansowanie i tych, które go nie otrzymały składają się: faktyczny przyczynowy efekt dotacji (może być zerowy) oraz różnice wynikające z procesu selekcji, który dzieli kwalifikujące się przedsiębiorstwa na beneficjentów i te, które beneficjentami nie zostały. Jest prawdopodobne, że firmy, które otrzymały dofinansowanie i te, które go nie otrzymały będą różnić się pod względem nakładów na działalność B+R, nawet w przypadku, gdyby firmy tej pierwszej kategorii nie otrzymały wsparcia.

Wylimitowanie obciążenia selekcyjnego stanowi duże wyzwanie dla ewaluacji oddziaływania interwencji. Zagadnienie to cieszy się dużym zainteresowaniem statystyków, ekonomistów i socjologów. Dostępnych jest wiele metod i technik umożliwiających poradzenie sobie z tym wyzwaniem. Aby zastosować odpowiednią metodę pomiaru oddziaływania interwencji, konieczna jest wiedza na temat procesu selekcji.

Randomizacja i eksperymenty: idealna strategia wylimitowania obciążenia selekcyjnego

Idealną strategią eliminowania obciążenia selekcyjnego jest losowy dobór podmiotów – beneficjentów i podmiotów niebędących beneficjentami. W tym przypadku mamy pewność, że obciążenie selekcyjne wynosi zero. Niestety nie zawsze można zastosować randomizację. Randomizacja umożliwia oszacowanie skutków działania spójnych wewnętrznie, ale trudnych do uogólnienia, a taka generalizacja jest kluczowa z punktu widzenia przydatności wyników dla kreowania polityk publicznych. Eksperymenty są zwykle kosztowne i wymagają uważnego monitorowania zapewniającego ich skuteczne przeprowadzenie. Ewentualna odmowa w zakresie udzielanego wsparcia często jest kontrowersyjna z etyczno-politycznego punktu widzenia. Kwestie te mogą spowodować, iż szanse na rozważenie eksperymentu jako metody ewaluacji programu zmniejszą się. Problematyczne może się również okazać zagwarantowanie współpracy z podmiotami odpowiedzialnymi za wdrożenie interwencji.

Randomizacja wymaga starannego planowania, wczesnego zaangażowania ewaluatora i pewnego stopnia stabilności środowiska, w którym przeprowadzany jest eksperyment. Te elementy są rzadko obecne w sektorze publicznym państw członkowskich UE. Randomizacja wymaga również, by interwencja miała dość prosty charakter, jednak polityki są najczęściej złożone, gdyż koncentrują się na problemach wielowymiarowych/ wielopoziomowych. Złożony charakter zagadnień jest zawsze utrudnieniem dla ewaluacji, lecz w przypadku randomizacji konflikt pomiędzy metodologią i kontekstem jest szczególnie uderzający.

Nieunikniony zwrot w kierunku metod nieeksperymentalnych

Biorąc pod uwagę powyższe, każda wiarygodna strategia ewaluacji oddziaływania danej polityki musi uwzględniać szereg metod nieeksperymentalnych (określanych również jako „quasi-eksperymentalne”). Pokrótkie przedstawione zostaną cztery główne nieeksperymentalne strategie, które okazały się w ostatnich latach najbardziej niezawodne i wiarygodne z punktu widzenia zmniejszenia obciążenia selekcyjnego.

Metoda *różnicy w różnicach* (podwójnej różnicy) to najprostszy sposób potraktowania problemu, opierający się na założeniu, że dysponujemy danymi dotyczącymi rezultatów dla beneficjentów i podmiotów niebędących beneficjentami zarówno sprzed interwencji, jak i po interwencji. Nieweryfikowalnym założeniem jest tutaj to, że trend wzrostu dla rezultatów obserwowanych u podmiotów niebędących beneficjentami stanowi rozsądne przybliżenie *tendencji kontrfaktycznej* („Co działoby się z beneficjentami w danym okresie, gdyby nie otrzymali dotacji?”). Jeśli uznamy założenie, że podmioty niebędące beneficjentami zachowują się podobnie do beneficjentów, szacunek oddziaływania interwencji na podstawie różnicy w różnicach lub podwójnej różnicy można otrzymać poprzez odjęcie różnicy rezultatów przed interwencją od różnicy po interwencji.

Kolejne dwie strategie można postrzegać jako dwutorowy sposób wyboru beneficjentów w celu zapewnienia ich większej reprezentatywności dla sytuacji kontrfaktycznej. Podejście oparte na metodach łączenia (*matching*) opiera się na koncepcji wyszukiwania innych podmiotów niebędących beneficjentami, które „przypominają” beneficjentów pod względem wszelkich (możliwych do zaobserwowania) charakterystyk: gdy odnajdziemy takie „klony”, efekt interwencji szacowany jest na podstawie średniej różnicy zmiennych dotyczących rezultatów beneficjentów i podmiotów przypominających beneficjentów, przy założeniu, że dopasowanie wyeliminowało różnice wynikające z selekcji. Założenia tego nie można zwerifikować, ale staje się ono tym bardziej wiarygodne, im więcej zmiennych – mogących mieć wpływ na proces selekcji – jest kontrolowanych w procesie dopasowania jednostek.

Podejście oparte na pojęciu nieciągłości (*discontinuity*) opiera się na koncepcji nieciągłości oddziaływania bodźca wokół wartości progowej. Dotyczy ono głównie sytuacji, w których niektóre podmioty kwalifikują się do korzystania z interwencji, a inne nie na podstawie precyzyjnie zdefiniowanej reguły, zwykle o charakterze administracyjnym. Obie grupy są do siebie podobne pod innymi względami, ale są wyraźnie rozdzielone w zależności od ich pozycji w odniesieniu do wartości progowej, przy czym po jednej stronie progu mamy podmioty objęte polityką, a po drugiej te nią nieobjęte. Efekt rozróżnienia uzyskujemy na podstawie różnicy rezultatów osiągniętych wokół wartości progowej. Na potrzeby identyfikacji przyjmuje się tutaj założenie (bardziej wiarygodne niż większość innych), że obciążenie selekcyjne wokół wartości progowej wynosi zero. U podstaw tego założenia leży koncepcja, że wokół wartości progowej osiągamy sytuację podobną do randomizacji. Należy zauważyć, że szacowany efekt ma charakter lokalny: jest on bardziej wiarygodny (większa spójność wewnętrzna), ale w mniejszym stopniu umożliwia generalizację (mniejsza spójność zewnętrzna).

Wracając do naszego przykładu. Możliwość zastosowania analizy nieciągłości zależy od tego, czy proces selekcji na potrzeby przyznania dotacji na działalność B+R polegał na stworzeniu rankingu wnioskodawców według ciągłej punktacji. W takim przypadku poprzez porównanie przedsięwzięć krańcowych, tych, które uwzględniono w rankingu i tych, które z niego wykluczono, uzyskujemy dość wiarygodne, aczkolwiek lokalne, szacunki dotyczące efektu otrzymanej dotacji.

Podejście oparte na zmiennych instrumentalnych (*instrumental variables*) bazuje na koncepcji *eksperymentu naturalnego*. Mowa tutaj o sytuacjach, w których czynnik zewnętrzny wyznacza poddanie danego elementu działaniu określonego bodźca. Strategię tę szczególnie trudno jest objaśnić w sposób intuicyjny bez zastosowania algebry. W praktyce potrzebne są dwa założenia. Pierwsze z nich dotyczy tego, że

na uczestnictwo w danej interwencji wpływa pewien czynnik zewnętrzny. Drugie, że czynnik zewnętrzny wpływa na rezultat tylko i wyłącznie pośrednio, poprzez wpływ na udział w danej interwencji. Założenia tego nie da się zweryfikować, a jego wiarygodność zależy od charakteru procesu selekcji. W ten sposób czynnik zewnętrzny wywołuje dwa efekty. Pierwszy dotyczy rezultatu, a drugi uczestnictwa. Żaden z tych efektów nie jest interesujący z punktu widzenia polityki. Interesuje nas efekt związany z udziałem w polityce. Można wykazać, że może być on oszacowany poprzez obliczenie stosunku dwóch efektów wywołanych przez czynnik zewnętrzny.

Trzy krople zdrowego rozsądku

Najbardziej interesującą miarą dla decydentów politycznych jest efektywność kosztowa. Przedstawienie oszacowanego efektu interwencji i jego istotności statystycznej jest niewystarczające. Dany program może sprawiać wrażenie efektywnego, ponieważ jego skutek jest pozytywny, jednak koszt gwarantujący istotny statystycznie efekt może być nieuzasadnienie wysoki.

Ewaluacja oddziaływania to proces kumulacyjny, który nigdy nie dostarcza ostatecznych odpowiedzi. Jeśli ma ona być procesem o charakterze kumulacyjnym, musi być w nią zaangażowane środowisko badaczy pracujących nad tym samym zagadnieniem, którzy w miarę możliwości udostępniają sobie te same dane. Ograniczanie dostępu do danych hamuje proces kumulacji wiedzy. Dzięki swobodnemu przepływowi danych w dłuższej perspektywie wszyscy odniesiemy korzyści.

Ważne, by decydenci polityczni byli świadomi niepewnego charakteru dokonywanych oszacowań. Ewaluatorzy powinni unikać kreowania nierealistycznych oczekiwań w odniesieniu do możliwości uzyskania „naukowych” i „obiektywnych” wyników. Wyniki ewaluacji mają charakter tymczasowy, mogą zawsze być udoskonalone, czasem są błędne, w szczególności w przypadku prób identyfikacji związków przyczynowo-skutkowych. *There is no magic counterfactual.*

Alberto Martini – Absolwent Wydziału Prawa Uniwersytetu w Turynie (1980), Doktor Ekonomii Uniwersytetu Wisconsin-Madison (1988). W latach 1988-1993 w Mathematica Policy Research – firmie badawczej specjalizującej się w ewaluacji polityk społecznych i wspierających zdrowie. W latach 1993-1998 – starszy badacz w Urban Institute, gdzie zajmował się kwestiami związanymi z dobrobytem oraz projektowaniem modeli mikrosymulacyjnych. Od 1998 – profesor statystyki i ewaluacji polityk na Uniwersytecie Piemonte Orientale. W latach 2001-2002 pełnił funkcję Prezesa Włoskiego Towarzystwa Ewaluacyjnego. Od 2007 r. członek Consiglio Italiano delle Scienze Sociali (Włoskiej Rady Nauk Społecznych), gdzie współprzewodniczy Komisji ds. Ewaluacji Oddziaływania.

Analiza kosztów i korzyści społecznych na przykładzie programów wspierania zatrudnienia

1. Analiza kosztów i korzyści – podstawowe założenia

Funkcja dobrobytu społecznego

Analiza kosztów i korzyści społecznych obejmuje finansowe efekty interwencji publicznej oraz ich wpływ na budżet państwa. Musi ona uwzględniać specyfikę efektów właściwych dla poszczególnych grup społecznych. Podstawowe założenie mikroekonomiczne, jakie przyjmujemy, mówi, że wszyscy ludzie, a tym samym społeczeństwo jako całość, kierują się w swoich działaniach zasadą maksymalizacji użyteczności⁴. Pomiar efektów opiera się zatem na zmianach poziomu dobrobytu, których ilustrację stanowi funkcja dobrobytu społecznego (SWF – *Social Welfare Function*). W najbardziej ogólnym ujęciu funkcja ta przyjmuje następującą postać⁵:

$$W = f[U_1(y_1), U_2(y_2), \dots, U_n(y_n)] \quad (1)$$

przy czym W = dobrobyt społeczny; U_i = użyteczność indywidualna danej osoby i oraz y_i = dochód danej osoby.

W literaturze występuje wiele definicji dobrobytu społecznego (Zob. Layard i Walters, 1978). Zgodnie ze standardową praktyką przyjmowaną dla analizy kosztów i korzyści opierać się będziemy na definicji utylitarnej (Brent, 2006; Boardman i in., 2001). Jej specyfika polega na przypisywaniu jednakowej wagi dobrobytowi każdej osoby. Dobrobyt społeczny równy jest wówczas sumie użyteczności indywidualnych⁶:

$$W = U_1(y_1) + U_2(y_2) + \dots + U_n(y_n) \quad (2)$$

Efekty interwencji publicznych możemy więc opisać następującym wzorem::

$$\Delta W = U_1(\Delta y_1) + U_2(\Delta y_2) + \dots + U_n(\Delta y_n) \quad (3)$$

Jeśli w wyniku interwencji publicznej $\Delta W > 0$, wówczas ma ona pozytywny wpływ na dobrobyt społeczny i tym samym, zgodnie z zasadą maksymalizacji użyteczności, należy wówczas rekomendować jej kontynuację.

⁴ Użyteczność odnosić się może do dobrobytu materialnego (*welfare*) lub subiektywnego poczucia dobrobytu (*well-being*). Zgodnie z założeniem o maksymalizacji użyteczności ludzie podporządkowują swoje działania maksymalizacji własnej użyteczności indywidualnej.

⁵ Ten ogólny format znany jest jako funkcja dobrobytu społeczeństwa Bergsona-Samuelsona.

⁶ Sugeruje to, że nie istnieje współzależność między różnymi ludźmi w zakresie wydajności.

Analiza kosztów i korzyści to praktyczne narzędzie pomiaru zmian dobrobytu społecznego (ΔW). Według definicji wskazanej w równaniu (2), wpływ polityki publicznej obliczamy jako bilans wszystkich indywidualnych kosztów i korzyści powstałych na skutek wdrażania danego programu.

Równanie (3) przedstawia pierwszy etap społecznej analizy, mający na celu oszacowanie efektów interwencji w kategoriach pieniężnych (tj. Δy). Wartość indywidualnych użyteczności przyjmowała będzie zarówno wartości dodatnie, jak i ujemne, które odpowiadać będą korzystnemu oraz niekorzystnemu wpływowi interwencji na poszczególne osoby. Na potrzeby analizy i programów wspierania zatrudnienia możemy uprościć wzór (2), zakładając, że społeczeństwo dzieli się na dwie główne grupy: uczestników programu (beneficjentów) oraz pozostałych obywateli, którzy jako podatnicy finansują interwencję publiczną. Ogólnie rzecz biorąc, korzyści uzyskuje pierwsza z wymienionych grup, a koszty ponoszone są przez drugą⁷:

$$W = U_1(y_1) + U_2(y_2) \quad (4)$$

Efekty interwencji wyrażane w wartościach pieniężnych

Szacując efekty interwencji w wartościach pieniężnych będziemy opierali się na nadwyżce konsumenta (*economic surplus*). Jest to lepsza miara dla analiz kosztów i korzyści niż ceny rynkowe, które zaniżałyby wpływ polityki publicznej na dobrobyt społeczny (Brent, 2006; Boardman i in., 2001). Wzrost i spadek nadwyżki można oszacować za pomocą skłonności do ponoszenia kosztów interwencji (WTP – *willingness to pay*), która przyjmuje wartości ujemne przy spadku nadwyżki. Funkcja dobrobytu społecznego w analizie kosztów i korzyści społecznych przyjmuje zatem następującą postać:

$$W = U_1(WTP_1) + U_2(WTP_2) \quad (5)$$

Różnego rodzaju efekty programów wspierania zatrudnienia mierzone są w kategoriach pieniężnych (dochody, świadczenia pieniężne, wpływy budżetowe), a więc mogą być wykorzystane jako miary skłonności do ponoszenia kosztów.⁸

Szacunki dotyczące wpływu zatrudnienia na przestępczość i zdrowie opierają się zwykle na cenach i wartościach rynkowych. Nie uwzględniają więc one skłonności społeczeństwa do ponoszenia kosztów interwencji w zamian za spadek przestępczości lub poprawę stanu zdrowia. Jest to słabość wielu analiz, której często nie da się uniknąć (Boardman i in., 2001).

W przypadku programów rynku pracy możemy jednak uzyskać oszacowanie finansowe również dla innych obszarów wpływu niż zdrowie i przestępczość. Wynika to z faktu, iż większość efektów tych programów ma z natury charakter pieniężny (np. koszty administracyjne, wysokość płac). Dzięki temu posiadamy dobre miary pozwalające na oszacowanie skłonności do ponoszenia kosztów interwencji.

Całkowity wpływ na dobrobyt społeczny

Skłonność do ponoszenia kosztów stanowi pieniężną ilustrację zmian poziomu dobrobytu społecznego związanych z konsumpcją usług dostarczanych w ramach programów publicznych. Aby oszacować faktyczną zmianę dobrobytu zgodnie z równaniem (5), potrzebne jest założenie dotyczące wartości pieniądza tj. określenie kształtu funkcji użyteczności $U_i(y_i)$.

⁷ Oczywiście wystąpią też pewne korzyści na rzecz grupy 2, takie jak niższy poziom przestępczości.

⁸ Wynika po prostu z faktu, iż gotowość do ponoszenia opłat w określonej wysokości wyrażana jest w jednostkach pieniężnych równych tej skłonności (np. wartość skłonności do ponoszenia kosztów w wysokości 1 £ również wynosi 1 £).

Ogólnie przyjętą praktyką jest pomijanie indywidualnych różnic w zakresie wartości pieniądza oraz bazowanie na założeniu, że wartość kosztu krańcowego pieniądza jest równa dla wszystkich ludzi. Funkcja społecznego dobrobytu wynosi wówczas:

$$W = WTP_1 + WTP_2 \quad (6)$$

a ogólny bilans kosztów i korzyści stanowi po prostu sumę indywidualnych WTP.

W Zielonej księdze Brytyjskiego Ministerstwa Gospodarki i Finansów (*HM Treasury Green Book*, 2003) zaleca się, aby w analizie kosztów i korzyści oraz w ewaluacji polityk publicznych wyraźnie uwzględniać specyfikę wpływu interwencji w obrębie różnych grup społecznych. Szacowana wartość pieniądza (marginalna użyteczność) uzależniona jest bowiem od wyjściowego poziomu dochodów. Ludzie z różnych grup dochodowych będą zatem w odmienny sposób postrzegać wpływ polityki publicznej na wysokość ich dochodów.

Rzeczywisty wpływ na dobrobyt społeczny można obliczyć wykorzystując miary względne lub przypisując różne wagi wpływu dla różnych grup dochodowych. Wagę dobrobytu⁹ możemy określić w oparciu o różnicę marginalnej użyteczności pieniądza (użyteczność zmiany dochodu na jedną jednostkę pieniężną). Taki sposób wyliczania zmian dobrobytu społecznego nazywać będziemy ujęciem pieniężnym, względnym. Należy zwrócić uwagę, że mimo wyższych wag przypisanych zmianom dochodowym w grupach o niższych zarobkach wyjściowych, analiza kosztów i korzyści nie odzwierciedla żadnych ocen normatywnych dotyczących dystrybucji dochodów w społeczeństwie (stąd jesteśmy bardziej skłonni posługiwać się terminem „waga dobrobytu” zamiast „waga dystrybucyjna”)¹⁰.

W analizie kosztów i korzyści, ważąc rezultaty pieniężne polityki publicznej (zgodnie ze zmienną skłonnością do ponoszenia kosztów) oraz bazując na różnicach względnej użyteczności marginalnej dochodów, otrzymujemy precyzyjne oszacowanie funkcji społecznego dobrobytu (patrz wzór 5). Wpływ programów wspierających zatrudnienie mierzymy według poniższego wzoru (przy czym α_1 to waga dobrobytu dla grupy I):

$$\Delta W = \alpha_1 (WTP_1) + \alpha_2 (WTP_2) \quad (7)^{11}$$

Ustalone w ten sposób ramy analizy kosztów i korzyści zapewnią dobre oszacowanie ogólnego wpływu interwencji publicznej na dobrobyt społeczny¹².

1.1. Obszary wpływu w społecznej analizie kosztów i korzyści

W 2007 r. Departament Pracy i Emerytur (*The Department for Work and Pensions* – DWP) zlecił Instytutowi Badań Politycznych (*Policy Studies Institute* – PSI) przeprowadzenie przeglądu metodologii analizy kosztów i korzyści dla programów w zakresie zatrudnienia. Po przeprowadzeniu przeglądu wskazano kluczowe obszary wpływu, które powinny być uwzględniane w analizach: wpływ na rynek podstawowy i wpływ na rynek wtórny.

⁹ Wymiennie stosuje się określenie „waga dystrybucyjna”.

¹⁰ Więcej informacji można znaleźć w dokumencie na temat wag w analizie CBA.

¹¹ Jedną z wag będzie równa 1. Zazwyczaj jest to waga stosowana w przypadku grupy o wyższych dochodach (α_2).

¹² Należy zwrócić uwagę, że forma SWF w równaniu (7) nie oznacza, że odeszliśmy od założeń użyteczności funkcji SWF. Równanie (7) wskazuje na fakt, iż dochody w dobrobycie społeczeństwa ukazują krzywą wklęsłą, nie zmienia to jednak założenia, że dobrobyt społeczeństwa w odniesieniu do wydajności ukazany jest jako linia prosta i tym samym jest użyteczny. Dokładniejsze informacje na ten temat w: Cowell i Gardiner (1999).

Rynek podstawowy to rynek objęty bezpośrednią interwencją publiczną. W przypadku programów dotyczących zatrudnienia jest to rynek pracy. Wpływ na tym rynku tworzą następujące efekty:

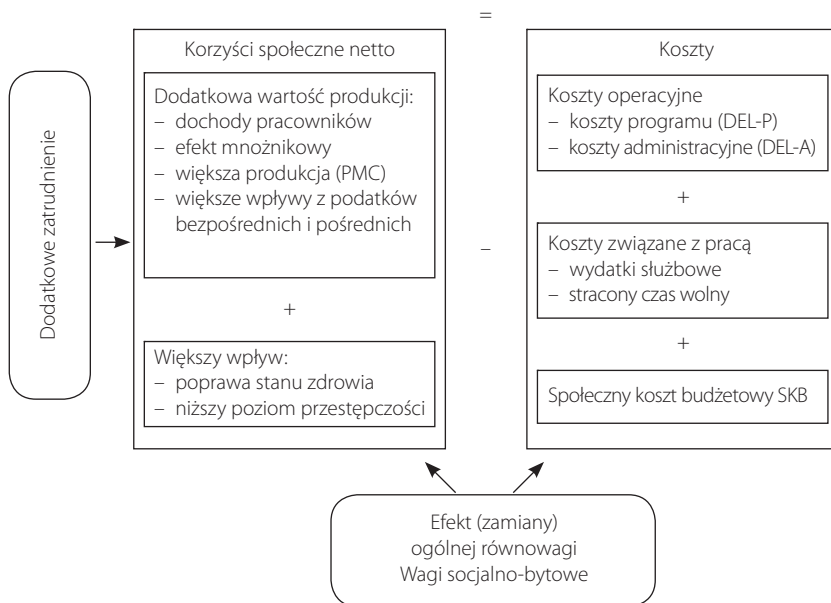
- zmiany dochodów¹³,
- koszty pracownicze (koszty podjęcia zatrudnienia),
- koszty utraty czasu wolnego,
- efekty równowagi na rynku pracy oraz
- społeczny koszt budżetowy (SKB – koszty związane z opodatkowaniem niezbędnym dla sfinansowania interwencji)¹⁴.

Rynek wtórny obejmuje pozostałe rynki, w obrębie których interwencje na rynku pracy mogłyby wywołać efekt domina. Zachodzić mogą tutaj efekty następujących rodzajów:

- efekt mnożnikowy na gospodarce,
- zmiany wielkości produkcji, tj. korektor rynku produktów [*Product Market Corrector – PMC*] oraz
- zmiany wskaźników stanu zdrowia i poziomu przestępczości.

W społecznej analizie kosztów i korzyści wpływ na rynki wtórne zaliczany jest do efektów podatkowych, a następnie podlega wazeniu, aby otrzymać ogólne oszacowanie efektów netto oraz ogólną ocenę efektywności interwencji. Sposób obliczania ogólnych korzyści netto przedstawia poniższy schemat.

Ryc. 1. Ramy analizy kosztów i korzyści społecznych



¹³ Zmianę dochodów zachodzącą w miarę, jak ludzie odchodzą od korzyści i zmierzają w kierunku pracy szacuje się przy użyciu wyników ewaluacji i można ją znaleźć w modelu Taxben.

¹⁴ Społecznych kosztów budżetowych – SKB [ang. *Social Cost of Exchequer Finance (SOCEF)*] nie uwzględniono w rekomendowanej analizie z Zielonej księgi, ale uwzględniono je jako część analizy wrażliwości w niniejszym dokumencie po konsultacjach z Brytyjskim Ministerstwem Gospodarki i Finansów.

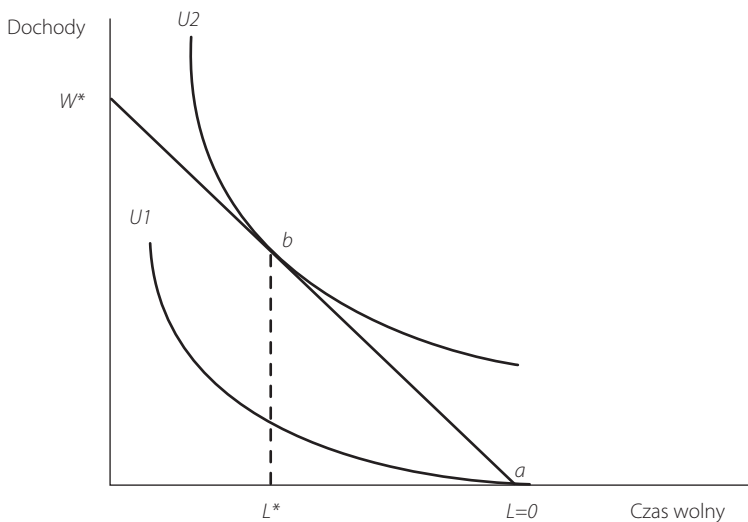
Szacowanie wartości transferów

Analizując wpływ interwencji na rynki wtórne należy rozstrzygnąć, które efekty związane są bezpośrednio ze zmianami dobrobytu społecznego oraz które stanowią jedynie formę transferów kosztów lub korzyści. W tym miejscu musimy powrócić do zagadnienia nadwyżki konsumenta. Włączenie nadwyżki konsumenta w ramy analizy kosztów i korzyści wymaga bowiem przyjęcia określonych założeń. Kluczowe założenie, jakie przyjmujemy na potrzeby analizy programów zatrudnieniowych mówi, że mamy do czynienia z rynkiem (rynkami) odbiegającym od modelu idealnego, na którym obserwujemy przypadki zawodności rynkowych (*market failures*). Założenie to opiera się na tezie, że występowanie zawodności rynkowych jest koniecznym warunkiem dla podejmowania interwencji publicznej. Należy przy tym zwrócić uwagę, że nie przyjmujemy żadnych założeń dotyczących rynków wtórnych.

Zawodności rynku sprawiają, że niektórzy bezrobotni nie podejmują zatrudnienia, pomimo iż byliby skłonni pracować za aktualne stawki rynkowe. Na przeszkodzie staje nierówny dostęp do informacji, brak dostosowania umiejętności do potrzeb rynku pracy oraz niekorzystna pozycja negocjacyjna w stosunku do potencjalnych pracodawców (*job search friction*). Dlatego też zakładamy, że bezrobocie może mieć charakter niedobrowolny (patrz Ryc. 2) Założenie to zgadza się z ustaleniami z prac w zakresie relacji pomiędzy subiektywnym poczuciem dobrostanu a bezrobociem (np. Clark i Oswald, 1994, 2002; Winkelmann i Winkelmann, 1998).

Biorąc pod uwagę płace (W^*) osoba bezrobotna wykonywałaby pracę w ilości równiej L^* , ale z powodu zawodności rynku pozostaje bez pracy czyli $L = 0$ (punkt a na Ryc. 2). Poprzez podjęcie zatrudnienia bezrobotny zwiększa podaż pracy oraz własne zasoby finansowe (punkt b na Ryc. 2). Uzyskuje tym samym nadwyżkę konsumenta, którą ilustruje wzrost użyteczności z wartości $U1$ do $U2$.

Ryc. 2. Bezrobocie o charakterze niedobrowolnym¹⁵



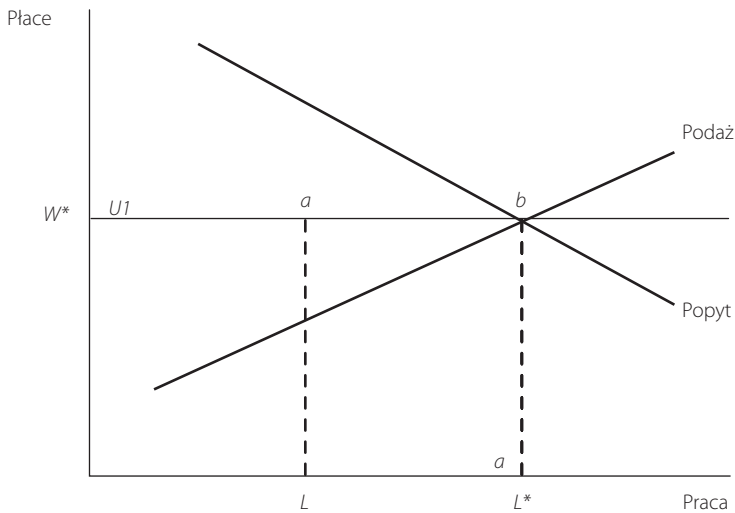
¹⁵ Bez straty dla ogólności zakładamy, że nie istnieją świadczenia pracownicze, co oznacza, że możemy założyć liniowy pułap ograniczenia budżetu.

Rycina 3 ilustruje zagregowane efekty zawodności występujące na danym rynku pracy. Sprowadzają one podaż pracy do poziomu L , a więc poniżej punktu równowagi (L^*), który byłby osiągnięty w warunkach rynku idealnego. Osoby wykonujące pracę w ilości ograniczonej przedziałem $L^* - L$ mogą uzyskać nadwyżkę konsumenta związaną z podjęciem zatrudnienia. W ich przypadku marginalne korzyści z pracy (płaca W^*) są bowiem większe niż marginalne koszty. Jeśli program wspierania zatrudnienia zdoła zapewnić pracę jednej dodatkowej osobie, możemy przyjąć założenie, że prowadzi to do wzrostu nadwyżki konsumenta, którą zapiszemy w naszej analizie po stronie korzyści.

Do korzyści (*społeczne korzyści budżetowe*) należy zaliczyć również zmniejszenie świadczeń pieniężnych oraz wzrost wpływów z podatków związane z podejmowaniem zatrudnienia.

Z Ryc. 3 wynika, że wzrost nadwyżki dotyczy również pracodawców. Koszt zatrudnienia dodatkowej osoby jest bowiem mniejszy od wynikających z tego korzyści (różnica między krzywą popytu na pracę a stawką płacy W^*). Ponieważ nie zakładamy występowania równowagi rynkowej, możemy przyjąć, że skłanianie ludzi do podjęcia pracy nie wpływa negatywnie na poziom płac¹⁶. W tym sensie nie musimy uwzględniać ani oceniać wpływu na osoby pracujące przed realizacją interwencji¹⁷. Gdybyśmy analizowali rynek idealny, wówczas wzrost nadwyżki wśród osób zatrudnionych w efekcie realizacji programu, współwystępowałby z obniżeniem nadwyżki niektórych pracowników. Negatywny wpływ w drugiej grupie związany byłby z obniżeniem wartości dochodów.

Ryc. 3. Zawodność rynku pracy



¹⁶ Może to nie mieć miejsca, jeśli w programie zatrudnienia będzie bardzo dużo dodatkowych osób do pracy w stosunku do aktualnego poziomu siły roboczej, lecz jest to mało prawdopodobne. Istnieje małe prawdopodobieństwo, aby realizacja programów prozatrudnieniowych powodowała presję na obniżenie płac.

¹⁷ Dotyczy to dochodów i efektu substytucji zmiany płac.

Drugie założenie, jakie przyjmować będziemy na potrzeby analizy programów wspierania zatrudnienia polega na wykluczeniu społecznie nieakceptowalnych preferencji¹⁸. Ma to duże znaczenie w przypadku mierzenia wpływu na poziom przestępczości i koszty z tym związane. Dlatego też, po stronie korzyści nie będziemy uwzględniali na przykład profitów, jakie przestępcy czerpią z kradzieży.

Powyższe założenia pozwalają określić obszary zachodzenia transferów. Zasadniczo mamy do czynienia z dwoma ich rodzajami. Po pierwsze, niektóre koszty mogą stanowić korzyści dla innych podmiotów lub osób. Po drugie, niektóre korzyści przypadające pewnej grupie mogą spowodować efekt domina w stosunku do innych osób lub grup. W przypadkach, gdy mamy do czynienia z prostym transferem korzyści, nie powinny być one uwzględniane w analizie.

W przypadku programów wspierania zatrudnienia transfery pierwszego rodzaju dotyczą kosztów podjęcia pracy, kosztów utraty świadczeń społecznych oraz kosztów spowodowanych przestępczością. Wykazaliśmy, że wpływ na rynek pierwotny (rynek pracy) – w postaci wzrostu zatrudnienia – wyrażany jest we wzroście nadwyżki konsumenta. Wszelkie koszty ponoszone w celu podjęcia pracy (np. koszty dojazdów lub koszty opieki nad dziećmi) stanowią wówczas koszty utraconych korzyści¹⁹.

Kosztów tych nie uznajemy za transfery, które realizowane są w innych miejscach (np. w formie przychodów na rzecz dostawców opieki nad dziećmi lub dostawców usług transportowych), dlatego też nie są zliczane po stronie korzyści. Wynika to z faktu, iż nie dysponujemy wiedzą na temat struktury rynków, na których transfery byłyby realizowane. Nie możemy więc przyjmować żadnych założeń dotyczących tych rynków. Możemy jedynie stwierdzić, że w przypadku, gdyby były to rynki idealnie konkurencyjne, wówczas nie występowałaby na nich nadwyżka związana z interwencją na rynku pracy.

Mimo iż interesujące nas rynki wtórne (ryniki usług transportowych lub opieki nad dziećmi) mogą nie spełniać powyższego warunku brakuje nam podstaw merytorycznych do określenia kwoty potencjalnego wzrostu nadwyżki występującej na tych rynkach. Przyjmujemy więc, że całkowity efekt netto związany z kosztami podjęcia zatrudnienia będzie miał negatywny wpływ na poziom społecznego dobrobytu.

Drugi rodzaj możliwego wpływu programów prozatrudnieniowych na rynki wtórne związany jest z występowaniem efektów mnożnikowych, związanych ze wzrostem wydatków konsumpcyjnych osób podejmujących pracę. Jak wspomnieliśmy brakuje nam wiedzy dotyczącej struktury rynków wtórnych (w tym wypadku rynków, na których realizują swoje wydatki beneficjenci programu). Utrudnia to oddzielenie efektów, które faktycznie powodują wzrost nadwyżki, od prostych transferów korzyści, które w analizie nie powinny być uwzględniane. Dlatego też, aby uniknąć podwójnego naliczania korzyści, w naszej praktyce zdecydowaliśmy się nie uwzględniać efektu mnożnikowego.

Podsumowując, powyżej opisaliśmy zasady tworzenia odpowiednich miar na potrzeby analizy kosztów i korzyści społecznych oraz sposoby szacowania ich wartości. Jesteśmy w stanie określić skłonność do ponoszenia kosztów interwencji w przypadku większości potencjalnych efektów oraz możemy przy pomocy ważenia użyteczności indywidualnej uzyskać dobre oszacowanie ogólne wpływu programów wspierania zatrudnienia na poziom dobrobytu społecznego. Określiłiśmy również rodzaje wpływu charakterystyczne dla rynku pierwotnego oraz rynków wtórnych. Zaproponowaliśmy także przyjęcie założeń pozwalających określić, jakie rodzaje wpływu należy brać pod uwagę, a jakie wykluczać, aby uniknąć błędów podwójnego zliczania.

¹⁸ Jest to standardowa procedura w ekonomii i analizie CBA (Boardman i in., 2001).

¹⁹ Wynika to z faktu, że ludzie nie podejmują tych decyzji konsumpcyjnych (o opiece nad dziećmi i podróżami), aby bezpośrednio zmaksymalizować wydajność i tym samym uwzględniają one straty nadwyżki i powinny być odliczone od wzrostu płac.

2. Wpływ na rynku pierwotnym

2.1. Koszty podjęcia pracy

W momencie, gdy jakaś osoba podejmuje pracę, ponosi związane z tym koszty. Koszty te obejmować mogą między innymi dojazd do pracy i do domu oraz koszty opieki nad dziećmi. Są to koszty utraconych korzyści, które osoba podejmująca pracę musi pokryć z części uzyskanych dochodów rezygnując tym samym z innych wydatków, które mogłyby sfinansować z tej części dochodu. Prowadzi to do zmniejszenia nadwyżki konsumenta. Nadanie miar pieniężnych kosztom podjęcia pracy pozwoliłoby na bardziej precyzyjnie szacowanie indywidualnych korzyści netto związanych z zatrudnieniem.

Programy na rzecz zatrudnienia prowadzone przez Departament Pracy i Świadczeń (DWP) zaprojektowane są jako programy mające na celu pomoc przy przejściu ze stanu bezrobocia lub braku aktywności zawodowej do zatrudnienia. Uczestnicy tych programów, którzy podejmą pracę, poniosą wskazane powyżej koszty.

Instytut Badań Politycznych (PSI) (2007) sugeruje, że koszty te powinno się uwzględniać w analizie kosztów i korzyści wykorzystując „wyniki badań na temat pieniędzy wydawanych przez pracowników o niskich pensjach na dojazd do pracy i do domu po pracy”. Autorzy sugerują, aby do czasu uzyskania tych wyników „przynajmniej wspomnieć, że te koszty istnieją i mogą mieć jakieś znaczenie”.

Metodologia i zastrzeżenia

Wykorzystując dane z Departamentu Edukacji (*Department for Education – DfE*)²⁰ na temat korzystania z formalnej opieki dla dzieci w różnym wieku oraz dla różnych rodzajów rodzin, wyliczyliśmy tygodniowe koszty opieki nad dziećmi dla samotnych rodziców czynnych zawodowo oraz par z dziećmi. Biorąc pod uwagę odmienny zakres opieki nad dziećmi w okresach wakacyjnych oraz mnożąc uzyskane wartości przez liczbę dodatkowych lat pracy, otrzymujemy całkowity koszt opieki nad dziećmi związany z podjęciem pracy.

W naszej praktyce skorzystaliśmy z danych *Labour Force Survey* (LFS) i uzyskaliśmy informacje dotyczące form transportu do i z miejsca pracy wśród pracowników o niskich dochodach (dolny kwintyl według wielkości zarobków). Dane te wykorzystaliśmy do oszacowania kosztów użytkowania samochodu, opłat za korzystanie z metra i dróg krajowych oraz kosztów zakupu biletów autobusowych. Mnożąc uzyskane szacunki przez dodatkowe lata pracy dodane przez udział w programie wspierania zatrudnienia uzyskaliśmy oszacowanie całkowitych rocznych kosztów dojazdu do pracy.

Przyjmując powyższe podejście, warto wskazać na kilka zastrzeżeń:

- 1) Departament Zabezpieczenia Społecznego (*Department of Social Security*) (1999) sugeruje, że koszty związane z zakupem narzędzi i wyposażenia oraz ubrań roboczych niezbędnych dla danego stanowiska pracy, jak również koszty stałe, takie jak zakup żywności i koszty uczestniczenia w życiu towarzyskim załogi również mogą być analizowane jako koszty podjęcia zatrudnienia. Niemniej jednak dane dotyczące kosztów zakupu sprzętu i ubioru są skromne, natomiast pozostałe koszty (żywność i życie towarzyskie) nie stanowią kosztów dodatkowych niezbędnych do podjęcia pracy.

²⁰ Sprawozdanie wykorzystane na potrzeby oszacowania kosztów pracowniczych opracowano w 2009 r. i tym samym nosi poprzednią nazwę DfE – Departamentu ds. Dzieci, Szkół i Rodzin [*Department for Children, Schools and Families* (DCSF)].

2) Zagadnienia związane z opieką nad dziećmi:

- Dostępne w naszych warunkach dane na temat korzystania z opieki nad dziećmi nie uwzględniają statusu rodziców na rynku pracy. Mogłoby to prowadzić do zaniżania kosztów opieki ponoszonych przez rodziców (pary) podejmujących zatrudnienie.
- Zakładamy, że w przypadku korzystania z nieformalnej opieki nad dziećmi koszty są zerowe. Z tego powodu nie uwzględniamy kosztów utraconych korzyści w przypadku czasu poświęcanego na opiekę. Omawiane dane i założenia dotyczą Anglii jako całości. Nie uwzględniają one różnic regionalnych. Dane na temat opłat są dostępne jedynie w przypadku wybranych typów opieki nad dziećmi. Przyjmujemy założenie, że dane te mogą być uogólnione na wszystkie formy opieki. Dane pochodzące z programów Departamentu Zatrudnienia i Świadczeń (*Department of Work and Pension*), dotyczą głównie kosztów opieki nad dziećmi w wieku do lat 11. Zakłada się, że koszty opieki nad dziećmi starszymi są nieistotne. Istnieją bowiem przesłanki empiryczne wskazujące, że koszty opieki spadają drastycznie w momencie rozpoczęcia przez dziecko edukacji na poziomie szkoły średniej (przyj. red. w angielskim systemie oświaty edukację w szkole średniej rozpoczynają dzieci w wieku 11 lat). Nieuwzględnienie w analizie dzieci w wieku powyżej 11 lat może spowodować drobne odchylenie w kierunku niedoszacowania kosztów opieki.
- Zakłada się, że liczba członków rodziny nie ma wpływu na decyzję o korzystaniu z formalnej opieki nad dziećmi (tj. rodziny z większą liczbą dzieci nie wydają więcej na opiekę nad dziećmi). Dane DCSF (*Department for Children, Schools and Families*) (2009) sugerują, że dzieci w wieku przedszkolnym mające dwoje lub więcej rodzeństwa rzadziej objęte są formalną opieką; nie istnieje taki związek w przypadku dzieci w wieku szkolnym. Założenie to może spowodować drobne odchylenie analizy w kierunku przeszacowania kosztów opieki.
- Trzy- i czterolatkom przysługuje 12,5 godziny bezpłatnej opieki przedszkolnej tygodniowo przez okres 38 tygodni rocznie. Według danych DCSF (2009) z opieki tej korzysta trzy czwarte uprawnionych dzieci. Dzięki temu spadają koszty ponoszone przez rodziców. Z drugiej strony jednak dokonywane są płatności transferowe ze strony państwa na rzecz dostawców opieki. Transfery tego rodzaju powinny być uwzględniane w statystykach podatkowych, tak, abyśmy mogli w pełni określić koszty opieki nad dziećmi. Podejmowane są obecnie działania w tym kierunku. Przy aktualnym stanie danych, koszty te możemy przeszacowywać.

3) Zagadnienia dotyczące dojazdów:

- W kosztach dojazdów nie uwzględniamy różnic regionalnych. Oznacza to, że prawdopodobnie przeszacowujemy koszty podróży w przypadku programów, wśród uczestników, których liczniej reprezentowane są osoby zamieszkujące obszary, gdzie koszty komunikacji są niższe (np. w miastach).
- Przy wyliczaniu liczby dni w pracy założono, że wszyscy beneficjenci podejmując pracę wykonują ją przez pięć dni w tygodniu. Prowadzi to do zawyżonego wyliczenia kosztów dojazdów. Założenie to będzie skutkowało różnymi obciążeniami w przypadku różnych programów. Na przykład niektóre programy ukierunkowane są na samotnych rodziców. Prawdopodobieństwo, że beneficjenci tych programów podejmą zatrudnienie w pełnym wymiarze jest stosunkowo niskie. Programy tego rodzaju będą więc szczególnie obciążone błędem przeszacowania kosztów dojazdów do pracy.
- W przypadku niektórych grup, na przykład osób niepełnosprawnych, sposób dojazdu do pracy, a tym samym jego koszty będą odmienne od standardowych. Założenie, jakie przyjęliśmy na potrzeby naszych analiz pomija tę specyfikę i opiera się na kosztach dojazdów ponoszonych przez „przeciętnego” uczestnika wszystkich programów wspierania zatrudnienia.
- Kolejne założenie, jakie przyjęliśmy, dotyczy dystansu pokonywanego w drodze do pracy. Założyliśmy, że osoby podejmujące zatrudnienie w efekcie uczestnictwa w programach Departamentu

Pracy i Świadczeń pokonują dziennie odległość około 12 mil. Koszty związane z przejazdami na tej odległości traktujemy w całości jako koszt podjęcia zatrudnienia. Przyjmujemy, że wydatki na ten cel nie byłyby ponoszone, gdyby beneficjenci nie pracowali. Istnieje prawdopodobieństwo, że w tym przypadku wystąpi odchylenie w kierunku zawyżenia kosztów. Osoby dojeżdżające do i z pracy, starają się bowiem połączyć tę obowiązkową podróż z innymi, niezwiązanymi z pracą sprawami (np. z zakupami), a więc przynajmniej część kosztów byłaby przez nich ponoszona również w sytuacji pozostawania bez zatrudnienia.

Z uwagi na wspomniane wyżej zastrzeżenia, zaleca się uwzględnianie w analizach kosztów i korzyści wartości pieniężnych kosztów podjęcia zatrudnienia²¹.

Koszty opieki nad dziećmi oraz koszty dojazdów do pracy należy uwzględniać w analizach dotyczących programów na rzecz zatrudnia, biorąc pod uwagę ograniczenia tych szacunków.

Najlepszym sposobem szacowania rocznych kosztów opieki nad dziećmi generowanych przez programy prozatrudnieniowe Departamentu Pracy i Świadczeń jest mnożenie liczby dodatkowych lat pracy (ustalonej odrębnie dla samotnych rodziców i par) przez średnie roczne wydatki na opiekę formalną nad dziećmi. Przyjąć należy następujące wzory obliczeń:

liczba dodatkowych lat pracy samotnych rodziców x 744 GBP;

liczba dodatkowych lat pracy par z dziećmi x 694 GBP.

W kosztach należy uwzględnić inflację. Najlepszym sposobem oszacowania rocznych kosztów dojazdów do pracy jest mnożenie liczby dodatkowych lat pracy przez średnie roczne koszty podróży. Należy zastosować następujące wzory obliczeń:

dotatkowe lata pracy x 433 GBP.

Przy obliczaniu całkowitego wpływu interwencji, sumę kosztów podjęcia zatrudnienia odejmujemy od wartości zsumowanych korzyści.

Podsumowując, koszty podjęcia zatrudnienia rozumiemy jako dodatkowe, nieuniknione wydatki związane z rozpoczęciem wykonywania pracy przez osoby uprzednio bezrobotne. Mają one wpływ na bilans kosztów i korzyści zatrudnienia. Zgodnie z metodologią przedstawioną w niniejszym artykule, możemy uzyskać pieniądze miary kosztów podjęcia pracy, a następnie wykorzystać je na potrzeby analizy kosztów i korzyści programów wspierania zatrudnienia.

²¹ Przykładem zastosowania analizy są dane na temat programu New Deal 25+ (ND25+) sugerujące, że dzięki jego wdrożeniu powstaje 10 324 dodatkowych miejsc utrzymujących się średnio przez rok. Korzystając z tych danych oraz z danych dotyczących profilu rodzin objętych programem, obliczyliśmy wysokość kosztów podjęcia zatrudnienia.

Koszty opieki nad dziećmi oszacowane zostały w oparciu o dane empiryczne, które pozwoliły wyznaczyć roczne koszty opieki formalnej. Wartość ta została pomnożona przez liczbę „dodatkowych lat pracy” określoną dla samotnych rodziców i par z dziećmi. Uwzględniono także intensywność korzystania z opieki formalnej specyficzną dla tych grup. Całkowity koszt opieki oszacowaliśmy na poziomie około 1,67 mln GBP.

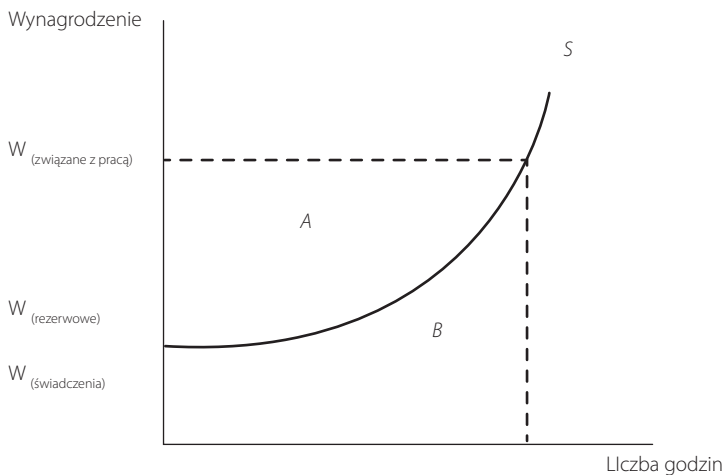
Aby uwzględnić koszty dojazdów do pracy pomnożyliśmy „dodatkowe lata pracy” przez łączne koszty rocznego korzystania z samochodu, kolei (*National Rail*), metra (londyńskiego) i przejazdów autobusami. Dało to wartość około 4,47 mln GBP. Całkowite koszty podjęcia zatrudnienia związane z programem ND25+ oszacowaliśmy więc na 6,14 mln funtów (4,47 mln GBP + 1,67 mln GBP). Wartość tę należy odjąć od zsumowanej wartości korzyści programu. Uzyskamy wówczas oszacowanie rzeczywistego wpływu interwencji na dobrobyt społeczny (wpływ netto).

2.2. Czas wolny w analizie kosztów i korzyści

Podjęcie zatrudnienia przez osobę uprzednio bezrobotną wiąże się z koniecznością rezygnacji z części czasu wolnego. Strata ta uznawana jest za koszt utraconych korzyści. Właściwą metodą pomiaru skutków programu rządowego dla dowolnej grupy osób jest zastosowanie wskaźnika zmiany netto w ich nadwyżce (lub rencie ekonomicznej) zamiast zmiany netto w dochodzie (Greenberg i Robins, 2008)²². W klasycznych analizach kosztów i korzyści oblicza się z reguły całkowitą zmianę dochodów netto. Istnieją jednak argumenty za tym, że takie podejście może skutkować przecenieniem indywidualnych korzyści z programów na rzecz zatrudnienia.

Obszar A + B na Ryc. 4 pokazuje wzrost dochodu netto osób rezygnujących ze świadczeń i podejmujących zatrudnienie w wymiarze h^* godziny przy wynagrodzeniu o wartości W (związane z pracą). Obszar B jest ilustracją utraty czasu wolnego. Zatem faktyczna nadwyżka netto jest równoważnością wzrostu dochodu pomniejszoną o wartość utraconego czasu wolnego. Na Ryc. 4 wartość tę przedstawia Obszar A – obszar powyżej krzywej podaży²³.

Ryc. 4. Zmiany w wartości nadwyżki netto i czasu wolnego wynikające z podjęcia zatrudnienia



Metodologia i zastrzeżenia

W proponowanej metodologii (Greenberg & Robins, 2008) wykorzystuje się dane dotyczące wartości netto dochodów uzyskiwanych z pracy i poza pracą oraz wzrostu liczby przepracowanych godzin wynikającego z podjęcia zatrudnienia. Metodologia ta zakłada, że pomiędzy wartością dochodów z pracy i uzyskiwanych poza pracą znajduje się wartość graniczna, będąca najniższym akceptowalnym poziomem wynagrodzenia. Metodologia ta wyznacza również indywidualne krzywe podaży pracy dla poszczególnych beneficjentów. W oparciu o powyższe założenia przeprowadzić można obliczenia rzeczywistej zmiany wartości dochodów związanej z podjęciem pracy. Wartość tę ilustruje Obszar A na Ryc. 4.

²² Greenberg, D. i Robins, P. (2008).

²³ Greenberg i Robins (2008) dokonali oceny wielu programów na rzecz zatrudnienia realizowanych w Kanadzie. Zgodnie z ich szacunkami, ponad jedna trzecia zysków związanych ze wzrostem dochodów kompensowana jest utratą czasu wolnego. Jest to zgodne z wynikami wcześniejszych badań, co z kolei sugeruje, że korzyści dochodowe powinniśmy redukować o co najmniej jedną czwartą. (np. Bell i Orr, 1994 ; Greenberg, 1997).

Z tym podejściem wiążą się dwa podstawowe problemy:

- Pierwszy z nich to niepewny charakter zarówno szacunków dotyczących granicznej wartości wynagrodzenia, jak i krzywej wartości podaży. Należy zauważyć, że są to elementy silnie warunkujące wartość szacowania (Greenberg i Robins, 2008).
- Po drugie w literaturze przedmiotu zakłada się jedynie dodatnią wartość ekonomiczną czasu wolnego. Niemniej jednak wartość czasu spędzonego w pracy może również osiągnąć wartość dodatnią i jeśli tak się dzieje, równoważy to utratę czasu przeznaczanego na wypoczynek. Ludzie często cenią swój czas spędzany w pracy w związku z szeregiem potencjalnych korzyści zdrowotnych i bytowych. Korzyści te wynikają prawdopodobnie z ukrytych funkcji pracy – kwestię tę po raz pierwszy poruszył Jahoda (1982). Zgodnie z teorią ukrytej deprywacji zatrudnienie zapewnia dostęp do pięciu kategorii doświadczeń, które są ważne z punktu widzenia dobrostanu psychicznego. Do kategorii tych zaliczamy: ustrukturyzowanie czasu, aktywność, kontakty społeczne, wspólny cel i status. Są to korzyści, które ze względu na nierówny dostęp do informacji, w ograniczonym zakresie wpływają na wartość graniczną wynagrodzenia. Gdybyśmy mieli pełną wiedzę o korzyściach pozafinansowych wówczas wartość ta byłaby niższa, co skutkowałoby obniżeniem indywidualnych krzywych podaży pracy. Ryc. 4 pokazuje, że w praktyce wpływ ten nie jest tak wyraźny. Niedawno wydane publikacje dotyczące ekonomii i psychologii, w których wykorzystano wskaźnik subiektywnego dobrego samopoczucia dowodzą, że korzyści niefinansowe zatrudnienia są duże i mogą być znacznie większe niż wynagrodzenie (Clark i Oswald, 1994, 2002; Winkelmann i Winkelmann, 1998).

Najlepiej byłoby uwzględnić wszystkie skutki dla dobrego samopoczucia w innym miejscu analizy, na przykład przy obliczaniu wskaźnika skutków zdrowotnych zatrudnienia. Ze względu na brak odpowiednich danych i informacji na temat dobrobytu i stanu zdrowia psychicznego, korzyści te nie są w pełni uwzględnione w naszych szacunkach. Dane te najczęściej dotyczą oszczędności związanych z kosztami ponoszonymi przez NHS (*National Health Service* – brytyjski odpowiednik NFZ – przyp. red.).

Dodatnia wartość czasu pracy powinna być zatem uwzględniona w tym miejscu analizy. Nie dysponujemy jednak solidnymi szacunkami dotyczącymi wartości czasu pracy, ale najlepsze dostępne dane sugerują, że mogą one być wyższe niż wartość czasu wolnego. Uwzględnienie zatem wyłącznie negatywnych aspektów ograniczenia ilości czasu wolnego spowodowałoby zaniżenie wartości korzyści związanych z podejmowaniem zatrudnienia.

Szacunki dotyczące kosztów związanych z czasem poświęcanym pracy są niedokładne i jest mało prawdopodobne, aby odpowiednio uwzględniały jego pozytywną wartość. Biorąc pod uwagę wspomniane wątpliwości i problemy wynikające z braku odpowiedniej literatury, zalecane jest, by szacunki te nie były uwzględniane w analizie kosztów i korzyści.

2.3. Efekt równowagi na rynku pracy w analizie kosztów i korzyści

Programy na rzecz zatrudnienia mogą generować efekty na poziomie całej gospodarki. PSI dokonał obszernego przeglądu literatury pod kątem efektu równowagi i wskazał jego główne rodzaje:

- **Efekt substytucji po stronie podaży.** Poprzez podniesienie umiejętności zawodowych i wzrost liczby osób poszukujących pracy oraz zwiększenie intensywności tych poszukiwań nasilić się może konkurencja o dostępne miejsca zatrudnienia. A zatem beneficjenci programu mogą ostatecznie podjąć zatrudnienie na stanowiskach, które w przeciwnym wypadku zajmowane byłyby przez osoby nieuczestniczące w programie (Johnson, 1972; Schiller, 1973). Jeśli zastępowani pracownicy stają się bezrobotni lub podejmują gorzej płatną pracę, wówczas korzyści badanego programu dla całego społeczeństwa jako całości będą mniejsze niż korzyści odnoszone przez beneficjentów.

Skala tego rodzaju wpływu zależy od liczby dostępnych miejsc pracy. Jeśli stosunek wolnych miejsc pracy do liczby osób poszukujących pracy jest wysoki, efekt substytucji będzie niski; jeśli jednak proporcje te będą odwrotne, wpływ programu może być znaczny. Efekt substytucji będzie zatem większy w okresie złej koniunktury. Należy jednak zauważyć, że efekt substytucji ma znaczenie zwykle wyłącznie w krótkiej perspektywie. Wraz z rozwojem gospodarki powinien on słabnąć²⁴.

- **Efekt substytucji po stronie popytu.** Efekt tego rodzaju mogą wywołać programy, które obejmują subsydiowane zatrudnienie. W ich efekcie pracę mogą uzyskać beneficjenci programu kosztem innych kandydatów, których zatrudnienie byłoby dla pracodawcy bardziej kosztowne.
- **Efekt przeniesienia.** Efekt przeniesienia następuje, gdy w ramach programu na rzecz zatrudnienia dofinansowywane są niektóre firmy, które dzięki temu mogą się rozwijać szybciej niż konkurencja.
- **Efekt równowagi wynagrodzeń.** Programy na rzecz zatrudnienia, które mają duży wpływ na podaż pracy mogą obniżyć poziom równowagi wynagrodzeń. Wystąpić mogą dwa rodzaje skutków: 1) obniżone wynagrodzenia zwiększają popyt na pracę, a tym samym wzrasta poziom zatrudnienia, 2) obniżeniu ulegają zarobki pracowników nieobjętych programem.

Efekt równowagi na rynku pracy

Efekty substytucji oszacowano wykorzystując trzy podejścia badawcze:

- 1) badania sondażowe, w których pytano pracodawców, czy praca wykonywana przez uczestników programu byłaby wykonywana w sytuacji, gdyby program nie byłby realizowany
- 2) ilościowe analizy makroekonomiczne mające na celu oszacowanie związków między zatrudnieniem a różnymi miarami skali programu, oraz
- 3) modele równowagi ogólnej.

Każda z tych metod ma znaczące wady (zob. Greenberg i In., 2010)²⁵.

Pomimo rozbieżności w badaniach pewne obserwacje się powtarzają. Programy o charakterze popytowym (np. dotacje dla pracodawców) skutkują zazwyczaj większym efektem substytucji niż programy podażowe. Greenberg i in. (2010) sugeruje na podstawie istniejących badań, aby w ramach analizy kosztów i korzyści stosować testy wrażliwości oparte na założeniu o występowaniu dużego 45-procentowego efektu substytucji²⁶. Jest to współczynnik dodatkowego wpływu programu na zatrudnienie. Dla programów o charakterze podażowym (np. programów, które służą ułatwieniu poszukiwania pracy lub zawierają system zachęt do podejmowania pracy) Greenberg i in. (2010) sugeruje analizę wrażliwości, opartą na założeniu o 20-procentowym efekcie substytucji.

Efekty substytucji zmniejszają szacowaną liczbę dodatkowych miejsc pracy powstałych dzięki interwencji publicznej. Osoby, które zostały wyparte ze swoich miejsc pracy przez beneficjentów programu ponoszą koszty, przy założeniu, że wykonując dotychczasową pracę maksymalizowały swoją indywidualną użyteczność. Koszty warunkują wpływ interwencji na dobrobyt społeczny i dlatego też powinny być uwzględniane w analizie kosztów i korzyści²⁷. Nie znaleziono oddzielnych badań empirycznych dotyczą-

²⁴ Liczba wolnych miejsc pracy może wzrosnąć jeszcze przed nastąpieniem ożywienia gospodarczego, gdy pracodawcy zatrudniają nowe osoby spodziewając się wzrostu popytu. Spowoduje to jeszcze szybsze zmniejszenie efektu substytucji.

²⁵ Greenberg i in. (2010) podsumowuje dane uzyskane tymi metodami. Tylko trzy badania dotyczą Wielkiej Brytanii, a ich wyniki znacznie się różnią, ponieważ badania prowadzone były w różnym czasie, w różnych miejscach oraz w oparciu o odmienne metody i dane. O ile wiele badań sugeruje brak znaczącego efektu substytucji, o tyle niektóre opracowania szacują go na poziomie przekraczającym nawet 50 procent (oznaczałoby to, że korzyści netto interwencji publicznej należałoby zredukować o co najmniej połowę w stosunku do szacunków brutto). Żadne z badań nie brało pod uwagę gospodarczych uwarunkowań kontekstowych.

²⁶ W połowie proponowanego przez Greenberga zakresu 30-60 procent.

²⁷ Powyższe oszacowania mają charakter bardzo ogólny i należy mieć na uwadze, że programy realizowane w innych warunkach (np. w różnych regionach) mogą cechować się różną siłą efektów substytucji.

cych efektu przeniesienia. Efekt ten jednak będzie ograniczony przede wszystkim do programów subsydiowanego zatrudnienia. Efekt przeniesienia jest poza tym trudny do odróżnienia od popytowego efektu substytucji. Greenberg i in. (2010) nie zaleca zatem uwzględniania efektu przeniesienia w kalkulacji kosztów i korzyści społecznych.

Dowody dotyczące równowagi wynagrodzeń są niepewne i w wielu badaniach nie uzyskano statystycznie znaczących wyników²⁸. Teoretycznie programy na rzecz zatrudnienia nie powinny mieć znaczącego wpływu na płace, ponieważ: po pierwsze przeciętny beneficjent podejmuje pracę nisko płatną, a możliwość obniżania zarobków innych pracowników jest wówczas ograniczana przez wysokość płacy minimalnej; po drugie grupy docelowe programów są zazwyczaj niewielkie i stanowią mały ułamek populacji osób w wieku produkcyjnym. Greenberg i in. (2010) zalecają więc, aby w analizie kosztów i korzyści nie zakładać wpływu interwencji na równowagę wynagrodzeń.

Jeśli analiza kosztów i korzyści wykorzystuje wyniki badań ewaluacyjnych, należy:

- upewnić się, czy w badaniach tych nie przeprowadzono już szacowania efektów substytucji,
- do standardowej analizy kosztów i korzyści dołączyć analizy wrażliwości dotyczące efektu substytucji,
- określić, czy dany program (lub jego komponenty) ma charakter podażowy (intensyfikacja poszukiwania pracy, rozwijanie umiejętności albo zwiększanie w inny sposób efektywnej podaży siły roboczej), czy popytowy (zwiększanie zachęt dla pracodawców do zatrudniania siły roboczej),
- na potrzeby testów wrażliwości zmniejszyć o 20 procent oszacowanie liczby dodatkowych miejsc pracy w programach na rzecz zatrudnienia o charakterze podażowym,
- obliczyć korzyści ekonomiczne netto przy uwzględnieniu tego założenia,
- na potrzeby testów wrażliwości zmniejszyć o 45 procent oszacowanie liczby dodatkowych miejsc pracy w programach na rzecz zatrudnienia o charakterze popytowym,
- obliczyć korzyści ekonomiczne netto przy uwzględnieniu tego założenia,
- wyniki testów wrażliwości zestawić z wynikami standardowej analizy kosztów i korzyści oraz opisać wszystkie zastrzeżenia.

Warto wskazać na pewne zastrzeżenia. Po pierwsze zakres efektu substytucji obarczony jest dużą dozą niepewności²⁹. Po drugie, wartości, których użycie rekomendujemy, należy traktować jako górne oszacowania (scenariusz pesymistyczny). Należy podkreślić, że efekty substytucji po stronie podaży najprawdopodobniej będą istnieć tylko w perspektywie krótkoterminowej.

2.4. Społeczne koszty budżetowe – SKB

Koszty budżetowe związane są ze zmniejszeniem efektywności ekonomicznej w wyniku transferu środków z sektora prywatnego do sektora publicznego. Istnieje wiele powodów utraty efektywności³⁰, wśród których można wyróżnić negatywne skutki większości form opodatkowania³¹. Zgodnie z klasyczną

²⁸ Zob. np. wydanie *Employment Outlook* z 1993 r., opublikowane przez Organizację Współpracy Gospodarczej i Rozwoju (OECD).

²⁹ Dla Wielkiej Brytanii istnieje bardzo mało danych pozwalających na jego pomiar. Literatura również nie dostarcza wielu wskazówek co do tego, jak efekty substytucji różnią się w zależności od warunków ekonomicznych oraz na ile są trwałe.

³⁰ Inne źródła utraty efektywności stanowią:

- i brak presji konkurencyjnej (maksymalizacja zysków) w sektorze publicznym; oraz
- ii brak możliwości ukierunkowania programów sektora publicznego na klientów, których dotyczą najmniejsze bariery związane z podjęciem pracy.

³¹ Nie stosuje się to w przypadku podatków zryczałtowanych.

teorią mikroekonomiczną ludzie są gotowi nabyć dany produkt pod warunkiem, że uznana przez nich wartość towaru (określana jako ich „skłonność do ponoszenia kosztów”) jest równa jego cenie rynkowej lub od niej wyższa. Podobne założenie dotyczy decyzji dotyczących podejmowania zatrudnienia. Ludzie skłonni są rozpocząć pracę, gdy wynagrodzenie, jakie mają otrzymać jest nie mniejsze niż szacowana przez nich wartość czasu wolnego, z którego musieliby zrezygnować.

Poprzez zmniejszanie wartości dochodów i wpływ na wzrost cen, podatki warunkują wybory rynkowe konsumentów. Z punktu widzenia rynku pracy, podatek dochodowy zmniejsza wysokość wynagrodzenia netto, co może mieć negatywny wpływ na subiektywną wartość przypisywaną czasowi pracy³². Wynika to z tego, że niektórzy ludzie bardzo cenią czas wolny i wolą przeznaczyć więcej czasu na wypoczynek niż na pracę, która ze względu na opodatkowanie nie mogłaby im dostarczyć satysfakcjonujących dochodów.

Wskaźnik SKB służy pomiarowi nadwyżki strat. Określenie „strata” wynika z tego, że gdyby nie istniał podatek dochodowy, ludzie wysoko ceniący swój czas wolny pracowaliby dłużej. Wytwarzaliby również dobra o większej wartości, co zwiększałoby ich użyteczność indywidualną oraz dobrobyt społeczny. Wartość SKB obliczana jest jako procent pobranego podatku dochodowego. Przykładowo, wartość na poziomie 0,2 oznacza, że w przypadku każdego funta brytyjskiego pobranego tytułem podatku, następuje strata 20 pensów wynikająca z utraty nadwyżki powodowanej negatywnym wpływem podatków na indywidualne wybory ekonomiczne.

Programy rządowe w różny sposób wpływają na dochody państwa, warunkując przy tym wysokość podatków niezbędnych dla ich sfinansowania. Można założyć, że interwencja publiczna, której realizacja wymaga wprowadzenia nowych podatków (lub podniesienia wysokości podatków obowiązujących) zwiększa SKB³³. Każdy funt przeznaczony na finansowanie programów wspierania zatrudnienia będzie pochodzić z podatków i dlatego zakres jego oddziaływania obejmie wszystkie rynki, z których podatek został pobrany (rynki tworzone przez podmioty objęte danym podatkiem). Zgodnie z literaturą przedmiotu przyjmujemy, że największe obciążenie stanowi podatek dochodowy³⁴.

Uwzględnienie kosztów podatkowych zapewni pełniejszą ocenę ogólnych korzyści ekonomicznych netto interwencji publicznej (Greenberg i Knight, 2007b).

Programy na rzecz zatrudnienia wdrażane przez DWP finansowane są z budżetu państwa. Zgodnie z ostatnio przeprowadzonymi analizami niektóre z tych programów generują społeczne koszty budżetowe. Oznacza to, że w celu realizacji tych programów konieczne jest nakładanie dodatkowych obciążeń podatkowych. Wprowadzenie dodatkowych danin wpływa z kolei niekorzystnie na decyzje osób bezrobotnych dotyczące podejmowania lub zaniechania zatrudnienia. Niektóre osoby, które byłyby skłonne rozpocząć pracę w zamian za minimalne akceptowalne przez siebie wynagrodzenie, powstrzymają się od tego. Obciążenia podatkowe sprawią bowiem, że wynagrodzenie netto spadnie poniżej ich granicznej wartości. Jeżeli natomiast program generuje dodatnie korzyści podatkowe netto, zmniejsza to ogólny poziom opodatkowania i zwiększa efektywność gospodarki.

Koszty budżetowe w dotychczasowych analizach kosztów i korzyści

W porównawczych analizach programów wspierania zatrudnienia nie uwzględniano dotąd w sposób systematyczny kosztów budżetowych. Zagadnienie to było jednak przedmiotem analiz kosztów i korzyści prowadzonych w innych obszarach. W dokumencie Ramy metodologiczne analizy kosztów i korzyści (*Cost*

³² Uznaje się, że opodatkowanie dochodu może w pewnych okolicznościach zwiększyć ilość pracy. W takich przypadkach efekt dochodowy objawiający się spadkiem przychodów netto byłby większy niż efekt substytucyjny.

³³ I odwrotnie – jeśli program powoduje ograniczenie wypłaty świadczeń umożliwiając zmniejszenie podatków, wartość SKB obniżyłaby się.

³⁴ Zob. na przykład: Mankiw (2008); Boardman i in. (2001); Ruggeri (1999).

Benefit Framework – CBF) (2007) oraz w raporcie PSI (2007) zaproponowano obliczanie wartości kosztów budżetowych na podstawie sumy wszystkich przepływów podatkowych netto.

W badaniu PSI dotyczącym Nowego Porozumienia w Sprawie Osób Niepełnosprawnych (2007) (ang. *New Deal for Disabled People*), zleconym przez DWP, w analizie wrażliwości uwzględniono koszt budżetowy na poziomie 25%. Ocena skutków prawnych ustawy o reformie systemu socjalnego z roku 2009 wykorzystuje tę samą miarę. Wartość tę zaczerpnięto z badania Melville'a (2002).

Dane liczbowe są spójne z szacunkami dotyczącymi kosztów budżetowych obserwowanych w innych krajach. W nowozelandzkiej publikacji *Cost Benefit Analysis Primer* (2005) zaleca się ich uwzględnianie w analizach, ilekroć istnieje przypuszczenie, że mogą one mieć znaczący wpływ na dobrobyt społeczny. Nowozelandzkie władze fiskalne zalecają przyjęcie wartości kosztu budżetowego na poziomie 20 procent. Oznacza to, że aby właściwie wyliczyć wartość kosztów społecznych należy pomnożyć przez 1,20 wydatki związane bezpośrednio z badanym programem.

Rząd australijski³⁵ zaleca stosowanie wartości 25-procentowej. Władze australijskie rekomendują również aplikowanie tego podejścia w analizach dotyczących wszystkich obszarów wydatków publicznych. Oszacowanie kosztu budżetowego zaczerpnięto w tym wypadku z badania Campbell i Bond (1997), którzy oparli swoje analizy na przykładzie podatku dochodowego.

Metodologia

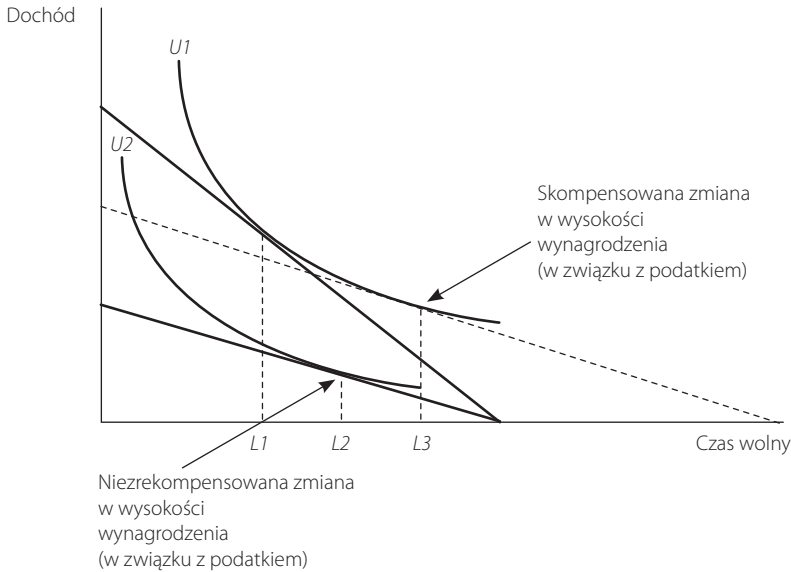
Efekty obowiązywania podatku dochodowego przedstawiono na Ryc. 5. Zgodnie z ilustracją osoba maksymalizująca swoją użyteczność staje przed wyborem odpowiedniej proporcji między czasem wolnym a pracą, będącą źródłem dochodów. Osoba dokonująca tego rodzaju wyboru może zdecydować się zrezygnować z części czasu wolnego, aby pracując przez większą ilość czasu podnieść swoje dochody. Możliwości finansowe (linia ograniczenia budżetowego) poszczególnych osób są warunkowane wysokością uzyskiwanych zarobków oraz płaconych podatków. Dla uproszczenia zakładamy, że podatek dochodowy ma charakter liniowy (zapobieganie „skręceniu się” linii ograniczenia budżetowego).

Podatek dochodowy wpływa na wybory ekonomiczne ludzi, którzy motywowani są do zwiększenia ilości czasu wolnego kosztem pracy. Przy obliczaniu wartości społecznego kosztu budżetowego redukcji czasu pracy nadawane są miary pieniężne, które przedstawione są następnie jako odsetek od wartości pobranych podatków.

Przy obliczaniu kosztu budżetowego decydujące znaczenie ma ustalenie, czy efekt substytucji lub pełnej ceny [ang. *full price effect*] (stanowiący połączenie efektu substytucji i efektu dochodowego) jest w danym przypadku istotny. W literaturze przedmiotu efekt substytucji mierzony jest z wykorzystaniem skompensowanej elastyczności podaży siły roboczej (Browning 1976, 1987; Thirsk i Moore, 1991). Zakłada się tutaj, że obniżenie wynagrodzenia jest skompensowane egzogenicznym wzrostem przychodów, dzięki czemu poziom wyjściowy użyteczności pozostaje niezmienny. W rozpatrywanym przez nas przykładzie oznaczałoby to, że dana osoba przeznaczalaby (1 – L2) czasu na pracę. Zmiana w liczbie przepracowanych godzin wynika zatem wyłącznie ze zmian w wartościach względnych pracy i wypoczynku. W innych badaniach analizowany jest całkowity efekt podatku dochodowego przy użyciu wskaźnika nieskompensowanej elastyczności podaży siły roboczej (Dahlby, 1994; Ballard i in. 1985; Fullerton i Henderson, 1989). W naszym przykładzie oznaczałoby to, że dana osoba musiałaby przeznaczyć (1 – L3) czasu na pracę

³⁵ Departament Finansów i Administracji, „Podręcznik do analizy kosztów i korzyści” (*Handbook of Cost-Benefit Analysis*) 2006, str. 37.

Ryc. 5. Wskaźnik SKB na rynku pracy³⁶



Z Ryc. 5 wynika, że przy wykorzystaniu nieskompensowanej elastyczności podaży siły roboczej otrzymujemy niższą wartość oszacowania SKB. Dzieje się tak dlatego, że redukcja liczby przepracowanych godzin pracy (wyrażona jako wydłużenie czasu wypoczynku) osiąga tutaj wyższą wartość niż w przypadku elastyczności skompensowanej. Innymi słowy, $(L2-L1) > (L3-L1)$.

Zastosowanie skompensowanej elastyczności podaży siły roboczej jest niewłaściwe, gdy podlegająca analizie polityka nie przynosi bezpośrednio korzyści podatnikowi (Ruggeri, 1999). Dzieje się tak przede wszystkim w przypadku programów na rzecz zatrudnienia, których koszt ponosi podatnik i które przynoszą ogólne korzyści osobom bezrobotnym. Podstawą do analiz powinna być wówczas nieskompensowana elastyczność podaży siły roboczej.

Ruggeri (1999) przeprowadził obszerny przegląd literatury na temat szacowania SKB. Badanie to miało bardzo duży wpływ na początkowe obliczenia przeprowadzane przez Brytyjskie Ministerstwo Gospodarki i Finansów, które oszacowało wartość kosztu na poziomie 30 procent (górną granicę) zgodnie z informacjami przedstawianymi przez Melville'a (2002). Melville (2002) dokonał obliczeń na podstawie szacunków dostępnych w literaturze przedmiotu. Wspomniane szacunki oparto zarówno na skompensowanej i nieskompensowanej elastyczności podaży siły roboczej, a więc zgodnie z powyższym zastrzeżeniem, nie możemy ich uwzględniać w naszych analizach.

W literaturze przedmiotu, której przeglądu dokonujemy na potrzeby naszych analiz, szacunki SKB opierają się wyłącznie na nieskompensowanej elastyczności podaży siły roboczej. Większość obliczeń dokonywanych jest na podstawie ogólnych zrównoważonych modeli. Najczęstszym punktem wyjścia jest maksymalizacja funkcji użyteczności (*utility*) w zależności od konsumpcji, przepracowanych godzin i egzogenicznych preferencji podlegających ograniczeniom budżetowym, które zależne są z kolei od dochodów z pracy po opodatkowaniu i dochodu uzyskanego ze źródeł innych niż praca. Jest to rozwiązanie umożliwiające maksymalizację użyteczności dla funkcji podaży siły roboczej, której argument stanowi wartość wynagrodzenia po opodatkowaniu. Może ono być zastosowane do obliczenia nieskompensowanego

³⁶ W trosce o zapewnienie możliwości uogólnienia danych, zakładamy liniowe ograniczenia budżetu. Przy uwzględnieniu korzyści niezwiązanych z pracą linia ograniczenia budżetu przebiega zgodnie z modelem liniowym.

wpływu podatku dochodowego na decyzje ekonomiczne dotyczące podejmowania pracy i ilości poświęcanego jej czasu (Kumar, 2004).

Tab. 1. Szacunki SKB wyrażone jako procent podatku dochodowego na podstawie badań wykorzystujących nieskompensowaną elastyczność podaży siły roboczej

Badanie	Kraj	SKB wyrażony jako procent dochodu z podatków (wartość uśredniona)
Dahlby (1994)	Kanada	9% – 38% (23,5%)
Stuart (1984)	USA	43% (43%)
Fullerton i Henderson (1989)	USA	6% – 17% (11,5%)
		16% – 31% (23,5%)*
Ballard i in. (1985)	USA	12% – 23% (17,5%)
		14% – 25% (19,5%)*
Campbell i Bond (1997)	Australia	19% ¹

Uwaga: Wszystkie dane szacunkowe dotyczą opodatkowania dochodów pracowniczych za wyjątkiem tych wyróżnionych symbolem*.

* Oznacza szacunki dotyczące SKB związane z podatkiem dochodowym od osób fizycznych (np. podatek od całkowitego dochodu osób fizycznych.)

¹ Wartość ta różni się od podawanej wcześniej w niniejszym tekście wartości dla Australii. Jest to spowodowane tym, że rząd australijski oparł swoje obliczenia na szacunkach pochodzących z publikacji Campbella i Bondy (1997), w której zastosowano skompensowaną elastyczność podaży siły roboczej. 19% wartość odnosi się do obliczeń autorów dotyczących SKB dokonanych przy użyciu nieskompensowanej elastyczności podaży siły roboczej.

Z zaleceń PSI i podręczników dotyczących analizy kosztów i korzyści wydanych w innych krajach OECD wynika ewidentna zgodność co do tego, że szacunki dotyczące SKB powinny być stosowane w odniesieniu do analizy kosztów (i oszczędności) podatkowych. Taki zabieg gwarantuje lepsze szacunki skutków gospodarczych danej polityki.

Przeanalizowaliśmy szacunki SKB na podstawie badań mogących wnieść wkład do analizy kosztów i korzyści. W badaniach tych przyjęto różne założenia dotyczące wysokości elastyczności nieskompensowanej podaży siły roboczej i krańcowych stawek podatkowych. Podaliśmy wyniki obliczeń oraz dodaliśmy oszacowania przedziałowe, mając na względzie niepewny charakter tych założeń³⁷.

Nasze szacunki dotyczące wartości podstawowych SKB opierają się na średnich obliczeniach z tabeli 1. Na tej podstawie otrzymujemy szacunek punktowy (20%). Przedział wartości możemy oszacować na podstawie średnich wartości niskich i wysokich szacunków z wyżej wymienionych badań³⁸. Wynik tej operacji umożliwi nam z kolei obliczenie przedziału wartości dla SKB (17-27 procent).

Powyższe dane liczbowe są porównywalne do danych zaczerpniętych z analizy CBA DWP i podręczników na temat CBA z Australii i Nowej Zelandii. Zgodnie z założeniami CBA, ekonomiści DWP wyliczyli, że wartość SKB przy zwiększeniu dochodów podatkowych w Wielkiej Brytanii wynosiłaby około 25%. W rzeczywistości jest to uśredniona wartość zaczerpnięta z obliczeń Melville'a (2002). Wartości naszych obliczeń są znacznie niższe niż (Greenberg i Knight, 2007b) wartość górnej granicy SKB oszacowanej przez PSI (40%). Opiera się ona na zaleceniach Boardmana i in. (2001), którzy powołują się na dwa badania ze Stanów Zjednoczonych, w których zastosowano elastyczność skompensowaną.

³⁷ Należy zauważyć niewielką dostępność danych dotyczących Wielkiej Brytanii. Dane pochodzące z innych krajów mogą być jednak wykorzystywane również w brytyjskim kontekście. Uważamy, że funkcje podaży siły roboczej i systemów podatkowych mogą być w tych krajach wystarczająco podobne, co pozwala na wykorzystanie uzyskiwanych w tamtejszych badaniach oszacowań. W opracowaniach PSI oraz brytyjskich analizach kosztów i korzyści wykorzystano dotychczas z oszacowań międzynarodowych. W naszych obliczeniach korzystaliśmy z dostępnych danych krajowych. Widzimy przy tym potrzebę bardziej systematycznego włączenia zagadnienia SKB w analizach kosztów i korzyści.

³⁸ Uwaga: Szacunki punktowe Dahlby'ego (1994) a także Campbella i Bondy (1997) bazują na elastyczności podaży siły roboczej, która równa jest elastyczności przyjętej w niższych szacunkach z innych badań. Są one w ten sposób uwzględnione w niższych szacunkach punktowych na potrzeby wspomnianych obliczeń.

Uważamy, że nasze obliczenia są lepsze z punktu widzenia założeń prowadzonej przez nas analizy CBA, gdyż szacunki opierają się wyłącznie na badaniach, w których zastosowano nieskompensowaną elastyczność podaży siły roboczej. Szacunki Melville'a (2002) i Boardmana i in. (2001) są wyższe, ponieważ uwzględniają obliczenia oparte na skompensowanej elastyczności podaży siły roboczej. Zalecamy analitykom stosowanie następującej procedury przy prowadzeniu analizy CBA:

- 1) W przypadku analizy wrażliwości: należy obliczyć całkowite korzyści podatkowe netto z dodatkowych miejsc pracy i wartość tę pomnożyć przez 0,20, aby w ten sposób uzyskać wysokość SKB.
- 2) Należy zauważyć, że w przypadku gdy korzyści podatkowe netto programu są dodatnie, wartość SKB również będzie dodatnia.
- 3) Analiza powinna zostać uzupełniona o pewne zastrzeżenia:
 - a. w związku z tym, że w przytaczanych badaniach przyjęto bardzo różne założenia musimy zachować pewną ostrożność przy stosowaniu pochodzących z nich danych liczbowych.
 - b. różnice założeń dotyczą stawek podatkowych i elastyczności podaży siły roboczej. Aby uwzględnić ten problem, nasze szacunki opieramy na różnych badaniach.
 - c. szacunki dotyczące elastyczności podaży siły roboczej nie zostały zdezagregowane według płci. Kumar (2004) sugeruje, że wartość elastyczności podaży siły roboczej dla kobiet może być wyższa niż dla mężczyzn. Z uwagi na to, że brytyjska siła robocza cechuje się większą liczebnością mężczyzn (którzy z kolei płacą podatek dochodowy), może to oznaczać, że nasze szacunki dla wartości SKB są zawyżone.

Powyższe podejście obrazuje przykład program ND25+, całkowite korzyści podatkowe netto z dodatkowych miejsc pracy wynoszą około 48 500 000 GBP. Oznacza to, że konieczne są dodatkowe dochody z podatków w celu finansowania programu – programu, który wiąże się z kosztami społecznymi.

Tab. 2. Szacunki dotyczące SKB dla ND25+ przy różnych założeniach

	Podstawowe obliczenia	Przedział wrażliwości
Obecne założenia	12 125 000 GBP	12 125 00 – 19 400 000 GBP
Zalecane nowe założenia	9 700 000 GBP	8 245 000 – 13 095 000 GBP

Uwaga: Bieżące założenia zaczerpnięto z modelu Tax Ben. Zgodnie z podstawowymi obliczeniami wartość SKB wynosi 25 procent a przedział wartości wskaźnika: 25-40%.

3. Wpływ na rynkach wtórnych

3.1. Wpływ programów na rzecz pracy i zatrudnienia na zdrowie

Istnieją przesłanki empiryczne świadczące o tym, że osoby zatrudnione są na ogół zdrowsze od osób bezrobotnych. W literaturze medycznej dominuje tradycyjnie niepewność dotycząca kierunku tej zależności, tzn. brakuje rozstrzygnięcia, czy to praca powoduje poprawę stanu zdrowia, czy też jest tak, że dobry stan zdrowia zwiększa prawdopodobieństwo zatrudnienia.

Poprawa jakości dostępnych danych, jaka zaszła w ostatnim czasie oraz rozwój nowatorskich badań empirycznych wniosły istotny wkład w wyjaśnienie tej przyczynowości oraz wytyczyły ramy dla ewaluacji wpływu programów wspierania zatrudnienia wdrażanych przez Departament Pracy i Świadczeń (DWP).

Zatrudnienie, zdrowie oraz analiza kosztów i korzyści

Kluczową kwestią w ocenie kosztów i korzyści jest oszacowanie wpływu poprawy stanu zdrowia związanej z podjęciem zatrudnienia na wzrost indywidualnego dobrobytu i użyteczności. Na gruncie ekonomii wiodącym modelem analizy decyzji dotyczących spraw zdrowotnych oraz zmian stanu zdrowia jest model popytowy (*demand for health*) Grossmana (Grossman 1972a). Zgodnie z jego założeniami osoby podejmując decyzje dotyczące konsumpcji i czasu wolnego określają swój optymalny poziom zdrowia. Zazwyczaj przyjmuje się tu założenie o rynku idealnym, które implikuje, że aktualny stan zdrowia poszczególnych osób znajduje się na poziomie optymalnym w rozumieniu wspomnianego modelu popytowego (zob. np. Grossman 1972a; Nozal i in. 2004; Sickles, 1998). W takim przypadku jakkolwiek zmiana stanu równowagi spowodowana na przykład podjęciem zatrudnienia, miałaby niekorzystny wpływ na indywidualną użyteczność (nawet jeśli zmiana prowadzi do poprawy stanu zdrowia danej osoby).

Na potrzeby naszych analiz przyjęliśmy jednak założenie, że decyzje tego rodzaju nie mają charakteru optymalnego. Przyjmujemy bowiem, że dostęp do informacji jest nierówny. Wówczas stan zdrowia i indywidualna użyteczność osób bezrobotnych utrzymują się poniżej poziomu optymalnego. Dlatego też wykazujemy, iż podjęcie pracy wpływa na poprawę stanu zdrowia i prowadzi tym samym do zwiększenia nadwyżki konsumenta. Wpływ w tym obszarze powinien być uwzględniany w analizach kosztów i korzyści.

Grossman (1972a) zakłada, że osoby w długiej perspektywie czasowej maksymalizują swoją użyteczność zgodnie z następującą funkcją³⁹:

$$\max_{c,t} \sum_t \beta^t u_t(c_t, h_t, l_t) \quad (1)$$

Maksymalizacja prowadzona jest w ramach indywidualnych ograniczeń budżetowych. Zdrowie stanowi w tym układzie zmienną systemową (*state variable*⁴⁰), a funkcja produkcji (krzywa dynamiki) przyjmuje dla stanu zdrowia następującą postać:

$$h_{t+1} = f_t(\delta_t, h_t, I_t)^{41}$$

przy czym c = konsumpcja w czasie t ; l_t = czas wolny (w niniejszym przypadku $(1 - l_t)$ to czas poświęcany pracy); h_t = stan zdrowia; δ_t = uszczerbek na zdrowiu; I_t = inwestycje w zdrowie (inwestycje w zdrowie są efektem decyzji o alokacji zasobów, w ramach których alternatywną opcję stanowi konsumpcja).

Zdrowie traktowane jest jako trwałe dobro kapitałowe, które z czasem traci na wartości. W modelu Grossmana osoby mogą inwestować czas i zasoby w poprawę swojego stanu zdrowia. Marginalna korzyść wynikająca z inwestycji w zdrowie pozostaje niezmienna we wszystkich okresach życia.

W modelu Grossmana standardowo nie przyjmuje się założenia o wpływie zatrudnienia na poziom zdrowia, tzn. zakłada się, iż praca nie wchodzi w zakres funkcji produkcji. Biorąc pod uwagę ustalenia naryskowane w niniejszym dokumencie rozwijamy model Grossmana, twierdząc, że ludzie mogą „inwestować” czas poświęcony pracy w celu poprawy swojego zdrowia. Funkcja produkcji w równaniu Bellmana wygląda zatem następująco:

$$h_{t+1} = f_t(\delta_t, h_t, I_t, (1 - l_t)), \text{gdzie } \frac{\delta h_{t+1}}{\delta (1 - l_t)} > 0$$

³⁹ Formułę modelu Grossmana (1972) dostosowujemy do języka i formatu problemu związanego z dynamicznym programowaniem, który jest standardowym narzędziem do optymalizacji dynamicznej.

⁴⁰ Zmienne systemowe (*state variable*) – zmienne opisujące aktualny stan dynamicznego systemu, np. liczebność populacji, temperatura, ciśnienie atmosferyczne (przyj. red).

⁴¹ Po szczegółowe informacje na temat metodologii programowania dynamicznego odsyłamy do Bergmana (2005) oraz Ljungqvista i Sargenta (2004).

Optymalizacja decyzji alokacyjnych w ramach tej formuły modelu Grossmana możliwa będzie wyłącznie w warunkach pełnego dostępu osób do informacji o wpływie pracy na stan zdrowia. Zakładamy, że ludzie nie są świadomi korzyści zdrowotnych płynących z zatrudnienia. W konsekwencji, nie są oni w stanie podejmować optymalnych decyzji dotyczących ilości czasu poświęconego pracy. Tym samym ich stan zdrowia będzie utrzymywał się na poziomie suboptymalnym.

Aby tego dowieść na satysfakcjonującym poziomie ogólności, model Grossmana odniesiemy tylko do jednego okresu czasowego⁴². Zakładamy, że krzywa dynamiki stanu zdrowia przyjmuje następującą postać:

$$h_{t+1} = I_t - \delta_t h_t + h_t + \phi(1 - l_t) \quad (2)$$

przy czym ϕ to funkcja określająca wpływ pracy na zdrowie. Zakładamy, że $\phi_{(1-l)}$ $(1 - l_t) > 0$ i że nie obserwujemy żadnego uszczerbku na zdrowiu ($\delta = 0$). Wówczas równanie (2) uzyska kształt:

$$h_t = I_t - h_{t+1} + \phi(1 - l_t) \quad (3)$$

Podstawiając do wzoru (1) wartość zdrowia zgodnie z równaniem (3) uzyskujemy dla określonego czasu t następującą postać funkcji maksymalizacji użyteczności:

$$\max_{c_t, l_t} u_t \{c_t, l_t, [I_t - h_{t+1} + \phi(1 - l_t)]\} \quad (4)$$

Przyjmując bardzo proste założenie dotyczące ograniczeń budżetowych: $w \cdot L_t = c_t$, przy czym w to wynagrodzenia rynkowe, c to konsumpcja ogółem, którą można przydzielić do konsumpcji dóbr lub inwestycji w zdrowie, a $L_t = (1 - l_t)$ to czas przeznaczany na pracę.

W tym ujęciu modelu Grossmana osoba dysponująca pełnym dostępem do informacji będzie podejmowała decyzje biorąc pod uwagę ilość czasu wolnego warunkującego poziom jej użyteczności oraz stan zdrowia.

Poniższe równania określają warunki niezbędne dla optymalizacji konsumpcji, ilość czasu wolnego alokowaną przez jednostkę posiadającą pełny dostęp do informacji oraz cenę kalkulacyjną (czasu wolnego) przy indywidualnych ograniczeniach budżetowych⁴³:

$$\frac{\delta u}{\delta c} = u_1 = -\lambda \quad (5)$$

$$\frac{\delta u}{\delta l} = u_2 + u_3 \cdot \phi_1(1 - l_t) = \lambda \cdot w \quad (6)$$

$$\frac{\delta u}{\delta \lambda} = w \cdot L - c = 0 \quad (7)$$

Oznacza to, że:

$$u_2 + u_3 \cdot \phi_1(1 - l_t) = \lambda \cdot \frac{c}{L} \quad (8)$$

⁴² W kompletnym modelu dynamicznym konieczne byłoby jednoczesne rozwiązanie wszystkich okresów t – zazwyczaj przy użyciu równania Bellmana. Prosty przykład przypadku jednookresowego stosuje się tutaj na potrzeby klarowności bez straty ogólności (Bergmann, 2005).

⁴³ Przy czym u , odnosi się do pochodnej u w odniesieniu do drugiego argumentu i tak dalej w przypadku u_3 .

⁴⁴ $\phi_1(1 - l_t) = \frac{\delta h_{t+1}}{\delta(1 - l_t)}$

Natomiast optymalna ilość czasu pracy w przypadku osoby z pełnym dostępem do informacji (L_{PI}^*) przyjmowała będzie wartość zgodną z poniższym wzorem:

$$\rightarrow L_{PI}^* = \frac{\lambda \cdot c}{u_2 + u_3 \cdot \phi_1(1 - l_t)} \quad (9)$$

Odnieśmy powyższe równania do sytuacji, w której dostęp do informacji jest ograniczony. Wówczas $\phi(1 - l_t)$ przyjmuje wartość zero a równania (5), (6) i (7) przyjmują postać:

$$\frac{\delta u}{\delta c} = u_1 = -\lambda \quad (10)$$

$$\frac{\delta u}{\delta l} = u_2 = \lambda \cdot w \quad (11)$$

$$\frac{\delta u}{\delta \lambda} = w \cdot L - c = 0 \quad (12)$$

W tym przypadku, ze względu na ograniczenia w dostępie do informacji (L_{NI}^*), ilość czasu poświęcanego na pracę określa wzór:

$$\rightarrow L_{NI}^* = \frac{\lambda \cdot c}{u_2} \quad (13)$$

Ponieważ zakładamy, że zdrowie ma pozytywny wpływ na użyteczność, to $u_3 > 0$. Wydłużenie czasu wolnego ma negatywny wpływ na zdrowie, ponieważ wiąże się z mniejszą ilością czasu poświęcanego na pracę, która zapewnia większe korzyści zdrowotne. Tym samym $\phi_1(1 - l_t)^{45}$. Jeśli $u_3 > 0$ a $\phi_1(1 - l_t) < 0$, to wówczas $\phi_1(1 - l_t) < 0$ i możemy stwierdzić, że:

$$u_2 > u_2 + u_3 \cdot \phi_1(1 - l_t)^{46} \quad (14)$$

Oznacza to, że $L_{PI}^* > L_{NI}^*$

Marginalna użyteczność czasu poświęcanego pracy jest w sytuacji niepełnego dostępu do informacji mniejsza niż gdyby dostęp ten był nieograniczony. Wynika to z braku kompletnej wiedzy dotyczącej korzystnego wpływu zatrudnienia na stan zdrowia. Osoba pozbawiona pełnego dostępu do informacji przeznaczająca więc będzie na pracę mniejszą ilość czasu ($L_{PI}^* > L_{NI}^*$). Osoba podejmująca racjonalną decyzję w oparciu o pełen zakres informacji dotyczących wpływu zatrudnienia na zdrowie optymalizowałyby swoją użyteczność pracując przez większą ilość czasu. Wartość $(1 - l_t)$ będzie zatem w jej przypadku większa niż wśród osób opierających się na informacji niepełnej.

Ponieważ zakładamy, że $\frac{\delta u_{t+1}}{\delta(1 - l_t)} > 0$, możemy stwierdzić, że osoby dysponujące pełną informacją będą zdrowsze ze względu na większą ilość czasu przeznaczanego na pracę. Ponieważ założyliśmy występowanie zawodności rynku związanych z luką informacyjną, podjęcie zatrudnienia i w konsekwencji poprawa stanu zdrowia przekładają się na wzrost indywidualnej nadwyżki konsumenta. Zmiany te powinny być więc uwzględniane w analizie kosztów i korzyści⁴⁷.

⁴⁵ Uwaga wysnuta na podstawie założenia, że: $u_2 > u_3 \cdot \phi_1(1 - l_t)$

⁴⁶ Ponieważ przepracowane godziny nie mogą być ujemne, zakładamy następujące ograniczenie: $u_2 > u_3 \cdot \phi_1(1 - l_t)$

⁴⁷ Proszę zwrócić uwagę na fakt, iż oczywiście jest możliwość poprawy wiedzy ludzi na temat korzyści zdrowotnych płynących z pracy, co zgodnie z przewidywaniami naszego modelu prowadziłoby do wzrostu zatrudnienia. W każdym razie wydłużamy czas spędzany przez ludzi w pracy, co prowadzi do wzrostu nadwyżki.

Możemy przyjąć, że biorąc pod uwagę ograniczenie budżetowe $w \cdot L_t = c_t$, przy stałej wartości wynagrodzenia (w), kwota pieniędzy dostępnych na ogólne cele konsumpcyjne będzie większa w przypadku osób dysponujących pełnym dostępem do informacji, ponieważ będą one pracować przez większą ilość czasu. Ponieważ inwestycje w zdrowie stanowią pozytywną funkcję kwoty pieniędzy dostępnych do ogólnej konsumpcji, wydłużenie czasu pracy również poprawi stan zdrowia poprzez zwiększenie dochodów.

Założenie o niepełnym dostępie do informacji ma kluczowe znaczenie dla analiz zamieszczonych w niniejszym tekście oraz ogólnie dla analizy kosztów i korzyści w obszarze programów wspierania zatrudnienia. Jeśli osoby bezrobotne dysponowałyby nieograniczonym dostępem do informacji, mogłyby maksymalizować użyteczność poprzez powstrzymanie się od podjęcia zatrudnienia, a tym samym efekty interwencji publicznej w postaci nowych miejsc pracy miałyby niekorzystny wpływ na indywidualną użyteczność tych osób⁴⁸.

Zatrudnienie i zdrowie

Opracowanie Jahody (1982) i Fryera (1986) stanowi fundament badań empirycznych w zakresie relacji między zatrudnieniem i zdrowiem. Jahoda w swojej teorii ukrytej deprivacji (*Latent Deprivation Theory* 1982) wprowadza pięć kategorii doświadczenia związanych z pracą, które mają wpływ na stan zdrowia. Są to: ustrukturyzowane wykorzystanie czasu, aktywność, kontakty społeczne, wspólny cel i status. Pozbawianie tych doświadczeń wpływa niekorzystnie na dobrostan psychiczny. Fryer w publikacji *Agency Restriction Model* (1986) wykazuje, że powodem zaburzeń dobrostanu psychicznego osób bezrobotnych jest utrata dochodów. Zjawisko to w literaturze anglojęzycznej określa się terminem deprivacji dochodowej (*economic deprivation*). Wedle tej teorii wzrost dochodów przekłada się na poprawę jakości żywienia i stylu życia, redukuje napięcia związane z niedostatkiem materialnym oraz pozwala na łatwiejszy dostęp do usług medycznych.

Model Grossmana w zaproponowanej przez nas formie, uwzględnia relacje pomiędzy zatrudnieniem a zdrowiem, w tym ukryte i ekonomiczne korzyści wynikające z poprawy jego stanu. Model ten dostarcza więc odpowiednich ram dla analizy kosztów i korzyści w zakresie wpływu programów rynku pracy na zdrowie publiczne.

Objęcie analizą korzyści zdrowotnych (ukrytych i ekonomicznych) wiąże się z ryzykiem popełnienia błędu podwójnego liczenia. Poprawa stanu zdrowia może wpływać na wzrost dochodów, który uwzględniany jest na innym etapie analizy. Aby uniknąć pomyłki powinniśmy uwzględnić wyłącznie korzyści związane z ukrytymi funkcjami zatrudnienia. Do kategorii tej moglibyśmy zaliczyć na przykład korzyści następujących rodzajów⁴⁹:

- zmniejszenie wydatków publicznych na opiekę zdrowotną;
- zmniejszenie kosztów ponoszonych prywatnie na opiekę zdrowotną (na przykład kupowane przez pacjentów leki lub koszt prywatnej opieki zdrowotnej);
- korzyści niematerialne wynikające z lepszego stanu zdrowia i samopoczucia psychicznego.

Powyższe kategorie nie stanowią kompletnej typologii. Mogą one jednak stanowić przybliżoną miarę nadwyżki konsumenta generowanej przez poprawę stanu zdrowia osób podejmujących zatrudnienie.

Badania wydłużone

1) Korzyści ukryte wynikające z zatrudnienia

Graetz (1993) przeprowadzając analizy stanu zdrowia psychicznego, wykorzystuje wyniki czterech badań Australian Longitudinal Survey oraz sondażu General Health Questionnaire. Stosuje on jednokierunkową analizę wariancji (ANOVA) w celu oceny statystycznej istotności różnic międzygrupowych. Kontrolując wpływ wyjściowego stanu zdrowia Graetz (1993) stwierdza, że osoby bezrobotne, które podjęły

⁴⁸ Miałoby to miejsce w przypadku, gdyby nie występowały zawodności rynkowe.

⁴⁹ W literaturze wyjaśniono czwartą potencjalną korzyść: większą produktywność w wyniku mniejszej zachorowalności i nieobecności w pracy. Zakładamy, że pracodawcy biorą tę niewątpliwą korzyść pod uwagę w procesie negocjacji płacowych. Wynik w zakresie wynagrodzenia jednostki obejmuje wówczas wartość większej produktywności dzięki obecności w pracy i nie ma potrzeby wylizania go oddzielnie na potrzeby analizy kosztów i korzyści.

pracę, deklarowały znaczącą poprawę swojego psychicznego samopoczucia. Istotne zastrzeżenia dla tego rodzaju wnioskowania stanowi fakt, iż wpływ zatrudnienia na kondycję psychiczną jest uzależniony również od rodzaju pracy. Według Graetza „konsekwencje związane ze znalezieniem pracy, jakie obserwujemy na poziomie zdrowia psychicznego są uwarunkowane jakością podejmowanego zatrudnienia”.

Powyższe wnioski znajdują potwierdzenie w wynikach badania przeprowadzonego przez zespół pod kierunkiem Roya (1987) wskazujących, że bycie bezrobotnym zwiększa skalę zaburzeń psychicznych. Z kolei według wyników badania zespołu Thomasa (2005), w którym wykorzystano skalę oceny zaczerpniętą z General Health Survey oraz wyniki badania panelowego gospodarstw domowych (*British Household Panel Survey*) z lat 1991-1999, podjęcie zatrudnienia stanowi czynnik ograniczający ryzyko rozwoju chorób psychicznych. Gerdtham i Johannesson (2002) podejmują bardziej kompleksowe badanie oceniające wpływ zarówno ukrytych, jak i ekonomicznych korzyści wynikających z zatrudnienia. Weryfikują oni korelację pomiędzy bezrobociem i dochodami z jednej strony oraz wskaźnikiem śmiertelności z drugiej. Analiza bazuje na danych dotyczących 30 000 osób w wieku 20-64 lata, gromadzonych na przestrzeni 17 lat. Badacze uwzględniając czynniki demograficzne i geograficzne, stwierdzają, że bezrobocie podnosi ryzyko śmierci z 5,36% do 7,33%. Poziom dochodów jest również statystycznie istotny oraz umiarkowanie negatywnie skorelowany ze śmiertelnością. Bardzo podobne korelacje między bezrobociem a wskaźnikiem zgonów stwierdzili Iverson i in. (1987) na podstawie analiz dotyczących Danii⁵⁰ oraz Martikainen i Volkonen (1996) w odniesieniu do danych pochodzących z Finlandii⁵¹.

2) Korzyści ekonomiczne wynikające z zatrudnienia

Benezal i Judge (2001) przeprowadzają pełen przegląd literatury pod kątem badań w zakresie związków dochodów ze stanem zdrowia. Obejmuje on 14 badań wydłużonych przeprowadzonych w USA, Niemczech, Kanadzie i Szwecji. W mniej więcej połowie badań dokonuje się pomiaru wskaźnika zgonów. Prawie wszystkie pozostałe badania zawierają subiektywne miary kondycji psychicznej. We wszystkich badaniach kontrolowane było występowanie odwróconej relacji przyczynowo-skutkowej oraz wpływ innych niż zatrudnienie i dochody determinant stanu zdrowia. Wyniki wszystkich analizowanych badań wskazują, że dochody mają relatywnie niewielki, lecz statystycznie istotny, pozytywny wpływ na stan zdrowia. Z niektórych badań wynikało, że wpływ dochodów rośnie wraz z utrzymywaniem się ich na określonym poziomie przez dłuższy okres (*long-term/permanent income*).

Prace, w których wykorzystuje się subiektywne miary dobrostanu psychicznego, ujawniły również wpływ dochodów na stan zdrowia, choć bezpośredni wpływ na samopoczucie psychiczne okazał się bardzo słaby (zob. na przykład Easterlin, 1974; Ferriera i Moro, 2009; Groot i in., 2006). W istocie, silniejszy wpływ na poziom zdrowia wykazują ukryte korzyści związane z zatrudnieniem (Winkelmann i Winkelmann, 1998)). Według wyników badania, które przeprowadzili Clark i Oswald (1994; 2002) wpływ korzyści ukrytych jest niemal dziesięciokrotnie silniejszy niż korzyści ekonomicznych.

Badania oparte na eksperymentach naturalnych

1) Korzyści ukryte i ekonomiczne wynikające z zatrudnienia

Likwidacja zakładów pracy wykorzystywana jest w badaniach jako sytuacja eksperymentu naturalnego mającego na celu określenie wpływu zatrudnienia na stan zdrowia. Sytuacje te są wykorzystywane ze względu na brak ryzyka związanego z wystąpieniem efektu odwróconej relacji przyczynowo-skutkowej.

Badania prowadzone w tych warunkach obciążone są jednak innego rodzaju ograniczeniami. Rozmiar próby jest zazwyczaj mały, a krótki czas trwania badania terenowego (*follow-up research*) uniemożli-

⁵⁰ Iverson i in. (1987).

⁵¹ Martikainen i Volkonen (1996).

wia prowadzenie analiz osadzonych w dłuższej perspektywie. Wykorzystywane są również różne wskaźniki rezultatów, co utrudnia prowadzenie analiz porównawczych.

W badaniach tych stwierdza się na ogół istotny statystycznie negatywny wpływ bezrobocia na zdrowie psychiczne i fizyczne w przypadku zachorowań rejestrowanych klinicznie oraz deklarowanych przez osoby badane. Iversen i Sabroe (1989) odnotowali po zamknięciu dużej miejscowej stoczni rosnącą liczbę hospitalizacji w badanej próbie duńskich robotników, a Keefe i in. (2002) opisali nadmierne ryzyko samo-okaleczenia prowadzącego do hospitalizacji lub śmierci na przykładzie byłych pracowników zlikwidowanego zakładu mięsnego. Podobne rezultaty zaobserwowano w przypadku zamknięcia dużego zakładu produkującego meble w Austrii (Studnicka i in., 1991). Badacze dochodzą do wniosków, że wraz z zatrudnieniem wzrastały subiektywne wskaźniki zdrowotne i stopień korzystania z usług medycznych. Sullivan i von Wachter (2006), wykorzystując dane urzędowe z dwóch stanów USA, szacują, że ryzyko zgonu w okresie kolejnych 20 lat jest w przypadku osób bezrobotnych o 15-20% większe.

Beale i Nethercott (1985, Wiltshire, Anglia) oraz Iversen i Sabroe (1989, Nordhavn, Dania) odnotowali, że po zamknięciu zakładu pracy o około 50% wzrósł wskaźnik korzystania z konsultacji lekarzy pierwszego kontaktu. Wyniki te są zgodne z niedawnymi ustaleniami Mathersa (1994) dotyczącymi Australii. W badaniach tych wykorzystano zestaw ściśle powiązanych grup kontrolnych, który pomógł kontrolować wpływ innych czynników warunkujących stan zdrowia.

Burgard i in. (2005) stwierdzają, że przymusowe zwolnienia mają silny negatywny wpływ na subiektywne oceny własnego stanu zdrowia. W badaniach wspomnianych autorów wykorzystano zestaw danych długookresowych z USA, co świadczy o tym, iż wyniki oparte na analizach dotyczących określonych obszarów geograficznych lub branż mogą być uogólniane na całą populację tworzącą siłę roboczą danego kraju.

Dostępne są również dane świadczące o tym, że utrata pracy nie zawsze prowadzi do znaczących zmian w zakresie korzystania z usług medycznych i wydatków konsumpcyjnych. Kuhn i in. (2009) wyprawdzają tego rodzaju wnioski wobec dużej próby pracowników, którzy stracili pracę w wyniku zamknięcia jednego z austriackich zakładów pracy. Należy jednak zaznaczyć, że w omawianym przypadku wzrosła liczba wypłacanych zasiłków chorobowych. Z faktu, że podstawą do wypłaty zasiłków jest opinia lekarza pierwszego kontaktu, można wywnioskować, że utrata pracy miała negatywny wpływ na stan zdrowia osób, którym były one wypłacane.

2) Korzyści ekonomiczne wynikające z zatrudnienia

Wiedzy o relacjach pomiędzy poziomem dochodów i stanem zdrowia dostarczają także badania eksperymentalne niezwiązane bezpośrednio z tematyką zdrowia czy też zatrudnienia⁵². Lindhal (2005) korzysta z wyników szwedzkiego badania poziomu życia (*Swedish Level of Living Survey*) uwzględniającego dochody, wygrane na loteriach oraz wskaźniki stanu zdrowia. Autor stwierdza, że wygrane na loteriach prowadzą do poprawy szerokiego spektrum aspektów zdrowia fizycznego i psychicznego. Ponieważ wygrane na loteriach rozpatrywane są jako element egzogeniczny, sugeruje to, że dochody mają pozytywny wpływ na stan zdrowia. Ustalenie to potwierdza Frijters i in. (2005), którzy analizują egzogeniczne czynniki wzrostu dochodów w Niemczech Wschodnich po zjednoczeniu. Stwierdzają oni, że poziom dochodów ma pozytywny (choć niezbyt silny) wpływ na subiektywną ocenę stanu zdrowia. Badania, w których celem kontroli odwróconych zależności przyczynowo-skutkowych zastosowano dane długookresowe i eksperymenty naturalne wykazały pozytywny wpływ zatrudnienia na kondycję fizyczną i psychiczną. Prawidłowość tę potwierdzają obserwacje szerokiej gamy wskaźników dotyczących zdrowia, co w jeszcze większym stopniu podnosi wiarygodność wniosków. Wpływ zatrudnienia wynika

⁵² Należy zwrócić uwagę, że wpływ zmiany dochodów na stan zdrowia niekoniecznie musi być taki sam, jak wpływ zmiany dochodów w wyniku zmiany stanu zatrudnienia. Niemniej jednak badania te pomagają wykazać, że dochody stanowią istotny wyznacznik stanu zdrowia.

z korzyści ukrytych i ekonomicznych, jednakże dane empiryczne wskazują, iż pierwsze z nich mają większe znaczenie. Niektóre badania wykazały, że kluczowy jest długookresowy poziom dochodów, który oddziałuje na stan zdrowia znacznie silniej niż poziom krótkookresowy. Wiemy również, że ważnym czynnikiem jest nie tylko sama praca, ale także jej jakość. W kontekście analizy kosztów i korzyści nie wypracowano jednak dobrej definicji jakości zatrudnienia. W oparciu o dostępne przesłanki empiryczne możemy więc przyjąć, że programy wspierania zatrudnienia, pomagając ludziom podjąć pracę, mogą mieć korzystny wpływ na stan ich zdrowia.

Ponadto w literaturze pojawiły się w ostatnio dane świadczące o tym, że korzystny wpływ na stan zdrowia ma już samo uczestnictwo w programach prozatrudnieniowych. Uczestnictwo w szkoleniach oraz kontakty z innymi osobami znajdującymi się podobnej sytuacji, mogą dostarczać podobnych profitów, jak ukryte korzyści związane z zatrudnieniem⁵³. Badania weryfikujące powyższe hipotezy są nadal w fazie wstępnej i nie mogą być tutaj wykorzystane. Możemy jednak założyć, że programy wspierania zatrudnienia dostarczają zróżnicowanego rodzaju korzyści, wykraczających poza tworzenie miejsc pracy.

Powyżej zostały zaprezentowane mocne podstawy dowodowe, świadczące o tym, że zatrudnienie ma pozytywny wpływ na stan zdrowia. Wpływ ten wynika z korzyści ukrytych i ekonomicznych. Zakładając, że ludzie nie dysponują pełną informacją dotyczącą wpływu zatrudnienia na ich stan zdrowia, wykazaliśmy, że zapewnienie pracy prowadzi do wzrostu nadwyżek konsumenta (beneficjenta programu), który należy uwzględnić w analizach kosztów i korzyści.

3.2. Wpływ programów prozatrudnieniowych na poziom przestępczości

Pomimo że brakuje jednoznacznych dowodów potwierdzających zależność pomiędzy bezrobociem a poziomem przestępczości⁵⁴, to już sama relacja pomiędzy przestępczością a dochodami wydaje się silna (istnieje z kolei oczywisty związek między zatrudnieniem a dochodami). Istnieją mocne podstawy empiryczne potwierdzające związek między poziomem dochodów a przestępstwami przeciwko mieniu. Nie znaleźliśmy natomiast mocnych dowodów świadczących o zależności między dochodami a przestępstwami z użyciem przemocy (Levitt, 2004).

Relacje pomiędzy przestępczością a zjawiskami zachodzącymi na rynkach pracy okazują się więc skomplikowane. Poziom bezrobocia jest wypadkową procesów typowych dla rynku pracy oraz uwarunkowań makroekonomicznych. Aby określić wpływ programów prozatrudnieniowych na przestępczość należy wyodrębnić poszczególne rodzaje efektów generowanych przez bezrobocie. Odwołam się w tym celu do różnych podejść ekonomicznych.

W klasycznej teorii ekonomicznej decyzje o popełnieniu przestępstwa analizowane są z punktu widzenia maksymalizacji indywidualnej użyteczności (ang. *utility maximisation*)⁵⁵. Zgodnie z teorią użyteczności oczekiwanej racjonalnie postępująca osoba popełnia przestępstwo, gdy oczekuje, że korzyści zwią-

⁵³ Zob. pełną dyskusję: Coutts (2009).

⁵⁴ Badania statystyczne wykazują związek między bezrobociem a przestępczością. Ta statystyczna relacja została podważona przez analizy prowadzone w latach 80. Dostępne były wówczas już doskonalsze metody szacowania skutków działalności przestępczej i poczyniono postępy w zakresie teoretycznym w tej dziedzinie. Ekonomisci są obecnie zgodni co do tego, że nie ma niezbitych dowodów na to, że poziom bezrobocia jest pozytywnie związany z przestępczością (np. Hansen i Machin, 2002, str. 678; Gould i in., 1998 str. 1; Wilson i Cook, 1985, str. 4). Z drugiej strony eksperci podzielają pogląd o negatywnym wpływie bezrobocia na ogólny stan gospodarki. W warunkach wysokiego bezrobocia, skłonność do popełniania przestępstw może wzrastać. Jednocześnie maleje jednak ilość dóbr, które mogłyby stać się przedmiotem kradzieży. W konsekwencji w wielu dokumentach stwierdzono negatywny wpływ bezrobocia na poziom przestępczości.

⁵⁵ Becker (1968); Ehrlich (1973).

zane z nim przewyższą korzyści, jakie uzyskiwałby w sytuacji zaniechania tego czynu (zaniechanie w naszym przypadku oznaczałoby ograniczanie się do uzyskiwania dochodu z legalnej pracy lub ze świadczeń społecznych)⁵⁶.

Osoba podejmująca decyzje dotyczące ewentualnego popełnienia przestępstwa, bierze pod uwagę:

- prawdopodobieństwo wykrycia przestępstwa;
 - surowość ewentualnej kary;
 - korzyści finansowe popełnienia przestępstwa;
 - korzyści finansowe legalnej działalności.
- } Czynniki wpływające na oczekiwane korzyści w przypadku popełnienia przestępstwa

Wyższe prawdopodobieństwo wykrycia przestępstwa, wyższa surowość kary i niższe korzyści finansowe przestępstwa zmniejszają skłonność danej osoby do jego popełnienia ze względu na mniejsze oczekiwane korzyści. Wyższe dochody ze źródeł legalnych również zmniejszają skłonność do popełnienia przestępstwa.

Zgodnie z tą teorią wpływ zatrudnienia na przestępczość wiąże się z wynagrodzeniem. Zatrudnienie zwiększa dochód, a dochód z kolei zwiększa (przy założeniu o niezmienności pozostałych warunków) atrakcyjność pracy w porównaniu do nielegalnych form uzyskiwania dochodu. Rosnący poziom zatrudnienia zmniejsza więc wskaźnik przestępczości⁵⁷.

Przyjęliśmy, że w analizie kosztów i korzyści pomijając będziemy preferencje nieakceptowalne społecznie (np. korzyści związane z popełnianiem przestępstw). Obniżenia wskaźnika przestępczości nie traktujemy więc jako transferu korzyści ze strony potencjalnych przestępców (którzy ponoszą koszty alternatywne zaniechania przestępstwa) w kierunku potencjalnych ofiar przestępstw. Dlatego też, spadek przestępczości traktowaliśmy jako wzrost nadwyżki konsumenta⁵⁸.

Dane empiryczne wskazują, że zależność pomiędzy skłonnością do naruszania prawa a poziomem dochodów jest silniejsza w grupach mniej zamożnych (Doyle i in., 1999; Hansen i Machin, 2002). Oznacza to, że zależność między przestępczością a dochodami może nie być liniowa. Krańcowy wpływ wynagrodzenia z pracy będzie wówczas największy wśród osób o najniższym wyjściowym poziomie dochodów.

W związku z powyższym badania statystyczne koncentrują się na segmencie niskopłatnych miejsc pracy. Wysokość wynagrodzenia uwzględniania jest jako główna zmienna objaśniająca. Stosuje się dodatkowo modele wieloczynnikowe obejmujące takie zmienne jak: poziom wykształcenia, miejsce zamieszkania i wysokość korzyści związanych z popełnieniem przestępstwa.

Niektórzy socjologowie (np.: Felson (1998) i Hirischi (1969)) twierdzą, że skłonność do popełnienia przestępstw może być warunkowana także innymi, nieekonomicznymi czynnikami. W tym kontekście wskazują oni następujące kryminogenne efekty bezrobocia – mniejszą rutynizację codziennych aktywności, słabszą strukturyzację czasu oraz beczynność i frustrację (Regea i in. 2009).

⁵⁶ Potencjalni przestępcy biorą pod uwagę fakt, że działalność przestępcza może zmniejszyć ich korzyści z tytułu pracy legalnej w przyszłości. Nie bierzemy jednak pod uwagę tego typu dynamicznych skutków oczekiwań. Zakładamy bowiem, że osoby, w przypadku których popełnienie przestępstwa jest prawdopodobne (tj. osoby o niskich kwalifikacjach i wynagrodzeniu) przeceniają korzyści, jakie mogą odnieść w przyszłości.

⁵⁷ Jest to zgodne z modelem prognozowania przestępczości Home Office (brytyjskiego departamentu spraw wewnętrznych) (Field, 1999). Stosuje się w nim wskaźnik konsumpcji, który umożliwia obliczenie kosztów z tytułu utraty korzyści wiążących się z popełnieniem przestępstwa i który jest ściśle skorelowany z dochodem. W celu uwzględnienia prawdopodobieństwa wykrycia przestępstwa Home Office bierze również pod uwagę wskaźnik liczby osadzonych i zwiększenie liczebności oficerów policji. Zmienne te umożliwiają nam zrozumienie charakteru kompromisów pomiędzy oczekiwanymi korzyściami z pracy i tymi z popełnienia przestępstwa.

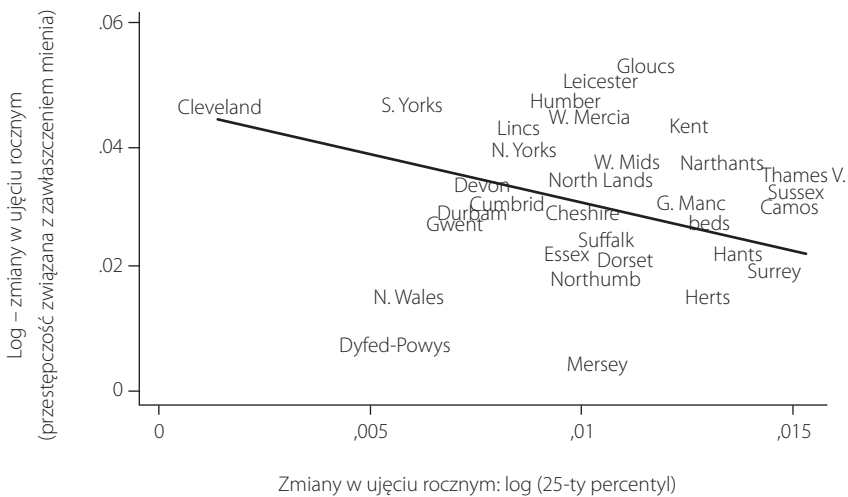
⁵⁸ W niektórych analizach CBA zyski z ograniczenia przestępczości dla społeczeństwa równoważone są utraconymi korzyściami finansowymi dla przestępców (Boardman i in., 2001).

Wpływ programów prozatrudnieniowych na poziom przestępczości w dotychczasowych analizach kosztów i korzyści

Mimo że wyniki badań dotyczących powiązań pomiędzy bezrobociem i przestępczością nie dostarczają rozstrzygających dowodów, istnieją mocne przesłanki sugerujące negatywną korelację między dochodami i przestępczością.⁵⁹ (Porównaj: Machin i Meghir (2000); Gould i in. (1998); Hansen i Machin (2002); Doyle i in. (1999) oraz Grogger (1997)).

Machin i Meghir (2000) badali powiązanie przestępstw przeciwko mieniu z sytuacją na rynku pracy w segmentach zajmowanych przez pracowników o niskich kwalifikacjach. Badanie dotyczyło Anglii i Walii, a analizowane dane pochodziły z okresu od 1970 r. do połowy lat dziewięćdziesiątych XX w. Autorzy dowiedli, że zwiększenie stawki godzinowej najgorzej zarabiającej części społeczeństwa (pierwszego kwartylu dochodowego) wyraźnie obniża poziom przestępczości. Zjawisko to zademonstrowano na Ryc. 6.

Ryc. 6. Zmiany w wysokości wskaźników przestępczości związanej z zawłaszczeniem mienia i wysokości wynagrodzenia w różnych okręgach działalności policji w Anglii i Walii, lata: 1975-1996



Uwaga: Dane zaczerpnięto z publikacji Machina i Meghira (2000)

Hansen i Machin (2002) badają wpływ wprowadzenia płacy minimalnej na wskaźniki przestępczości w Anglii i Walii. Przepisy te przyniosły korzyści znacznej liczbie pracowników uzyskujących niski poziom dochodów.

Badacze stwierdzili, że spadek wszystkich form przestępczości był najbardziej widoczny w obszarach zamieszkałych przez dużą liczbę osób wykonujących prace niskopłatne. Wprowadzenie płacy minimalnej podwyższyło ich dochody, zmniejszając tym samym motywację do popełniania przestępstw. Doprowadziło to do obniżenia lokalnych poziomów przestępczości. Wnioski z tego badania są na tyle mocne, że pozwalają na kontrolowanie wpływu innych czynników warunkujących przestępczość.

Grogger (1997) zbadał powiązania pomiędzy wynagrodzeniem i popełnianiem przestępstw przeciwko mieniu popełnianych przez młodych mężczyzn w Stanach Zjednoczonych. Wykorzystał on indywidualny model probitowy i stwierdził, że 10-procentowy wzrost wynagrodzeń zmniejsza liczbę popełnianych

⁵⁹ Dowody te odnoszą się do poziomu dochodów i zmian w dochodach.

przestępstw o 1,8 punktu procentowego. Również ten wniosek ma wystarczająco mocne podstawy, aby mógł być podstawą dla kontroli wpływu innych determinant przestępczości. Po uwzględnieniu początkowo oszacowanej skłonności do popełnienia przestępstw w grupie analizowanej przez Groggera, dziesięcioprocentowy wzrost wynagrodzeń przekłada się na sześcioprocentowy spadek liczby przestępstw w przeliczeniu na jednostkę.

Gould i in. (1998) badali zależność pomiędzy zmianami we wskaźnikach przestępczości a dostępnością zatrudnienia (*labour market opportunities*) w grupach cechujących się największym prawdopodobieństwem popełnienia przestępstw (młode osoby o niskich kwalifikacjach, zdolne do podejmowania wyłącznie prac niskopłatnych). Autorzy wykorzystali dane pochodzące z ogólnokrajowego badania panelowego prowadzonego przez okres 17 lat. W badaniu kontrolowano wpływ efektów stałych (*country fixed-effects*), ogólny poziom zamożności oraz czynniki demograficzne.

Powyższe wyniki są zgodne z wnioskami pochodzącymi z badania Groggera, w którym wykorzystano model probitowy. W obu przypadkach, zarówno z teoretycznego, jak i empirycznego punktu widzenia, wynagrodzenie okazuje się bardziej adekwatną, niż poziom bezrobocia, miarą szacowania kosztów alternatywnych popełnienia przestępstwa. Obserwacja ta tłumaczy, dlaczego poprzednie badania opierające się wyłącznie na parametrach rynku pracy dostarczały sprzecznych wyników.

Największy wpływ na skłonność do popełnienia przestępstw ma więc wysokość dochodów uzyskiwanych z legalnych źródeł. Stopa bezrobocia sama w sobie nie wpływa bezpośrednio na poziom przestępczości.

W naszych analizach ograniczyliśmy się wyłącznie do przestępstw przeciwko mieniu. Może to wpływać na zaniżanie szacunków redukcji kosztów przestępczości. Proponujemy, by analizy wpływu zatrudnienia na przestępczość oparte były na podejściu motywacyjnym, zgodnie z którym wyższe przychody zmniejszają skłonność do popełnienia przestępstwa. Badania przeprowadzone do tej pory konsekwentnie wykazały negatywną relację między dochodami a wskaźnikami przestępczości. Brak jest jednoznacznych dowodów na to, że bezrobocie jest skorelowane ze wskaźnikami przestępczości w sposób bezpośredni.

3.3. Efekt mnożnikowy w analizie kosztów i korzyści

Wpływ interwencji publicznej może rozprzestrzeniać się na inne sektory gospodarki niż te do których interwencja jest bezpośrednio kierowana. Skutki pośrednie, o których mowa, określa się terminem „efektu mnożnikowego”. W przypadku programów wspierania zatrudnienia występują dwa rodzaje efektów mnożnikowych.

Pierwszym z nich jest mnożnik ekonomiczny, który oznacza wpływ wydatków nowozatrudnionych uczestników programu na gospodarkę. Wydatki te stanowią przychody przedsiębiorstw i poprzez reakcję łańcuchową mogą skutkować zwiększeniem zatrudnienia i dochodów w innych sektorach gospodarki.

Drugi rodzaj efektu mnożnikowego stanowi tak zwany „mnożnik keynesowski”. W tym przypadku również zachodzi efekt domina (reakcja łańcuchowa). Różnica polega jednak na źródle początkowego wzrostu wydatków. Mnożnik keynesowski odnosi się do wydatków związanych z wdrażaniem interwencji publicznej. Są to więc wydatki finansowane z wynagrodzeń wypłacanych osobom i firmom zaangażowanym w obsługę programu, tj. pracownikom Departamentu Pracy i Świadczeń, agencji rządowej ds. zatrudnienia – Jobcentre Plus oraz prywatnym podwykonawcom i dostawcom usług (np. firmom szkoleniowym, dostawcom usług cateringowych).

Jeśli faktycznie zachodzą powyższe efekty mnożnikowe, powinny być one uwzględniane w analizie kosztów i korzyści. Wynagrodzenia beneficjentów podejmujących zatrudnienie stanowią wówczas tylko jedną składową ekonomicznych korzyści interwencji publicznej.

Związek z programami na rzecz zatrudnienia

Istnieją mocne argumenty przemawiające za uwzględnieniem obu efektów mnożnikowych w ewaluacji programów wdrażanych przez Departament Pracy i Świadczeń⁶⁰.

Należy jednak wskazać kilka istotnych zastrzeżeń:

W przewodniku dotyczącym zasady dodatkowości (*Additionality Guide*) nie przedstawiono w sposób czytelny punktów odniesienia, uzasadniających oszacowanie efektów mnożnikowych na sugerowanym poziomie. Uważamy jednak, że istnieją mocne przesłanki, aby uznać zakres 1,3 – 1,7 jako zawyżony. Po pierwsze autorzy opracowania nie wyjaśniają w jakim stopniu efekty mnożnikowe odzwierciedlają dodatkowy wpływ analizowanych interwencji. Można racjonalnie uargumentować, że obserwowane efekty mnożnikowe mogłyby zostać osiągnięte również dzięki alternatywnym formom wykorzystania środków przeznaczonych na realizację programów publicznych⁶¹. Szacunki mnożnikowe agencji English Partnerships uwzględniają gospodarcze i keynesowskie efekty mnożnikowe. Stoimy na stanowisku, że efekty keynesowskie możemy pomijać. Są one bowiem pochodną każdego działania angażującego wydatki publiczne.

Po drugie ekonomiczne efekty mnożnikowe występują na rynkach wtórnych. Uwzględniając założenia neoklasycznej ekonomii, możemy te efekty pominąć. Korzyści netto na rynkach wtórnych obserwować możemy bowiem tylko wówczas, gdy występują na nich mechanizmy zawodności rynkowej oraz gdy wpływ interwencji publicznych jest znaczący.

Powyższy argument zilustrujemy przywołując teoretyczne podstawy analizy wpływu programów prozatrudnieniowych. W ramach analizy kosztów i korzyści jako rynek pierwotny przyjmuje się rynek pracy. Jako powód do prowadzenia interwencji publicznej możemy przyjąć występowanie zawodności rynkowych (np. nierównego dostępu do informacji). Jak przedstawiono na Ryc. 7 równowaga na rynku pracy jest wówczas zaburzona, a zatrudnienie L1 plasuje się poniżej poziomu równowagi. Osoby skategoryzowane jako L2-L1 chciałyby pracować za aktualną stawkę rynkową (W^*). Dla tych osób marginalne korzyści z podjęcia pracy przewyższają koszty marginalne. W związku z tym wzrost podaży siły roboczej do L2 jako efekt programu doprowadzi do wzrostu nadwyżki w przypadku osób podejmujących pracę. Wartość nadwyżki określa na wykresie obszar powyżej krzywej podaży pracy oraz poniżej rynkowej wartości wynagrodzeń. Powstaje tym samym korzyść na rynku pierwotnym, którą uwzględniamy w analizie jako wzrost dochodów z pracy.

Dodatkowo nowozatrudnieni pracownicy (L2-L1) dzięki większym dochodom zwiększą również swoje wydatki konsumpcyjne co wpłynie z kolei na rynki wtórne⁶². Poniżej przeanalizujemy dwa możliwe scenariusze.

Po pierwsze, jeżeli rynek wtórny nie jest w żaden sposób zaburzony (tzn. ma on charakter idealnie konkurencyjny), występowanie dodatkowych nadwyżek (korzyści społecznych) ograniczone jest do rynku pierwotnego. Firmy na rynkach wtórnych reagują natychmiast na bodźce popytowe ze strony rynku

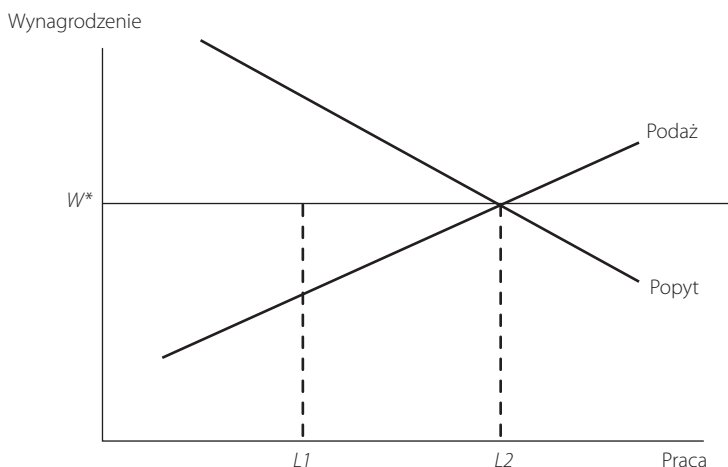
⁶⁰ Znaczna część ich beneficjentów znajduje pracę. Ponadto Departament zatrudnia dużą liczbę osób, które zajmują się bezpośrednio planowaniem i realizacją programów.

⁶¹ Departament Finansów i Administracji, Australia (2006). Podręcznik do analizy kosztów i korzyści (*Handbook of Cost-Benefit Analysis*).

⁶² Ta analiza może być stosowana w odniesieniu do programów zatrudnienia o charakterze popytowym. Nie zostało to uwzględnione w niniejszym dokumencie, ponieważ rezultat będzie taki sam, jak w przypadku modelu o charakterze podażowym. Programy o charakterze popytowym, np. dotacje, zwiększą popyt siły roboczej oraz, uwzględniając schemat podaży siły roboczej, zwiększą zatrudnienie podobnie jak w przypadku efektu L1-L2 przedstawionego na rysunku.

pierwotnego i utrzymują ceny na poziomie równym kosztom marginalnym⁶³. W związku z tym, nie trzeba uwzględniać żadnego wzrostu nadwyżki producenta. Wszystkie pieniądze wydane przez beneficjentów programu są w istocie jedynie transferami korzyści.

Ryc. 7. Struktura rynku pracy



W ramach bardziej realnego scenariusza, rynki wtórne nie będą jednak spełniały cech rynków idealnie konkurencyjnych. Jednym z powodów występowania zawodności rynkowych może być obecność podmiotów posiadających pozycję monopolistyczną. Wówczas wzrost popytu wywołany efektami programu (zatrudnieniem i wzrostem dochodów beneficjentów) przełoży się na nadwyżkę producenta. Innym możliwym rodzajem zawodności na rynkach wtórnych może być występowanie dużej ilości niewykorzystywanych zasobów (np. w związku z recesją). W takim przypadku wzrost wydatków również może wytworzyć nadwyżki na rynkach wtórnych.

Występowanie nadwyżek na rynkach wtórnych może również wywoływać efekty o charakterze negatywnym. Przykładem takiej sytuacji będzie zakup przez nowozatrudnionego uczestnika programu towarów wywołujących negatywne efekty zewnętrzne⁶⁴. Występowanie negatywnych efektów zewnętrznych oznacza z założenia, że społeczne koszty marginalne przekraczają prywatne koszty marginalne. W związku z tym wzrost popytu na rynku wtórnym wywołuje spadek nadwyżki społecznej. Ponadto część wydatków na rynku wtórnym jest związana z importem i nie należy jej uwzględniać w analizach prowadzonych na szczeblu krajowym.

Podsumowując, na rynkach nieidealnie konkurencyjnych może wystąpić zarówno wzrost, jak i spadek nadwyżek. Ustalenie tego nie jest jednak łatwe bez dokładnych informacji dotyczących wzorców konsumpcyjnych beneficjentów interwencji.

⁶³ Zakładamy, że w sytuacji idealnej konkurencji wszystkie firmy są takie same i mają takie same krzywe kosztów. Takie założenie ma również zastosowanie w przypadku normalnych zysków – minimalny poziom zysku wymagany do utrzymania działalności przez dłuższy czas. Jeżeli ponosi się średnie koszty, otrzymuje się normalne zyski. W przypadku idealnej konkurencji średnie koszty = koszty marginalne = otrzymywana cena, w związku z czym nie występuje żadnego rodzaju wzrost nadwyżki.

⁶⁴ Dotyczy tytoniu lub produktów powodujących zanieczyszczenie.

Efekty występujące na rynkach wtórnych należy z zasady oceniać osobno oraz uwzględniać w analizie kosztów i korzyści. W praktyce jest to bardzo trudne. Rzadko kiedy jesteśmy w stanie dokładnie oszacować zmiany popytu zachodzące na tych rynkach oraz określić czy zmiany generują wzrosty, czy też spadki nadwyżek. Co więcej, w przypadku programów zatrudnienia, zmiany popytu na rynkach wtórnych mogą być niewielkie. Dzieje się tak z dwóch powodów. Po pierwsze, większość beneficjentów podejmuje prace niskopłatne. Po drugie, liczba osób podejmujących zatrudnienie jest stosunkowo niewielka w stosunku do ogółu populacji osób pracujących⁶⁵.

Mając na względzie powyższe ograniczenia, zalecamy w przypadku analizy kosztów i korzyści programów prozatrudnieniowych wyłączenie efektów mnożnikowych.

Daniel Fujiwara jest ekonomistą oraz doktorantem London School of Economics and Political Science. Obecnie zatrudniony w Ministerstwie Obrony Wielkiej Brytanii, pracował również dla Departamentu Pracy i Emerytur (*Department for Work and Pensions* – DWP). W swoich badaniach zajmuje się analizą kosztów i korzyści oraz ewaluacją polityk publicznych. W szczególności: technikami oceny wartości dóbr nierynkowych oraz wykorzystaniem metod eksperymentalnych w analizie kosztów i korzyści. W przeszłości był również zaangażowany w projektowanie ram dla ewaluacji polityk publicznych dla Ministerstwa Finansów w Tanzanii.

Daniel Fujiwara is an economist and a PhD candidate at the London School of Economics and Political Science. His current research focuses on Cost-Benefit Analysis and policy evaluation. In particular, he is examining techniques for valuing non-market goods and the use of experimental methods in Cost-Benefit Analysis. Previously, he was also involved in designing policy evaluation frameworks for the Ministry of Finance, Tanzania. Currently works in Ministry of Defence.

Bibliografia:

- Ballard C.L. (1990), *Marginal Welfare Cost Calculations: Differential Analysis vs. Balanced-Budget Analysis*, Journal of Public Economics, vol. 41, s. 263-276
- Beale N., Nethercott S. (1987), *The Health of Industrial Employees Four Years After Compulsory Redundancy*, Journal of the Royal College of General Practitioners, vol. 37 (302), s. 390-394
- Boardman A., Greenberg D., Vining A., Weimar D. (2001), *Cost Benefit Analysis: Concepts and Practice*, Prentice Hall
- Brent R. (2006), *Applied Cost-Benefit Analysis*, Edward Elgar: UK
- Browning E. (1987), *On the marginal Welfare Cost of Taxation*, American Economic Review, vol. 77, s. 11-23
- Browning E. (1976), *The marginal Cost of Public Funds*, Journal of Political Economy, vol. 84, s. 283-298
- Buggard S.A., Brand J.E., House J.S., (2005), *Causation and Selection in the Relationship of Job Loss to Health in the United States*, Mimeo, University of Michigan
- Campbell H., Bond K. (1997), *The Cost of Public Funds in Australia*, The Economic Record, vol. 73, s. 22-34
- Clark A., Oswald A. (1994), *Unhappiness and Unemployment*, Economic Journal, vol. 104, s. 648-659
- Clark A., Oswald A. (2002), *A Simple Statistical Method for Measuring How Life Events Affect Happiness*, International Journal of Epidemiology, vol. 31, s. 1139-1144
- Dahlby B. (1994), *The Distortionary Effect of Rising Taxes*, w: Robson W., Scarth W. (red.), *Deficit Reduction What Pain, What Gain*, Toronto: C.D. Howe Institute
- *Department for Children Schools and Families (DCSF) (2009)*, Childcare and Early Years Survey of Parents 2008
- Doyle M., Ahmed E., Horn R. (1999), *The Effects of Labour Markets and Income Inequality on Crime: Evidence from Panel Data*, Southern Economic Journal, vol. 65 (4), s. 717-738
- Easterlin R. (1974), *Does Economic Growth Improve the Human Lot? Some Empirical Evidence*, w: David R., Reder R. (red.), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, Nowy Jork: Academic Press

- Fullerton D., Henderson Y. K. (1989), *The Marginal Excess Burden of Different Capital Tax Instruments*, Review of Economics and Statistics, vol. 71, s. 435-442
- Gould E., Weinberg B., Mustard D. (1998), *Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997*, Review of Economics and Statistics, vol. 84 (1), s. 45-61
- Greenberg D., Knight G., Speckesser S., Hevenstone D. (2010), *Improving DWP Assessment of the Relative Costs and Benefits of Employment Programmes*, Department for Work and Pensions Working Paper
- Greenberg D. and Robins P. (2008), *Incorporating Non Market Time into Benefit Cost Analyses of Social Programs: An Application to the Self Sufficiency Project*, Journal of Public Economics, vol. 92, s. 766-794
- Grogger J. (1997), *Market Wages and Youth Crime*, NBER Working Paper 5983
- Grossman M. (1972a), *On The Concept of Health Capital and the Demand for Health*. Journal of Political Economy, vol. 80 (2), s. 223-255
- Grossman M. (1972b), *The Demand for Health: A Theoretical and Empirical Investigation*. National Bureau of Economic Research, Occasional Paper 119, Columbia University Press, Nowy Jork
- Treasury H.M. (2003), *Green Book: Appraisal and Evaluation in Central Government*, HMSO Londyn
- Hansen K. and Machin S. (2002), *Spatial Crime Patterns and the Introduction of the UK Minimum Wage*, Oxford Bulletin of Economics and Statistics, vol. 64, s. 677-697
- Iversen L., Sabroe S., and Damsgaard M.T. (1989), *Hospital Admissions Before and after Shipyard Closure*, British Medical Journal, vol. 299, s. 1073-1076
- Jahoda, M. (1982), *Employment and Unemployment: A Social-Psychological Analysis*, Cambridge University Press
- Johnson, G. E. (1972), *Does Public Service Employment Pay Off?* Unpublished paper, University of Michigan
- Keefe V., Reid P., Ormsby C., Robson B., Purdie G., Baxter J., N.K.I. (2002), *Serious Health Events Following Involuntary Job Loss in New Zealand Meat Processing Workers*, International Journal of Epidemiology, vol. 31, s. 1155-1161
- Kuhn A., Lalive R. and Zweimüller J. (2009), *The Public Health Costs of Job Loss*, Journal of Health Economics, vol. 28 (6), s. 1099-1115
- Levitt S. (2004), *Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do not*, The Journal of Economic Perspectives, vol. 18 (1), s. 163-190
- Machin S. and Meghir C. (2000), *Crime and Economic Incentives*, IFS Working Paper 00/17
- Martikainen P. and Volkonen T. (1996), *Excess Mortality of Unemployed Men and Women during a Period of Rapidly Increasing Unemployment*, The Lancet, vol. 348, s. 909-912
- Mathers C. (1994), *Health Differentials Among Adult Australians Aged 25-64 Years*, Canberra: Australian Institute of Health and Welfare
- Nozal A., Lindeboom M. and Portrait F. (2004), *The Effect Of Work on Mental Health: Does Occupation Matter?*, Health Economics, vol.13 (10), s. 1045-1062
- Regea M., Skardhamar T., Telleb M. and Votrubic M. (2009), *The Effect of Job Loss on Crime: Evidence from Plant Closure Events*, Working Paper
- Roy P., Emond A., Pineault R., Mayer F. and Bodson P. (1987), *La Relation Chômage Santé: Une Etude Prospective*. Infante-Rivard, Marie-Josée Legault et Louise Boileau Santé mentale au Québec, vol. 12 (2)
- Ruggeri G. (1999), *The Marginal Cost of Public Funds in Closed and Small Open Economies*. Fiscal Studies, vol. 20 (1), s. 41-60
- Schiller Bradley R. (1973), *Empirical Studies of Welfare Dependency: A Survey*. Journal of Human Resources, vol. 8, s.19-32
- Sickles R. and Yazbeck A. (1998), *On The Dynamics of Demand for Leisure and the Production of Health*. Journal of Business and Economic Statistics, vol. 16 (2), s. 187-197
- Studnicka M., Studnicka-Benke A., Wögerbauer G., Rastetter D., Wenda R., Gathmann P., Ringel E. (1991), *Psychological Health, Self-Reported Physical Health and Health Service Use*. Social Psychiatry and Psychiatric Epidemiology
- Sullivan D.G. and von Wachter T. (2006), *Mortality, Mass Lay-offs, and Career Outcomes: An Analysis Using Administrative Data*. FRB of Chicago Working Paper
- Winkelmann L. and Winkelmann R. (1998), *Why are the Unemployed so Unhappy? Evidence from Panel Data*, Economica, vol. 65 (257), s. 1-15

Ocena oddziaływania oparta na teorii: zasady i praktyka⁶⁶

Wstęp

Ostatnie lata przyniosły wzrost zainteresowania wykorzystaniem metod ilościowych do pomiaru wpływu programów rozwoju. Programy organizacji takich, jak Poverty Action Lab (J-PAL) i Innovations in Poverty Action (IPA)⁶⁷, portfolio badań finansowanych w ramach inicjatywy Banku Światowego na rzecz oceny wpływu na rozwój (*Development Impact Evaluation Initiative* – DIME) i hiszpańskiego funduszu oceny oddziaływania (*Spanish Impact Evaluation Fund* – SIEF)⁶⁸, oraz środki finansowe udostępnione przez międzynarodową inicjatywę na rzecz oceny oddziaływania (3ie)⁶⁹ wskazują, że w ciągu kolejnych pięciu lat przeprowadzone zostaną setki takich badań, w porównaniu do zaledwie kilku wymienionych w opracowaniach z ostatnich lat (np. Centrum Globalnego Rozwoju, 2006). Jednak większość z tych, którzy wspierają realizację rygorystycznej oceny oddziaływania podkreśla, że prowadzone ewaluacje powinny pomagać zrozumieć nie tylko, czy oceniane działania przynoszą zamierzony skutek, ale także dlaczego tak się dzieje. Wiedzy takiej nie można uzyskać ograniczając się jedynie do ustalania i raportowania średniego efektu ocenianej interwencji. Stąd oświadczenie przedstawione przez *Network of Networks of Impact Evaluation* – NONIE: „Zastosowanie podejścia opartego na teorii oznacza, że dobrze zaprojektowana ewaluacja oddziaływania obejmuje zarówno pytania dotyczące procesu, jak i wielkości uzyskanych efektów. Trafność danej polityki jest uwydatniona, jeśli badanie nie ogranicza się tylko do wskazania, czy oceniana interwencja wywarła zamierzony wpływ, ale odpowiada również na pytanie, dlaczego miało – albo dlaczego nie miało – to miejsce” (NONIE, brak daty). Podobnie przewodnik 3ie na temat praktyk oceny oddziaływania stwierdza, że „badania powinny jasno wskazywać, w jaki sposób interwencja (w postaci określonych nakładów) wywarła wpływ na ostateczne efekty. Badania powinny również testować wszystkie założenia przyjmowane dla danej interwencji w odniesieniu do relacji pomiędzy nakładami a efektami (czasem określane jako teoria programu). Projekt ewaluacji powinien obejmować analizę łańcucha przyczynowego poczynając od nakładów, a kończąc na oddziaływaniu” (3ie, brak daty).

Podejście propagowane w niniejszym artykule, a służące zrozumieniu, dlaczego dany program wywarł (lub nie) określony wpływ, określane jest tu jako ocena oddziaływania oparta na teorii (*Theory Based Impact Evaluation* – TBIE). Nie jest to nowe podejście. Ewaluacja oparta na teorii, oznaczająca badanie założeń leżących u podstaw łańcucha przyczynowego od nakładów po rezultaty i wpływ, jest dobrze ugrun-

⁶⁶ Niniejszy dokument roboczy nr 3 został napisany w czerwcu 2009 roku przez dr. Howarda White'a, dyrektora wykonawczego 3ie w ramach *Międzynarodowej inicjatywy na rzecz oceny oddziaływania*; © 3ie, 2009

Międzynarodowa Inicjatywa na rzecz oceny oddziaływania (3ie) ma na celu poprawę warunków życia w krajach rozwijających się poprzez wspieranie tworzenia i wykorzystania dowodów na to, które rozwiązania stosowane w rozmaitych programach rozwoju są rzeczywiście skuteczne, a w szczególności ustalanie, co, kiedy, dlaczego i za ile działa. 3ie jest nową inicjatywą, która odpowiada zapotrzebowaniu na coraz lepsze dowody na sprawdzone rozwiązania i która zwiększy efektywność rozwoju poprzez wspieranie bardziej świadomych polityk publicznych. 3ie finansuje wysokiej jakości oceny oddziaływania i kampanie na rzecz lepiej zaprojektowanych programów i polityk w krajach rozwijających się.

⁶⁷ Zob. odpowiednio www.povertyactionlab.org i <http://poverty-action.org>

⁶⁸ Zob. odpowiednio www.worldbank.org/dime i www.worldbank.org/sief

⁶⁹ Zob. www.3ieimpact.org

towanym podejściem (zob. np. Weiss 1998, oraz Carvalho i White, 2004, przy zastosowaniu do kwestii rozwojowych). Opracowania w zakresie teorii programu od dawna stosowane były przez niektórych badaczy stosujących podejście eksperymentalne i quasi-eksperymentalne, jako sposób wyjaśniania otrzymanych wyników (Blackman i Reich, 2009). W swojej rozprawie na temat możliwych projektów oceny oddziaływania w odniesieniu do interwencji w zakresie rozwoju Rogers (2009) zauważa, że podejście oparte na teorii byłoby w zasadzie właściwe w każdym przypadku.

Chociaż ewaluacja oddziaływania oparta na teorii znajduje powszechne uznanie, to faktycznie wykorzystywana jest w niewielu badaniach. Niniejszy dokument ma na celu uzupełnienie tej luki przez wskazanie kroków, czy też zasad będących podstawą do przeprowadzania ewaluacji opartej na teorii. Na początku części 2 podaję jako przykład Zintegrowany Projekt Odżywiania w Bangladeszu (BINP), do którego następnie odwołuję się w kolejnych przykładach, w celu zilustrowania zasad omówionych w części 3. W części 4 krótko porównuję TBIE z podejściem tzw. „czarnej skrzynki”, a w części 5 podsumowuję tę kwestię.

Przykład – Zintegrowany Projekt Odżywiania w Bangladeszu

Niniejsza część zawiera krótki przegląd ewaluacji Projektu Zintegrowanego Odżywiania w Bangladeszu (*Bangladesh Integrated Nutrition Project* – BINP). Przypadek ten jest następnie wykorzystywany do zilustrowania zasad leżących u podstaw TBIE omówionych w kolejnej części. Szersze omówienie tego projektu można znaleźć w opracowaniach: Banku Światowego (2005), White i Masset (2006) i White (2005).

Projekt BINP – wzorowany na cieszącym się uznaniem Zintegrowanym Projekcie Odżywiania w Tamil Nadu (*Tamil Nadu Integrated Nutrition Project* – TINP) w Indiach – był projektem monitorowania wzrostu u dzieci. Dzieci uczestniczące w projekcie były ważone co tydzień w lokalnej stacji pomiarów, w której zatrudnione były kobiety ze wsi wyszkolone do wykonywania zadań specjalisty ds. żywienia społeczności. Każdorazowo waga dziecka była zestawiana z jego wiekiem na tzw. wykresie wzrostu. Programem objęto dzieci, które rosły zbyt wolno (wzrost niepewny) lub które znalazły się znacznie poniżej przyjętej normy (niedożywione). Na program składało się zarówno doradztwo w zakresie odżywiania oraz samo dożywianie. Podkreślić jednak należy, że z dokumentów projektowych jasno wynikało, że główny wpływ programu miał zostać osiągnięty poprzez element doradztwa. Uzasadnieniem dla tej logiki było założenie, że to niewiedza, bardziej niż samo ubóstwo, przyczynia się do złego odżywiania. Argument ten poparto danymi wskazującymi na niedożywienie, nawet w najbogatszym kwintylu populacji, jak również występującymi w społeczeństwie przekonaniami, takimi jak „celowe niedojadanie”, tj. pogląd, wg którego kobieta powinna jeść mniej w czasie ciąży. Program obejmował również doradztwo w zakresie odżywiania i dożywianie dla kobiet będących w ciąży. BINP był programem pilotażowym, kontynuowanym następnie przez Narodowy Program Żywienia (*National Nutrition Programme* – NNP).

BINP początkowo uznano za sukces. Dane z monitoringu wykazały znaczny spadek przypadków niedożywienia, zwłaszcza skrajnego niedożywienia na obszarach objętych projektem. Na podstawie tych danych, Bank postanowił zwiększyć skalę programu rozpoczynając realizację NNP w połowie trwania BINP i przed dokonaniem jakiegokolwiek ewaluacji projektu. Organizacja Save the Children UK opublikowała krytyczny w swym wydźwięku raport wobec tej decyzji – przedstawiając własne dane z prostego porównania *ex post*, bazującego na porównaniu efektów w grupie objętej wsparciem i dobranej grupie porównawczej. W przeprowadzonym badaniu nie stwierdzono różnicy pomiędzy analizowanymi dwoma grupami (Save the Children, 2003).

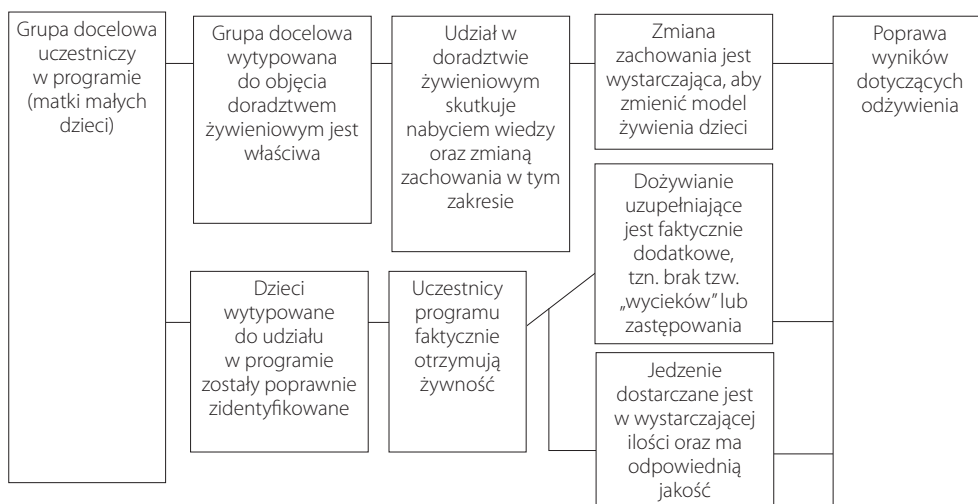
Analizy przeprowadzone przez Departament Oceny Operacji Banku (OED, aktualnie niezależna grupa ds. oceny, IEG), wykorzystywały technikę statystyczną *propensity score matching*, łącząc dane z obszaru projektu (dla beneficjentów wsparcia) z danymi ogólnokrajowego badania dotyczącego odżywiania

przeprowadzonego przez Helen Keller International, wykorzystanymi w celu stworzenia grupy kontrolnej. Analiza ta nie wykazała istotnego wpływu programu na poziom odżywienia dzieci, chociaż stwierdziła pozytywny efekt w przypadku grupy najbardziej niedożywionych.

W konstrukcji program BINP znalazło się wiele założeń na temat mechanizmu przyczynowego stojącego za uzyskiwaniem pozytywnych efektów w obszarze odżywiania. Część z nich przedstawiono na Ryc. 1.

Pierwsza kwestia dotyczy tego, czy ludzie rzeczywiście są poinformowani o programie, a następnie biorą w nim udział – wiele projektów rozwojowych nie jest w stanie poradzić sobie z przeszkodami w tym zakresie, ponieważ nie podejmuje się wystarczających starań, aby wyjaśnić grupie docelowej sensu interwencji lub dokonać racjonalnej oceny kosztów i korzyści z punktu widzenia beneficjentów wsparcia. Jednak BINP dobrze sobie poradził w tym zakresie: około 90 procent kobiet uprawnionych do udziału w programie przystąpiło do niego wraz z dziećmi, choć zdarzały się, opisane dalej, wyjątki od tej sytuacji.

Ryc. 1. Łańcuch przyczynowy dla projektu odżywiania: doradztwo w zakresie odżywiania i dożywiania



Po drugie jednostki wskazane do objęcia wsparciem muszą zostać wytypowane prawidłowo. Program był skierowany do matek małych dzieci. Jednak, jak się okazuje, w wielu przypadkach to nie matki podejmują decyzje w zakresie zdrowia i odżywiania własnych dzieci. Ma to kilka przyczyn. Po pierwsze kobiety w Bangladeszu nie chodzą na targ – to mężczyźni są tymi, którzy robią zakupy. Ponadto w przypadku kobiet mieszkających we wspólnych gospodarstwach domowych – tzn. mieszkających razem z teściową, to teściowa prowadzi dom. W praktyce, poziom udziału w projekcie w przypadku kobiet mieszkających razem z teściową, w bardziej konserwatywnej części kraju, jest znacznie niższy.

Po przeprowadzeniu ważenia dzieci, typowano te, które powinny zostać przyjęte do programu, tj. te z zaburzeniami wzrostu lub niedożywione. Jednak dane pokazują znaczące niedopasowanie będące konsekwencją wystąpienia błędów I rodzaju (dzieci nie włączone do programu, mimo że powinny w nim uczestniczyć) oraz II rodzaju (udział w programie dzieci, które nie powinny brać w nim udziału). W ramach przeprowadzonej ewaluacji przebadano specjalistów odpowiedzialnych za rekrutację do programu, wykorzystując przykładowe wykresy wzrostu (te same, które były używane podczas rekrutacji). Okazało się, że większość specjalistów nie potrafiła prawidłowo zidentyfikować dzieci, które powinny zostać włączone do programu, czego skutkiem było zaobserwowane niedopasowanie uczestników do zakładanej grupy docelowej. Miało to duży wpływ na oddziaływanie programu. W badaniu stwierdzono, że najbardziej nie-

dożywione dzieci skorzystały na udziale w programie, tak więc przeciętny efekt programu byłby większy, jeśli program koncentrowałby się właśnie na tej grupie. Jednak w rzeczywistości część zasobów programu była kierowana do dzieci, które z tego nie korzystały.

Ponadto, żeby dodatkowe żywienie miało korzystny wpływ, musi być faktycznie dodatkowe, podczas gdy w rzeczywistości miały miejsce przypadki „wyciekania” jedzenia (przekazywania jedzenia osobom, dla których nie było ono przeznaczone, dotyczyło to szczególnie posiłków wydawanych kobietom w ciąży) oraz przypadki zastępowania (tj. wydawania żywności w miejsce posiłku, który w przypadku braku dokarmiania byłby i tak dostępny).

Wracając do kwestii zmiany zachowań, udało się nawiązać komunikację, ale nie powiodło się uzyskanie pożądanego zmiany zachowań. To znaczy kobiety, które uczestniczyły w programie miały później znacznie lepszą wiedzę o „dobrych praktykach” w zakresie żywienia. Jednak odnotowano znaczącą lukę pomiędzy wiedzą a praktyką: wiele kobiet zwyczajnie nie korzystało ze zdobytej wiedzy. Powodem były m.in. ograniczone zasoby: w przypadku kobiet w uboższych gospodarstwach mniej prawdopodobne było, że podczas ciąży będą one jeść więcej, a w przypadku kobiet w gospodarstwach domowych dysponujących gruntami lub kobiet żyjących wspólnie ze starszymi krewnymi mniej prawdopodobne było, że w czasie ciąży będą częściej odpoczywać. Sytuacja taka miała miejsce również w przypadku, gdy kobieta mieszkała z teściową. Uczestniczki jednego ze zogniskowanych wywiadów grupowych (FGI) powiedziały badaczom wprost, że „kiedy nasze teściowe umrą wtedy być może będziemy robić te rzeczy, o których mówicie, ale do tego czasu będziemy postępować w tradycyjny sposób”. Wreszcie, mało prawdopodobne było, aby niektóre zmiany zachowania – zwłaszcza te mające na celu wyższy przyrost masy ciała podczas ciąży – miały duży wpływ na końcowy wynik dotyczący niskiej wagi narodzonego dziecka (to waga matki przed ciążą ma największe znaczenie).

Podsumowując, wpływ projektu został ograniczony przez słabe i brakujące ogniwa łańcucha przyczynowego. Ogólnie rzecz biorąc, projekt nie osiągnął żadnego efektu. Poprawa zaobserwowana w danych monitoringowych w rzeczywistości występowała w całym kraju, dlatego w badaniu Save the Children nie stwierdzono różnicy między obszarem projektu i obszarami kontrolnymi. W rzeczywistości była to tendencja wynikająca ze wzrostu plonów ryżu, wyższych dochodów i spadku cen ryżu, a nie z realizacji BINP.

Niemniej jednak, analiza wskazała na pewne możliwości usprawnienia realizacji programu: (1) włączenie teściowych i mężów do komponentu obejmującego doradztwo na temat odżywiania, (2) zawężenie grup docelowych programu oraz (3) lepsze ukierunkowanie działania poprzez lepsze szkolenie specjalistów ds. żywienia dla społeczności, czy wręcz bardziej selektywną rekrutację samych specjalistów. Jednak przeprowadzone obliczenia wykazały również, że jest to bardzo kosztowna interwencja – taka, którą trudno byłoby podjąć na dużą skalę, przede wszystkim ze względu na ograniczenia w zakresie możliwości sprawnego zarządzania oraz dostępnych zasobów.

Niestety, wnioski płynące z dokonanej oceny nie zostały ostatecznie wzięte pod uwagę. Zespół Banku ds. żywienia był bardzo przywiązany do modelu TINP/BINP (*Tamil Nadu Integrated Nutrition Project/ Bangladesh Integrated Nutrition Project*). Sądzono, że sprawdził się on w przypadku TINP (choć nie przeprowadzono żadnego rygorystycznego badania wg aktualnych standardów), a Bank potwierdził sukces w Bangladeszu, mimo że wniosek ten był podważany. Po pewnych debatach, zespół ds. żywienia Banku sporządził dokument, w którym określano BINP jako sukces bez zastrzeżeń (Bank Światowy, 2006). Opierając się na tym przekonaniu podjęto decyzję, aby realizować NNP w ramach tego samego modelu, co BINP, mimo dowodów z oceny stwierdzających, że model, który sprawdził się w Tamil Nadu może wymagać dostosowania do warunków panujących w Bangladeszu. Trzy lata później NNP (*National Nutrition Programme*) przedwcześnie zamknięto, a planowaną ewaluację wpływu odłożono ze względu na brak oddziaływania programu wynikający z błędów na etapie wdrażania. Raport Banku z zakończenia programu przedstawiał zalecenia dotyczące reform programu niezwykle podobne do tych przedstawionych przez OED dwa lata wcześniej.

Zasady

Sześć podstawowych zasad oceny oddziaływania opartej na teorii obejmuje:

1. Określenie łańcucha przyczynowego (teorii programu),
2. Zrozumienie kontekstu,
3. Przewidywanie zróżnicowania,
4. Przeprowadzenie rygorystycznej oceny oddziaływania za pomocą analizy wiarygodnych stanów kontrfaktycznych,
5. Rygorystyczną analizę faktów,
6. Wykorzystanie zróżnicowanych metod badawczych – „metody mieszane”.

Określenie łańcucha przyczynowego (teorii programu)

Łańcuch przyczynowy łączy nakłady z rezultatami i oddziaływaniem. Oznacza to, że łańcuch przyczynowy odzwierciedla tzw. teorię programu (czy inaczej teorię zmiany), w odniesieniu do sposobu, w jaki interwencja ma wywierać zamierzony wpływ. Taka teoria jest osadzona w tradycyjnej macierzy logicznej, choć ta ostatnia może nie określać wprost wszystkich przyjmowanych założeń, podczas gdy ich weryfikacja, jest centralnym elementem podejścia opartego na teorii.

Podejście odwołujące się do analizy zależności przyczynowych jest zwykle krytykowane za swoją liniowość, czy też jednokierunkowość lub determinizm podejście (patrz: White 2009) w odniesieniu do dyskusji na temat różnych znaczeń pojęcia „liniowy” w dyskursie dotyczącym ewaluacji. Jednak krytyka w tym zakresie nie jest właściwa. Choć może się zdarzyć, że faktycznie osoby zarządzające programem przewidują dość prosty schemat powiązań nakładów z działaniami, produktami, rezultatami oraz efektami, to rolę ewaluacji opartej na teorii jest badanie założeń leżących u podstaw tego łańcucha rozumowania. Jednym z takich powiązań jest to, że obserwowane rezultaty są efektem podjętych działań w ramach projektu i dostarczonych produktów, a nie odwrotnie. Jednak taka odwrócona lub dwukierunkowa przyczynowość znajduje się w centrum debaty na temat oceny oddziaływania: obciążenie selekcyjne wynikające ze sposobu przypisywania do programu oraz auto-selekcja do programu oznacza dokładnie, że zmienne wynikowe mają wpływ na to, kto uczestniczy w programie, a nie odwrotnie. Na przykład, społeczności o wysokim poziomie kapitału społecznego są bardziej skłonne do ubiegania się o fundusze na programy rozwoju społeczności. Programy te służą budowaniu kapitału społecznego, lecz zwykła obserwacja różnicy ex post w poziomie kapitału społecznego pomiędzy badanymi miejscowościami a grupą kontrolną ma większe szanse na odzwierciedlenie różnicy sprzed realizacji programu niż samego oddziaływania interwencji.

Bardziej zasadną krytyką jest to, że opisywane podejście ma raczej charakter statyczny, podczas gdy interwencje zazwyczaj zmieniają się w czasie – dostosowują do zmieniających warunków i rozwijają. Zasady opisane w dokumentacji projektowej mogą w praktyce mieć niewiele wspólnego z tym, jak dany program jest wdrażany. Może to wynikać z przeprojektowania interwencji lub z liberalnych interpretacji procedur dokonywanych przez zarządzających projektem (*field managers*). W pierwszym przypadku, teoria programu powinna odzwierciedlać nowy projekt, a ewaluacja powinna dokumentować zaistniały proces uczenia. W drugim przypadku, wszelkie rozbieżności między tym, co miało być zrobione i tym, co faktycznie wykonano, stanowią kluczowe pytanie ewaluacyjne: dlaczego pojawiły się takie różnice i jak one wpływają na realizację i efekty programu?

Przykład dostosowania projektu przedstawiono we wspomnianym wcześniej badaniu funduszy społecznych. Inna krytyka inwestycji finansowanych z funduszy społecznych dotyczy braku ich trwałości, co wynika z braku zabezpieczenia środków zapewniających ich obsługę i utrzymanie (*Operation and Maintenance – O&M*). Pierwotnie, w ramach funduszy społecznych powoływano centralną komisję, która zatwierdzała wszystkie działania, przy założeniu, że poszczególne resorty, przez swoją obecność w tej komisji, były

zobowiązane do pokrycia kosztów operacyjnych, jeśli zgodziły się na realizację projektu. Jednak ten system nie sprawdził się, więc fundusze społeczne zaczęły zawierać umowy kompleksowe, obejmujące wszystkie projekty dla poszczególnych ministerstw. System ten miał również wady, więc w ramach niektórych funduszy społecznych podpisywano umowy z ministerstwami indywidualnie dla każdego projektu, inne wymagały przygotowania lokalnych planów zrównoważonego rozwoju, natomiast w przypadku jeszcze innych odkładano środki na fundusz utrzymania (patrz Bank Światowy, 2002).

Teoria programu powinna być dynamiczna, tzn. powinna umożliwiać „naukę z terenu”, co jest wyrazem potrzeby konfrontacji teorii z danymi. W modelowym podejściu w analizie statystycznej przyjmuje się model jako dany, a następnie ocenia poziom dopasowania danych do tego modelu – badacze mają różne sposoby dbania o to, żeby dane były dopasowane do modelu, tak jak w stwierdzeniu Coase’a, że dane powiedzą wszystko, jeśli będziemy męczyć je wystarczająco długo (cytat Leamer, 1983). Jednakże analiza statystyczna pozwala danym poprowadzić teorię bazując na obserwowanych zależnościach i wzorach. Takie podejście wydaje się nieuporządkowane, ale oczywiście żadne ćwiczenie statystyczne nie może być pozbawione teorii, ponieważ to teoria określa, które dane są gromadzone i analizowane w pierwszej kolejności. To raczej teoria powinna być gotowa do dostosowania się do nieoczekiwanych danych i informacji. Takie podejście może wydawać się podobne do eksploracji danych typu *data mining*, ale w rzeczywistości jest zupełnie inaczej. Badacz eksplorujący dane w *data mining* wie, czego szuka i przegląda dane do czasu, aż to znajdzie. Z kolei w ewaluacji, analityk patrzy na dane oczekując na pojawienie się pewnych wzorców, zarówno tych spodziewanych, jak i niespodziewanych (zob. Mukherjee i in., 1998, rozdział 1).

Inną potencjalnie ważną krytyką tego podejścia jest to, że koncentrując się na łańcuchu przyczynowym, badanie może pominąć występowanie niezamierzonych skutków. Powyższego zagrożenia można uniknąć na dwa sposoby. Po pierwsze, uważne zastosowanie teorii programu może pomóc zidentyfikować również możliwe niezamierzone konsekwencje programu, na przykład skutki dla środowiska, które mogą być lekceważone przez autorów programu. Po drugie, wstępne badania terenowe, w tym np. obserwacje uczestniczące, stanowią ważną część przygotowywanego projektu ewaluacji. Tego typu wstępne badania mogą uchwycić takie niezamierzone skutki, które mogą być następnie włączone w ramy właściwej ewaluacji.

Kwestia niezamierzonych skutków wiąże się także z zagadnieniem „czyją teorię programu chcemy weryfikować?” Poprawny projekt ewaluacji opartej na teorii powinien brać pod uwagę różne konkurencyjne teorie dotyczące działania programu. Zarządzający programem będą mieć własną perspektywę, inną od pracowników terenowych, beneficjentów lub innych interesariuszy. Na przykład, projekty funduszy społecznych (fundusze na rozwój wydatkowane na szczeblu społeczności), jak twierdzą zarządzający programami, mają pozytywny wpływ na rozwój instytucjonalny na poziomie lokalnym i krajowym z uwagi na uczenie się na przykładach (osiągnięcia funduszu socjalnego) i uczenie się przez działanie na obszarach, gdzie instytucje były zaangażowane w realizację funduszu społecznego. Jednak krytycy twierdzili, że fundusze społeczne pomijają istniejące procedury rządowe, podważając je w ten sposób bezpośrednio (przesunięcia pracowników) i pośrednio przez zakłócenie optymalnej alokacji zasobów w poszczególnych ministerstwach. Ewaluacja brała więc pod uwagę zarówno oficjalną teorię programu, jak również konkurencyjną antyteorię (zob. Bank Światowy 2002 – pełne badanie, Carvalho i in. 2002 – podsumowanie, Carvalho i White 2004 – prezentacja zastosowanego podejścia teoretycznego).

Zazwyczaj punktem wyjścia dla odtworzenia teorii programu będzie dokumentacja projektowa. Jeśli istnieją matryce logiczne programu, to będą one ucieleśniać jego teorię. Jednak rzadko kiedy dokumenty projektowe określają wprost wszystkie założenia leżące u podstaw realizowanych działań, choć bywa, że niektóre z nich mogą się pojawiać pod postacią „zagrożeń”. Kolejnym krokiem jest uchwycenie teorii programu przedstawianej przez menedżerów programu. Nawet jeśli wcześniej kwestia ta nie była przedmiotem ich bezpośrednich rozważań, to będą oni posiadać opinię na ten temat. To ćwiczenie jest dobrą

okazją do zaangażowania menedżerów programu w realizację ewaluacji jeszcze na etapie opracowywania jej projektu, co powinno pozytywnie wpłynąć na jej jakość.⁷⁰ Drugim krokiem jest przegląd istniejących badań ewaluacyjnych i literatury naukowej na temat podobnych programów i ich oceny, w których dokonano identyfikacji słabych ogniw w łańcuchu przyczynowym analizowanych działań. Na przykład wspomiane niedopasowanie jest często przywoływanym problemem, zwłaszcza w przypadku programów mikrofinansowych (np. Mosley i Hulme, 1996). Kwestią pełną niuansów są np. mikrokredyty dla kobiet, które w rzeczywistości mogą być wykorzystywane przez męskich członków gospodarstwa domowego, co dalej ma wpływ na ostateczne efekty wsparcia, takie jak np. zdrowie dziecka i żywienie. Kolejne perspektywy wymagające uwzględnienia przy odtwarzaniu teorii programu to te przedstawiane przez pracowników terenowych i beneficjentów. Warto, aby każdy ewaluator zadał sobie pytania „jaki wpływ będzie mieć projekt na przeciętnego mieszkańca wioski? W jaki sposób mieszkańcy mają uzyskać informacje o projekcie? Dlaczego mieliby się zaangażować?”. Warto tego spróbować, antropologia rozwojowa uczy nas, że lokalne postrzeganie projektów może bardzo różnić się od naszych wyobrażeń w tym zakresie. Rozbieżności te mogą wynikać z odmiennych punktów widzenia, potrzeb lub po prostu błędów w komunikacji ze strony personelu projektu.

Zrozumienie kontekstu

Zrozumienie kontekstu jest niezbędne do zrozumienia oddziaływania programu, a więc i dla projektowania ewaluacji. Kontekst oznacza społeczne, polityczne i gospodarcze otoczenie, w którym jest realizowany program. Każdy wymiar może mieć wpływ na to, w jaki sposób przebiega proces przyczynowo-skutkowy. Wpływ takiego samego programu może różnić się w zależności od kontekstu: tak jak miało to miejsce w przypadku udanej realizacji modelu TINP w Indiach, który jednak nie poradził sobie już tak dobrze w Bangladeszu. „Identyczne programy” są czymś idealnym, rzadko spotykanym w warunkach terenowych – co samo w sobie jest ważnym elementem kontekstu. Ponadto, jak wskazano poniżej, zrozumienie kontekstu pomoże w przewidywaniu zróżnicowania oraz w procesie uogólniania.

Zrozumienie kontekstu oznacza uważną lekturę dokumentów projektowych przed przystąpieniem do projektowania ewaluacji, ale również zapoznanie się z szerszą literaturą (antropologia i ekonomia polityczna), o której mowa w części o wykorzystaniu metod mieszanych.

Zrozumienie kontekstu pomaga również w uogólnianiu. Badania wsparcia udzielanego przez Bank Światowy na rzecz podstawowej edukacji w Ghanie oraz zdrowia matki i dziecka w Bangladeszu wykazały, że projekty te odniosły sukces. W przypadku Ghany, odbudowa szkół i dostarczenie podręczników przyczyniły się do zwiększenia liczby uczestników i poprawy wyników nauczania (Bank Światowy, 2004). Dwa ważne aspekty kontekstowe odpowiadały za ten wynik. Po pierwsze po latach kryzysu, system szkolny był w bardzo złym stanie: nie posiadał wystarczającej infrastruktury i brakowało praktycznie wszystkich przyborów szkolnych. Dlatego też renowacja szkół i dostarczenie podręczników wywarły wpływ, jaki nie byłby możliwy, gdyby szkoły funkcjonowały stosunkowo dobrze. Po drugie istniało silne poparcie polityczne dla programu, które pomogło w jego realizacji (program był częścią szerszej reformy edukacji). Zaangażowanie rządu było również kluczowym elementem sukcesu planowania finansowanego ze środków pomocy i doprowadziło do przyspieszenia zmiany demograficznej w Bangladeszu ze znacznym spadkiem umieralności i płodności (Bank Światowy, 2005). Kraj w ciągu dziesięciu lat od odzyskania niepodległości przeszedł drogę od dotkniętego głodem bez praktycznie żadnych udogodnień do posiadania narodowego zdecentrali-

⁷⁰ Najczęściej menedżerowie programów odpowiadają, że nie jest to odpowiedni moment na ocenę programu, ponieważ został on właśnie przeprojektowany, właśnie przeprowadzili własne badania, doszło do zmiany w rządzie, zmiany ministra lub kierownika projektu itp. Te sprzeczny powinny być zwykle grzecznie ignorowane, tak samo jak wszelkie próby wpływania na wyniki. Ale warto sprawdzić, jakie kwestie dotyczące oceny są zdaniem menedżerów programu ważne.

zwanego systemu ochrony zdrowia i planowania rodziny, gdzie środki antykoncepcyjne można zamówić nawet z dostawą do domu. Bez odpowiedniego wsparcia rządu, równie ambitne programy mogą mieć problemy z realizacją.

Przewidywanie zróżnicowania

Zrozumienie kontekstu pomaga przewidzieć ewentualne zróżnicowanie oddziaływania. Oddziaływanie (to jest efekt podjętego działania) może się różnić w zależności od projektu interwencji, charakterystyki beneficjentów lub otoczenia społeczno-ekonomicznego. Odtworzenie teorii programu może pomóc w określeniu możliwej heterogeniczności efektów i pozwolić na odpowiednie zaprojektowanie ewaluacji w celu ich uchwycenia. Przewidywanie heterogeniczności jest ważne z dwóch powodów. Po pierwsze wielkość dobieranych prób badawczych musi odzwierciedlać poziom dezagregacji stosowany w analizie: im większy stopień dezagregacji tym większe wymagane próby (zarówno dla grupy beneficjentów, jak i grupy kontrolnej). Po drugie zasady rachunku prawdopodobieństwa sugerują, że jeśli przeprowadzimy badania wpływu w dwudziestu różnych, dowolnie zdefiniowanych podgrupach, stwierdzimy istotny statystycznie wpływ w jednej z nich przy pięcioprocentowym poziomie istotności. Dobre praktyki, stosowane w eksperymentach zrandomizowanych w badaniach medycznych, wymagają, aby podgrupy do badania zostały zdefiniowane przed przystąpieniem do zbierania danych. Podejście oparte na teorii pomaga w pre-identyfikacji tych grup i dostarcza prawdopodobnego wyjaśnienia dla ewentualnych różnic w oddziaływaniu. Należy mieć jednak na uwadze, że zagadnienie to wymaga znalezienia kompromisu pomiędzy stworzonym modelem, a danymi które należy pozyskać.

Rozważmy programy żywienia dzieci. Przyrost wagi dzieci niedożywionych będzie znacznie wyższy niż w przypadku dzieci dobrze odżywionych, chociaż bardzo źle odżywione dzieci mogą cierpieć na biegunkę, która uniemożliwia skuteczne dokarmianie i w konsekwencji przyrost masy ciała. Lepiej ukierunkowane programy będą zatem skutkować większym oddziaływaniem, a wpływ ten będzie największy w okresie między żniwami – jak to rzeczywiście miało miejsce w przypadku BINP. Młodsze dzieci mogą odnieść więcej korzyści; dzieci, których wzrost uległ zahamowaniu w dzieciństwie nie doświadczą takich korzyści w późniejszych latach. Podobnie, wydaje się, że korzyści poznawcze wynikające z lepszego odżywiania odnoszą głównie dzieci poniżej trzeciego roku życia. Dlatego też, oddziaływanie zależy od wieku beneficjenta oraz wcześniejszego stanu odżywienia; to ostatnie ma charakter sezonowy. Wpływ może się również różnić w zależności od statusu społeczno-ekonomicznego; na przykład substytucja (spożywanie dodatkowych posiłków zamiast istniejącego posiłku) jest bardziej prawdopodobna w ubogich gospodarstwach domowych.

Z tych powodów tendencja w programach żywienia przesunęła się od żywienia w szkołach do dożywiania dzieci w wieku do lat trzech, tak jak w programie w Bangladeszu omówionym powyżej. Jednak należy się spodziewać, że dożywianie w szkołach będzie korzystne z punktu widzenia przyswajania wiedzy. Niedobór kalorii sprawia, że dzieci są zmęczone i apatyczne, więc program żywienia może sprawić, że będą bardziej uważne na zajęciach; z jednym zastrzeżeniem, że większość ludzi staje się senna po dobrym posiłku, tak więc ważna jest również właściwa pora jego podawania. Otoczenie ma również znaczenie dla osiągnięcia korzyści z uczenia się w przypadku bardziej uważnych dzieci. Kluczowym warunkiem dla powodzenia wszystkich interwencji jest właściwe zdiagnozowanie ograniczeń. Uważne dziecko nie odniesie korzyści jeśli nie ma nauczyciela, i prawdopodobnie nauczy się mniej, jeśli nie ma materiałów do nauki. Tak więc wpływ programów żywienia może być większy w dobrze funkcjonujących szkołach niż w słabo wyposażonych, w których również powszechna jest nieobecność nauczycieli. Podobna obserwacja została poczyniona w odniesieniu do warunkowych transferów środków pieniężnych, które zwiększają popyt na edukację, ale niekoniecznie poprawiają wyniki nauczania, czy nawet liczbę zapisów, jeśli istnieją ograniczenia po stronie podaży (Ravallion, 2009).

Innym aspektem heterogeniczności jest możliwa komplementarność pomiędzy interwencjami; na przykład mikrofinanse mają duży wpływ, jeśli towarzyszą im usługi wsparcia biznesu. Czy przeciwnie oba typy wsparcia mogą być względem siebie substytucyjne, tzn. oddziaływanie obu łącznie jest mniejsze niż suma ich dwóch oddzielnie. Badania, które analizują tego typu zagadnienia mają ogromne znaczenie dla polityk publicznych.

Oddziaływanie może również zmieniać się w czasie, pomimo przyjmowanego zazwyczaj założenia (często nie danego wprost) o liniowej trajektorii oddziaływania (Woolcock, 2009). Liniowa trajektoria oddziaływania różni się od wcześniej omówionej krytyki jednokierunkowej przyczynowości lub od statycznego podejścia do ciągle zmieniającej się realizacji programu. Nawet wtedy, gdy projekt programu pozostaje bez zmian i kierunek przyczynowy został ustalony, wpływ interwencji może podlegać zmianom w czasie, przez co wyniki będą bardzo wrażliwe na moment, w którym jest mierzone oddziaływanie. Na przykład, w przypadku projektów, które starają się zwiększać udział oraz wzmacniać grupy marginalizowane, najbardziej prawdopodobnym kształtem oddziaływania w czasie, jak sugeruje literatura, jest krzywa J, czyli sytuacja pogarsza się, zanim ulegnie poprawie. Jest to obszar, który nie został dostatecznie zbadany za pomocą TBIE, jednak nadaje się do analiz w ramach proponowanego podejścia szczególnie dobrze. W przypadku programu BINP wspomnianego wcześniej, program mógł spowodować początkowe konflikty między kobietami i ich mężami i teściowymi, ze względu na wzrost świadomości kobiet, co może tłumaczyć brak oddziaływania w zakresie żywienia. Jednak być może ocena przeprowadzona w dłuższej perspektywie wykryje pozytywny wpływ ze względu na szersze zmiany społeczne, w tym poprawę statusu kobiet wiejskich w Bangladeszu.

Identyfikacja heterogeniczności jest powiązana z zagadnieniem uogólniania. Badania eksperymentalne w Kenii, Południowej Afryce i Ugandzie dotyczyły wpływu obrzezania na przenoszenie HIV/AIDS, stwierdzając, że obrzezani mężczyźni byli mniej skłonni do zarażenia się chorobą (patrz Wawer i in., 2008, Uganda). Jednym z aspektów heterogeniczności był wiek. Po obrzezaniu okres abstynencji powinien trwać jeden miesiąc, tak by rana się zagoiła, uprawianie seksu w tym okresie stanowi wyższe ryzyko. Przeprowadzenie zabiegu u, powiedzmy, 12-letniego chłopca nie niesie ze sobą wysokiego ryzyka w tym okresie. Jednak starsi mężczyźni często nie mogą się powstrzymać od seksu przez okres całego miesiąca, zmniejszając w ten sposób korzystne oddziaływanie zabiegu. Niemniej jednak badania wykazały zmniejszenie ryzyka przenoszenia choroby od 30 do ponad 50 procent w wyniku obrzezania. Ten poziom oddziaływania może być uogólniony na populacje z podobnymi wzorcami zachowań seksualnych. W społeczności, w której mężczyźni praktykują abstynencję, relacje monogamiczne lub stosują prezerwatywy, interwencja nie wywarłaby żadnego wpływu.

Oddziaływanie

Rygorystyczna ocena oddziaływania przy użyciu odpowiedniego stanu kontrfaktycznego jest oczywiście kluczowym elementem TBIE. Stan kontrfaktyczny jest najczęściej szacowany z wykorzystaniem grupy kontrolnej, która musi zostać określona w sposób pozwalający uniknąć obciążenia selekcyjnego, co w praktyce oznacza zastosowanie podejścia eksperymentalnego lub quasi-eksperymentalnego. Posiadanie danych panelowych dodatkowo wzmacnia szacunki oddziaływania, tak więc należy zachęcać do dokonywania oszacowań punktów bazowych w taki sposób, aby mieć możliwość ponownej identyfikacji próby gospodarstw domowych. W przypadku, gdy punkty bazowe nie są dostępne, mogą zostać odtworzone przy użyciu istniejących danych lub zapytań o przeszłość, jednak należy zachować ostrożność w tym drugim przypadku (zob. Bamberger, 2009). Poza obciążeniem selekcyjnym, ważnymi kwestiami do rozważenia przy projektowaniu są możliwości wystąpienia efektów *spill over* (interwencja ma wpływ na grupę kontrolną) i efektu kontaminacji lub zarażenia (inne interwencje mają wpływ na grupę kontrolną).

Analiza kontrfaktyczna oddziaływania musi być uzupełniona rygorystycznymi analizami faktów różnego rodzaju. Wiele z ogniw łańcucha przyczynowego jest opartych na tego typu analizie. W przypadku BINP obejmuje to wskazanie złego ukierunkowania wsparcia oraz przyczyny tego zjawiska, wskazanie problemu „wyciekania” jedzenia, a także faktu, że poprawa wiedzy nie przełożyła się na jej wykorzystanie w praktyce.

Analiza ukierunkowania wsparcia jest najbardziej rozpowszechnionym elementem analizy faktów i powinna być częścią większości, jeśli nie wszystkich, ocen oddziaływania: kluczowe są tu pytania, kto rzeczywiście korzysta z programu? Błędy w zakresie definiowania grupy docelowej czy ukierunkowania wsparcia można określić w sposób ilościowy, a ich źródła zidentyfikować, jak to miało miejsce w Bangladeszu. Analizy tego typu powinny zostać przeprowadzone na różnych poziomach. W przypadku funduszy socjalnych stwierdzono, że użycie map ubóstwa oznacza, iż fundusze społeczne w wielu krajach koncentrowały się na najbiedniejszych dzielnicach, jednak w dzielnicach tych było bardziej prawdopodobne, że to lepiej sytuowane społeczności uzyskują dostęp do zasobów projektu (Bank Światowy, 2002). Natomiast w przypadku elektryfikacji obszarów wiejskich, lepiej sytuowane społeczności miały większą szansę na podłączenie, a biedniejsze gospodarstwa domowe w podłączonych społecznościach pozostawały wciąż bez prądu przez wiele lat ze względu na brak środków na wniesienie opłaty za podłączenie (Bank Światowy, 2008).

Analizy ukierunkowania wsparcia muszą zostać wykonane przy użyciu reprezentatywnego zbioru danych. Z uwagi na problem obciążenia selekcyjnego analizowane zbiory danych mogą nie być reprezentatywne dla całej populacji, a więc nie mogą być używane do odpowiedzi np. na pytanie o to „jaki procent beneficjentów w grupie najbiedniejszych 20 procent skorzystał z projektu?”, chyba że dostępne są wagi dodatkowe dla próby, co może pomóc w otrzymaniu reprezentatywnego zbioru danych.

Drugie zagadnienie, na jakie należy zwrócić uwagę w kontekście problemu ukierunkowania wsparcia, to że jest to dwuwymiarowe ćwiczenie, wymagające zestawienia na wykresie lub w formie tabelarycznej zmiennej udziału z charakterystykami będącymi przedmiotem zainteresowania (cechy te mogą być indywidualne, dotyczyć gospodarstwa domowego lub społeczności). Quasi-eksperymentalne podejście wymaga analizy wielozmiennowej na temat udziału w programie, ale powszechnie uważa się za błąd używanie tych wyników przy analizie ukierunkowania wsparcia, która powinna opierać się raczej na statystyce opisowej. To, czy program dociera do dolnych 20 procent, jest stwierdzeniem opartym na wynikach rozkładu łącznego dwóch zmiennych, a nie na istotności statystycznej kwintyla w regresji wielokrotnej. To, co regresja może zrobić, to wskazać czynniki, które mogą mieć wpływ na udział w projekcie. Na przykład, wielowymiarowa analiza projektu w Indiach może wykazać znacznie niższy udział populacji plemiennych, które są jednymi z najbiedniejszych w niektórych obszarach programu, co wyjaśnia niską wydajność ukierunkowania wsparcia.⁷¹

Przykładem nie w pełni wykorzystywanej formy analizy faktów jest sprawdzenie, czy osoby, które przeszły szkolenia, nauczyły się i zastosowały w praktyce nowe podejście, zgodnie z przyjętymi założeniami programowymi. Badania BINP wykazały, że matki zdobywają nową wiedzę, ale wiele z nich wprowadza jej w życie. Specjaliści ds. żywienia dokonują ważenia dzieci, ale, co najważniejsze, nie nauczyli się poprawnie interpretować wykresów wzrostu. Analiza taka często nie jest w ogóle przeprowadzana, a z pewnością istnieją duże możliwości w tym zakresie. Czy przeszkoleni nauczyciele wiedzą o lepszych

⁷¹ W szczególności czynnik bogactwa jest istotny w przypadku wykluczenia zmiennej „status plemienny”, ale staje się nieistotny, gdy zmienna zostanie włączona do modelu. Stąd też status plemienny, a nie ubóstwo per se, ma wpływ na udział w programie. Takie podejście nie zawsze jest możliwe ze względu na wysoki stopień współliniowości wśród możliwych zmiennych objaśniających, takich jak te wymienione oraz edukacja i lokalizacja.

metodach nauczania i stosują je w praktyce? Badania Banku Światowego w Ghanie sugerują, że wielu tego nie robi.⁷²

Podobnie jak w przypadku Bangladeszu, analiza rzeczowa może wskazać przerwy w ciągu przyczynowym interwencji i w ten sposób wyjaśnić jej niskie oddziaływanie. Inne badania OED wykazały, że szkolenia i wizyty w zakresie doradztwa w rolnictwie w Kenii nie mają wpływu na jego wydajność. Projekt finansował nowe badania w dziedzinie rolnictwa w stacjach badawczych. Wnioski z badań miały być przekazywane pracownikom zewnętrznym, a następnie rolnikom. W praktyce wnioski nie były przekazywane pracownikom zewnętrznym, którzy doradzali rolnikom stosowanie jako nowych takich rozwiązań, które oni znali już od dawna (Bank Światowy, 2000).

Jednak czasem, coś, co wydaje się być wymogiem dla analizy faktów, może w rzeczywistości wymagać stanu kontrfaktycznego. Dotacje dla szkół mają na celu zwiększenie zarówno liczby studentów, jak i polepszenie wyników w nauce. Ale jak to zrobić? Wyjaśnienie z pewnością musi odwoływać się do wykorzystania pieniędzy. Może to wydawać się prostą analizą faktyczną – śledzeniem wykorzystania środków: sprawdzaniem ile środków rzeczywiście szkoły otrzymują i jak są one wydawane, obie kwestie stanowią rzeczywiście użyteczne części badania. Jeśli jednak szkoły posiadają już pewne fundusze, to istnieje możliwość występowania problemu zastępowania. Analizy wzorców wydawania przed i po otrzymaniu środków mogą dostarczyć wiedzy na temat potencjalnej sytuacji kontrfaktycznej w tym przypadku, choć analiza z wykorzystaniem grupy kontrolnej w odniesieniu do ulepszeń i zakupu materiałów może przynieść mocniejsze dowody w tym zakresie.

Wykorzystanie zróżnicowanych metod badawczych – „metody mieszane”

„Metody mieszane” to połączenie podejścia jakościowego i ilościowego w ramach jednej ewaluacji. Wszystkie badania ilościowe w pewnej mierze korzystają z analizy jakościowej – przynajmniej na etapie czytania dokumentacji projektowej – więc jest to raczej kwestia skali.

Wezwanie do stosowania mieszanych metod pochodzi zasadniczo od zwolenników podejścia jakościowego. Jednak w dziedzinie rozwoju do niedawna to właśnie podejście jakościowe dominowało w ewaluacji. Tak więc ważnym krokiem w kierunku stosowania mieszanych metod jest fakt coraz częstszego korzystania z rygorystycznych metod ilościowych w badaniach jakościowych. Jednak tu należy zwrócić uwagę na zwiększenie wykorzystania danych jakościowych w badaniach ilościowych, problem, którym zajmują się bardziej szczegółowo White (2008). Wyodrębnię tu trzy główne punkty.

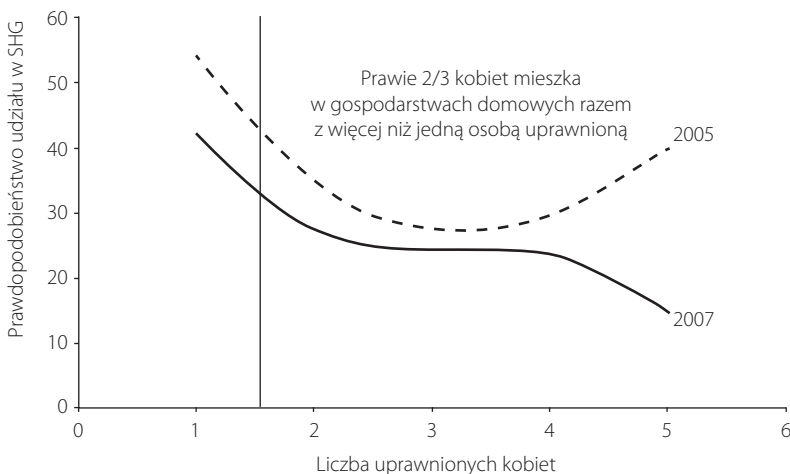
Po pierwsze, wykorzystanie danych jakościowych oznacza szereg działań, nie tylko organizację grup fokusowych (w moim przekonaniu jest to jedna ze słabszych postaci danych jakościowych, chyba że FGI przeprowadzone zostały naprawdę dobrze). Obejmuje ona, na przykład, przegląd literatury antropologicznej i politycznej powiązanej tematycznie z ocenianą interwencją, w celu wykorzystania dostępnej wiedzy jeszcze na etapie tworzenia projektu ewaluacji. W przypadku Bangladeszu, identyfikacja efektu „teściowej” bazowała na wynikach opisanych w literaturze antropologicznej (White, 1992). To spostrzeżenie doprowadziło nas do dokładniejszego przyjrzenia się podziałowi obowiązków domowych w celu identyfikacji kobiet mieszkających z teściowymi, i przeprowadzenia analizy ilościowej z udziałem elementów jakościowych, dających dodatkowy wgląd w to zagadnienie.

⁷² Obserwacja zajęć w szkołach byłaby najlepszym sposobem weryfikacji, czy założone działania są rzeczywiście podejmowane w praktyce, ale zostały wykluczone ze względu na koszt. Można by pomyśleć, że pytając nauczycieli o stosowane metody, otrzymamy obciążone wyniki, ponieważ zgłosiliby wykorzystanie ulepszonych metod, nawet jeśli tak nie jest, jednak w praktyce nauczyciele okazali się zaskakująco szczerzy, albo ich brak wiedzy był na tyle duży, że nie wiedzieli, które odpowiedzi były właściwe.

Zakres technik obejmuje z jednej strony „turystykę rozwojową” (spędzanie całego dnia lub dłużej w terenie), poprzez zestaw narzędzi PRA (*Participatory Rural Appraisal*), aż do osadzania antropologa na obszarze objętym projektem. To ostatnie jest podejściem nie w pełni wykorzystywanym, które mogłoby znaleźć zastosowanie w przypadku badań długookresowych. Drugą kwestią jest to, że mimo wyszydzenia wspomnianej „turystyki rozwojowej”, jest ona istotną częścią TBIE. Naprawdę nic nie zastąpi spędzania czasu w terenie i trudno jest sobie wyobrazić, jak można sensownie analizować dane bez tego typu doświadczenia (widać to, gdy ma to miejsce). Spędzenie choćby tylko kilku dni w terenie, w obszarze realizacji projektu, w tym w różnych warunkach jego realizacji – i najlepiej innych niż te wybrane przez personel projektu – pomoże zarówno w zaprojektowaniu, jak i w późniejszej realizacji badania. Pomocne jest także dotarcie do obszarów nie objętych projektem.

Można by podać wiele przykładów cennych spostrzeżeń z rozmów przeprowadzonych z personelem projektu, beneficjentami i innymi zainteresowanymi stronami w terenie – poniżej podam dwa. Pierwszy z nich pochodzi z ewaluacji projektu dla obszarów wiejskich, który dotyczył pożyczek w ramach grup samopomocy dla kobiet. Jeden człowiek skarżył się, że jego niezamężna córka w wieku 22 lat nie mogła dostać kredytu, ponieważ jego żona już jeden otrzymała. Spostrzeżenie to doprowadziło do wniosku, że mieszkańcy postrzegają kredyt w powiązaniu z całym gospodarstwem domowym, a nie jednostką. Jak się okazało fakt ten doprowadził do znacznie niższego udziału kobiet z gospodarstw domowych, w których mieszkała więcej niż jedna kobieta uprawniona do wzięcia udziału w SHG (*self-help group*) – Ryc. 2. Celem projektu było, by wszystkie uprawnione kobiety w nim uczestniczyły, ale był to cel nierealistyczny tak długo, jak uzyskiwane korzyści postrzegane były z perspektywy całego gospodarstwa domowego, a nie jednostki. Drugi przykład dotyczy tego, jak dobrze dobrany cytat może podkreślić ogólny punkt widzenia. W ramach pracy terenowej dla ewaluacji funduszu socjalnego w Zambii, uderzający był fakt, że wszyscy – od menedżerów do pracowników programu i mieszkańców wsi – mówili, że „społeczność” wybrała projekt, choć było oczywiste, że faktycznie miał miejsce bardziej selektywny proces (patrz White i Vajja, 2008). Jednakże fakt, że „społeczność” była w istocie dość ograniczonym bytem, oznaczającym komisję projektu, został dobrze podsumowany przez regionalnego pracownika programu, który, odbierając telefon, powiedział nam „muszę już iść, mam społeczność w moim biurze”.

Ryc. 2. Poziom uczestnictwa w programie grup samopomocy w Andhra Pradesh wg liczby uprawnionych kobiet w gospodarstwie domowym



Jeśli dostępne dane umożliwiają ocenę oddziaływania bez konieczności ponownego zbierania danych (coś, co z zasady należy wspierać, ponieważ zbyt często przystępujemy do zbierania nowych danych, podczas gdy istniejące zbiory danych nie są w pełni wykorzystywane), istnieje niebezpieczeństwo, że badacze będą prowadzić ocenę oddziaływania bez jej faktycznego powiązania z interwencją. Bardzo prawdopodobne, że takim badaniom brakować będzie trafności, patrząc z punktu widzenia projektowania polityk publicznych. Będzie to wynikiem słabego zrozumienia tego, w jaki sposób interwencja rzeczywiście funkcjonuje.

Wreszcie, budżet powinien pozwalać na pewne działania badawcze, które pozwolą na uzupełnienie niejasności w zgromadzonych danych podczas dodatkowych prac terenowych. Grupy fokusowe dotyczące wyjaśnienia przyczyn luk w zastosowaniu wiedzy w praktyce w Bangladeszu są przykładem takich prac. Inny przykład pochodzi z wyżej wymienionego badania finansowania dla grup samopomocy w Andhra Pradesh w Indiach. Posiadano dane panelowe, a badanie obejmowało standardowy moduł typu LSMS (*Living Standards Measurement Study*) dla gospodarstw domowych. Analiza tych danych wykazała niskie zwroty dla większości przedsiębiorstw, w tym znacząca mniejszość przynosiła straty. Jednak dane w module były zbyt ogólnym narzędziem, aby umożliwić zrozumienie, w jaki sposób analizowane gospodarstwa funkcjonują. Dlatego zlecono, swego rodzaju „ilościową etnografię”, aby ponownie odwiedzić wszystkie gospodarstwa domowe, które zostały przebadane i uznane za przedsiębiorstwa. Podczas powtórnych wizyt wykorzystano częściowo ustrukturyzowany kwestionariusz w celu identyfikacji codziennych przepływów środków pieniężnych przedsiębiorstw i nakładów pracy członków gospodarstw domowych (jak również pracowników, chociaż nie było ich wielu). Wyniki rzeczywiście potwierdziły niski poziom dochodów z tej działalności (20-30 rupii na dzień nie było niczym niezwykłym, w porównaniu do dziennego wynagrodzenia 50-70 rupii) oraz ryzykowny charakter wielu innych działań (zdychanie zwierząt domowych, zwłaszcza kóz i niewystarczająca wielkość rynku).

TBIE wobec podejścia czarnej skrzynki

Ocena oddziaływania oparta na teorii może być przeciwstawiona tzw. „podejściu czarnej skrzynki”. W podejściu tym często po prostu sprawozdawana jest wielkość oddziaływania – dotyczy istotności statystycznej współczynnika średniego efektu interwencji, jednak w ramach podejścia nie próbuje odpowiedzieć się na pytanie dlaczego przyjmuje on taką, a nie inną wartość. Niniejszy artykuł starał się pokazać, jak odnieść się do tego pytania i jakie mogą być z tego korzyści. Jednak niektóre zastrzeżenia wysuwane pod adresem opisywanego podejścia są zasadne.

Krytyki badań, których celem jest dokonanie oszacowania średniego efektu interwencji nie należy wyolbrzymiać. Heterogeniczność ma znaczenie, podobnie, jak zrozumienie kontekstu, w którym konkretne oddziaływanie miało miejsce. Jednak rzadko będzie tak, że sam efekt interwencji nie będzie przedmiotem zainteresowania (zwykle zarówno podjętej interwencji, jak i samego zamiaru interwencji). W rzeczywistości jest bardzo prawdopodobne, że będzie głównym przedmiotem zainteresowania. Mylące byłoby przedstawianie istotności, lub jej braku, efektu dla poszczególnych podgrup, jeżeli ogólny średni efekt interwencji byłby negatywny. Ponadto wartość przeciętnego oddziaływania interwencji jest podstawą do przeprowadzenia analizy efektywności kosztowej.

Po drugie, TBIE analizuje łańcuch przyczynowy na różne sposoby. Stara się rozwickłać różne etapy łańcucha przyczynowego, ale także wskazać, które elementy interwencji sprawdzają się, a które nie. Można tego dokonać poprzez analizę regresji. Na przykład, badanie BINP prezentuje regresje czynników wpływających na lukę między wiedzą a praktyką. Jednak takie podejście, które opiera się na modelach selekcji i parametrycznej specyfikacji badanej relacji, ma wielu krytyków, którzy opowiadają się po stronie metod eksperymentalnych i quasi-eksperymentalnych, takich, jak *propensity score matching* i nieciągłość w równaniu regresji. To rygorystyczne podejście może obejmować analizy tego, które elementy programu są

skuteczne, ale interwencja musi być skonfigurowana tak, by umożliwić zróżnicowanie w poszczególnych grupach objętych wsparciem – np. niektórzy przedsiębiorcy otrzymują kredyty, niektórzy korzystają z usług wsparcia dla biznesu, a niektórzy z obu tych możliwości. W praktyce TBIE będzie łączyć tego typu rygorystyczne oszacowania oddziaływania, tak jak może mieć to miejsce w innych podejściach służących do analizy łańcucha przyczynowego.

Wreszcie, struktura tego, co znajduje się w „czarnej skrzynce” może być tak skomplikowana, że czasami nie warto się w nią zagłębiać. Badania Banku Światowego dotyczące elektryfikacji obszarów wiejskich uwzględniały wpływ elektryfikacji na płodność. Dostęp do energii elektrycznej znacznie zmniejsza płodność (Bank Światowy, 2008). Badanie było w stanie wykazać jedną z możliwych kwestii, która wydawała się być istotna (telewizja zwiększa wiedzę na temat antykoncepcji), i jedną, która nie była istotna („alternatywy dla seksu” zmniejszają aktywność seksualną). Istnieje jednak wiele innych możliwych przyczyn, takich jak wpływ dochodów, innych świadczeń edukacyjnych i tak dalej. W takich przypadkach, gdzie nie można oddzielić wszystkich przyczyn, ograniczenie się do oszacowania oddziaływania interwencji może być najlepszym wyjściem.

Wnioski

Niniejszy dokument popiera wezwanie do większej liczby rygorystycznych badań ilościowych, weryfikujących skuteczność programów wspierających rozwój. Jednak trafność tego typu badań dla polityk publicznych będzie znacznie większa, jeśli wyjaśnią również, dlaczego interwencje okazały się skuteczne lub nie. Powszechnie uważa się, że ewaluacja oddziaływania oparta na teorii (TBIE) może przynieść niezbędną do tego celu insight. Jednak wiele nowych badań nie spełnia obietnicy podejścia opartego na teorii, prowadząc spekulacje co do przyczyn wpływu lub różnic w oddziaływaniu, zamiast przeprowadzić solidną analizę empiryczną w celu ich wyjaśnienia.

Przedstawiłem praktyczne zastosowanie TBIE oraz to, jak podejście prowadzi bezpośrednio do wniosków z zakresu projektowania polityk publicznych, w celu zwiększenia wpływu programu. Rozwiązanie to wymaga zastosowania szeregu zasad, które opisane zostały powyżej. Teoria programu musi zostać opracowana w sposób elastyczny, gotowy do adaptacji do zmieniających się warunków, i do przyswojenia konkurencyjnych teorii, jak również niezamierzonych konsekwencji. Rygor musi zostać włączony zarówno do analizy faktów, jak i analizy stanów kontrfaktycznych, co w praktyce oznacza wykorzystanie metod mieszanych. Teoria programu musi zostać osadzona w kontekście społecznym, politycznym i kulturalnym interwencji, co będzie jednym z elementów podkreślających spodziewaną różnorodność oddziaływania interwencji.

Howard White jest dyrektorem 3ie (*International Initiative for Impact Evaluation* – Międzynarodowa Inicjatywa dla Ewaluacji Oddziaływania), współprzewodniczącym Campbell International Development Coordinating Group oraz profesorem Instytutu Badawczego Alfreda Deakina Geelong University. W swojej karierze kierował programem ewaluacji oddziaływania w ramach Niezależnej Grupy ds. Ewaluacji (Independent Evaluation Group) Banku Światowego oraz licznymi ewaluacjami ponadkrajowymi. Był również kierownikiem dużych projektów, jak np. przygotowanie raportu Banku Światowego *African Poverty at the Millenium* czy rozwój szkoleń nt. ubóstwa dla pracowników Departamentu Międzynarodowego Rozwoju (*Department for International Development* – DFID) na całym świecie. Howard White zajmuje się przede wszystkim zagadnieniami związanymi z rozwojem krajów Afryki i Azji. Opublikował ponad 60 tekstów w międzynarodowych periodykach oraz kilka książek dotyczących redukcji ubóstwa oraz skuteczności

pomocy. Pełni funkcję redaktora naczelnego *Journal of Development Studies* oraz *Journal of Development Effectiveness*. Prowadzi zajęcia w Instytucie Studiów Społecznych The Hague University oraz w Instytucie Studiów Rozwoju na Uniwersytecie Sussex oraz liczne szkolenia i warsztaty dla decydentów politycznych dotyczące skuteczności rozwoju oraz ewaluacji oddziaływania.

Bibliografia

- 3ie guide for grantees (brak daty) '3ie impact evaluation practice: a guide for grantees', <http://www.3ieimpact.org/page.php?pg=overview> (ostatni dostęp z 1 czerwca 2009 r.).
- Bamberger, Michael (2009) 'Strengthening the evaluation of program effectiveness through reconstructing baseline data' *Journal of Development Effectiveness* 1(1): 37-59.
- Blackman, Leonard and Stephanie Reich (2009) 'Randomized control trials: a gold standard with feet of clay?' in Stewart Donaldson, Christina Christie and Melvin Mark (eds.) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* [Thousand Oaks, California: Sage].
- Carvalho, Soniya, Gil Perkins and Howard White (2004) 'Social funds: participation, social capital and sustainability' *Journal of International Development* 14 611-625, 2002.
- Carvalho, Soniya and Howard White (2004) 'Theory-based evaluation: the case of social funds' *American Journal of Evaluation* 25(2) 141-60, 2004.
- Centre for Global Development (2006) *When Will We Ever Learn?* [Washington D.C.: Centre for Global Development].
- International Initiative for Impact Evaluation, 3ie (no date) '3ie Impact Evaluation Practice: a guide for grantees'
- <http://www.3ieimpact.org/doc/3ie%20impact%20evaluation%20practice.pdf> (dostęp z 1 czerwca 2009 r.).
- Leamer, E. (1983), 'Let's take the con out of econometrics', *American Economic Review*, 23 (1), 31-43.
- Mosley, Paul and David Hulme (1996) *Finance Against Poverty* [London: Routledge].
- Mukherjee, Chandan, Marc Wuyts and Howard White (1994) *Econometrics and Data Analysis for Developing Countries* London: Routledge.
- NONIE (brak daty) 'NONIE statement on impact evaluation' <http://www.worldbank.org/ieg/nonie/members.html> (dostęp z 1 czerwca 2009 r.).
- Ravallion, Martin (2009) 'Evaluating three stylized interventions', *Journal of Development Effectiveness* 1(3).
- Rogers, Patricia (2009) 'Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation' *Journal of Development Effectiveness* 1(3).
- Save the Children (2003) *Thin on the Ground. Questioning the evidence behind World Bank-funded community nutrition projects in Bangladesh, Ethiopia and Uganda.* [London: Save the Children UK].
- Weiss, Carol (1998) *Evaluation: methods for studying programs and policies.* Prentice Hall: New York.
- Wawer M, Kigozi G, Serwadda D, et al. Trial of Male Circumcision in HIV+ Men, Rakai, Uganda: Effects in HIV+ Men and in Women Partners. 15th Conference on Retroviruses and Opportunistic Infections; 2008; Boston, MA; 2008.
- White, Howard (2005) 'Comment on Contributions Regarding the Impact of the Bangladesh Integrated Nutrition Project' *Health Policy and Planning* 20(6), 408-411
- White, Howard (2008) 'Of Probits and Participation: the use of mixed methods in quantitative impact evaluation' *IDS Bulletin*, 2008.
- White, Howard (2009) 'Some reflections on current debates in impact evaluation' *3ie Working Paper No. 1* [New Delhi: International Initiative for Impact Evaluation].
- White, Howard and Edoardo Masset (2006) 'The Bangladesh Integrated Nutrition Program: findings from an impact evaluation' *Journal of International Development* 19: 627-652, 2006.
- White, Howard and Anju Vajja (2008) 'Can the World Bank Build Social Capital?: Community Participation in Social Funds in Malawi and Zambia' *Journal of Development Studies* 44(8): 1145-1168.
- White, Sarah (1992) *Arguing with the crocodile: gender and class in Bangladesh*, London: Zed.
- Woolcock, Michael (2009) 'Toward a plurality of methods in project evaluation: a contextualised approach to understanding impact trajectories and efficacy' *Journal of Development Effectiveness* 1(1): 1-14.
- Bank Światowy (2000) *Agricultural extension: the Kenya experience* [Washington D.C.: OED, Bank Światowy].
- Bank Światowy (2002) *Social Funds: assessing effectiveness* [Washington D.C.: OED, Bank Światowy].
- Bank Światowy (2005) *Maintaining Momentum to 2015? An impact evaluation of interventions to improve maternal and child health and nutrition in Bangladesh* [Washington D.C.: OED, Bank Światowy].
- Bank Światowy (2006) *Repositioning Nutrition as Central to Development: a strategy for long term large-scale action* [Washington D.C.: Bank Światowy].
- Bank Światowy (2008) *The welfare impact of rural electrification: a re-assessment of the costs and benefits* [Washington D.C.: IEG, Bank Światowy].

Przykład ewaluacji efektów Aktywnych Polityk Rynku Pracy

Bazą dla niniejszego artykułu są analizy przeprowadzone w trakcie realizacji projektu „Wpływ polityki spójności na poziom i jakość zatrudnienia w Polsce” (2010), których celem był pomiar efektów szkoleń osób bezrobotnych współfinansowanych z Europejskiego Funduszu Społecznego w ramach perspektywy 2004-2006. Były to szkolenia w projektach realizowanych przez Powiatowe Urzędy Pracy, z reguły w ramach Priorytetu 1 Sektorowego Programu Operacyjnego Rozwój Zasobów Ludzkich (pojedyncze projekty były realizowane w ramach Inicjatywy Wspólnotowej EQUAL lub Działania 2.4 Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego). Beneficjentami tego typu projektów były niemal wszystkie Powiatowe Urzędy Pracy, dlatego nie będzie dalekie od prawdy stwierdzenie, że ich łącznym zasięgiem objęty został cały kraj.

Podstawowym celem szkoleń osób bezrobotnych, z punktu widzenia dalszych losów ich uczestników, jest zwiększenie szans na znalezienie pracy. Dlatego jako efekt szkolenia było traktowane zwiększenie prawdopodobieństwa podjęcia pracy przez osoby przeszkolone. Logicznie rzecz biorąc, w celu ustalenia, na ile zwiększyły się szanse znalezienia pracy, należałoby porównać prawdopodobieństwo znalezienia pracy po udziale w szkoleniu z oszacowanym prawdopodobieństwem znalezienia pracy w przypadku braku udziału w szkoleniu. Tego rodzaju sformułowanie problemu odsyła nas do kontrfaktycznych metod oszacowania wpływu interwencji, a zatem podejścia quasi-eksperymentalnego. W związku z tym za wskaźnik efektu szkoleń uznano różnicę w prawdopodobieństwie posiadania pracy między grupą ostatecznych beneficjentów wsparcia a grupą kontrolną, przy czym dobrana przy wykorzystaniu techniki PSM grupa kontrolna miała dostarczyć oszacowania zatrudnienia wśród uczestników szkoleń w kontrfaktycznej (tj. hipotetycznej i sprzecznej z rzeczywistością) sytuacji ich nieuczestniczenia w szkoleniach.

W artykule przedstawiono metodologię badania, przebieg analiz i uzyskane wyniki. Stanowią one grunt do refleksji metodologicznej na temat:

- konsekwencji zastosowanych technik gromadzenia i analizy danych, często mających formę ograniczeń w sferze uzyskiwanych wyników, formułowania wniosków i dokonywania interpretacji,
- środków służących przezwyciężeniu tych ograniczeń, jak również poszerzaniu i pogłębianiu posiadanej wiedzy na temat efektów szkoleń osób bezrobotnych.

Dane źródłowe i zastosowana metodologia

Dane niezbędne do przeprowadzenia analiz we wspomnianym wyżej badaniu zaczerpnięto z baz systemu informatycznego PULS, administrowanych przez Powiatowe Urzędy Pracy (PUP). Ze względów technicznych i logistycznych nie występowało o przekazanie danych do wszystkich PUP, lecz wylosowano próbę obejmującą 96 urzędów pracy. Aby wyniki były reprezentatywne dla całego kraju, zastosowano losowo-warstwowy schemat doboru próby. Warstwy zostały wyróżnione ze względu na województwo, stopę bezrobocia i poziom urbanizacji obszaru działania urzędu (tj. odsetek ludności miejskiej) oraz łączną wartość projektów unijnych realizowanych przez urząd. Niestety, nie od wszystkich wytypowanych w ten sposób Powiatowych Urzędów Pracy udało się pozyskać niezbędne dane. Zdarzało się też, że przekazane bazy zawierały usterki uniemożliwiające ich wykorzystanie w badaniu. Ostatecznie do zasobu danych źród-

dłowych weszły informacje o 1 292 145 osobach zarejestrowanych jako bezrobotne lub poszukujące pracy, pochodzące z 59 Powiatowych Urzędów Pracy. Wśród nich, na podstawie danych z systemu PULS, zidentyfikowano 18 490 osób, które jako bezrobotne ukończyły w okresie od 1 czerwca 2004 r. do 1 czerwca 2008 r. szkolenia współfinansowane ze środków EFS. Wszystkie te osoby zaliczono do próby uczestników (w dalszej części artykułu nazywanej „grupą beneficjentów ostatecznych”), która miała posłużyć do oszacowania efektów szkoleń.

Do tak wyodrębnionej grupy beneficjentów ostatecznych dobrano grupę kontrolną za pomocą metody Propensity Score Matching (PSM) przy zastosowaniu algorytmu doboru najbliższego sąsiada „1 na 1”, bez zwracania. Każdej osobie z grupy beneficjentów ostatecznych przyporządkowano zatem jedną i tylko jej właściwą osobę „kontrolną”. Przed wykonaniem tej operacji podjęto kilka istotnych decyzji metodologicznych. Pierwszą z nich było określenie populacji, z której dobierana była grupa kontrolna. W pierwotnej wersji analiz do populacji tej zaliczono osoby nieuczestniczące w szkoleniach współfinansowanych z EFS i zarazem bezrobotne w okresie, w którym osoby z grupy beneficjentów ostatecznych rozpoczynały udział w szkoleniu. Zastosowano zatem klasyczną logikę eksperymentalną, zgodnie z którą między grupą zasadniczą a kontrolną powinna występować tylko jedna różnica, polegająca na obecności lub braku bodźca w postaci uczestnictwa w szkoleniu współfinansowanym z EFS. Późniejszą modyfikacją tego założenia, polegającą na zawężeniu populacji, z której dokonywano doboru, opisano w dalszej części niniejszego artykułu.

Innym istotnym wyborem metodologicznym było określenie zbiorowości, w ramach których dokonywano doboru grup kontrolnych. Założenia badania przewidywały możliwość dokonywania porównań międzywojewódzkich, dlatego wyodrębnienia grup kontrolnych dokonano w ramach poszczególnych województw. Trzeba bowiem podkreślić, że zasadność dokonywania analiz w rozbiciu na poszczególne kategorie respondentów była uzależniona od algorytmu przeprowadzenia doboru grupy kontrolnej. Metoda PSM pozwala przyporządkować odbiorcy wsparcia osobę podobną pod względem wielu cech wpływających na prawdopodobieństwo wzięcia udziału w projekcie oraz na osiągnięcie pożądanego efektu. Nie można jednak oczekiwać, że będzie to osoba identyczna ze względu na wszystkie te cechy⁷³ (por. Konarski, Kotnarowski, 2007, s. 187; Trzcirski, 2009, s. 27). Oznacza to, że gdyby dokonywać doboru na poziomie całego kraju, grupa kontrolna dla uczestników szkoleń z danego województwa składałaby się nie tylko z mieszkańców tego samego województwa, ale również, choć w mniejszym stopniu, z mieszkańców innych województw. Różnica w efektach między podgrupą beneficjentów ostatecznych a podgrupą kontrolną dla danego województwa mogłaby zatem wynikać nie tylko z uczestnictwa w szkoleniu, ale także z innego składu ze względu na województwo (i prawdopodobnie również inne kontrolowane cechy). Nie pozwoliłoby to interpretować tej różnicy jako efektu szkoleń w określonym województwie. Dlatego doboru grupy kontrolnej dokonano w ramach poszczególnych województw (a więc jako potencjalnych członków grupy kontrolnej traktowano tylko te osoby, które pochodziły z tego samego województwa, co dany uczestnik szkolenia).

Kolejną decyzją był wybór zmiennych wykorzystanych do szacowania *propensity score*, czyli zmiennych, pod względem których skład grupy kontrolnej był zrównywany ze składem grupy beneficjentów ostatecznych. W tym charakterze użyto informacje o 22 cechach, odpowiadających czynnikom, które hipotetycznie wpływają zarówno na prawdopodobieństwo udziału w projektach szkoleniowych, jak i na

⁷³ Takie dopasowanie identycznej osoby było mało prawdopodobne ze względu na samą liczbę zmiennych uwzględnionych w analizie. Strukturę próby kontrolowano ze względu na 22 zmienne. Liczba kombinacji występujących wśród uczestników szkoleń różnych wartości tych 22 zmiennych, czyli potencjalnie możliwych charakterystyk osoby, wynosiła nieco ponad osiem i pół kwadrylionu. Gdyby każda kombinacja wartości cech była równie prawdopodobna, prawdopodobieństwo znalezienia identycznej osoby jak określony odbiorca wsparcia w zbiorze liczącym nieco ponad 1,2 miliona różnych osób wynosiłoby więc około siedmiu trylionowych. Pokazuje to, jak trudne byłoby znalezienie osoby identycznej jak uczestnik szkolenia ze względu na wszystkie uwzględnione zmienne kontrolne.

szanse znalezienia zatrudnienia – albo korelatom takich czynników. W szczególności starano się za pomocą korelatów takich jak odsetek stawień się na wezwanie czy udział w stażach, innych pracach i szkoleniach uchwycić poziom motywacji respondenta do podjęcia pracy (czynnika niewątpliwie ważnego, ale trudnego w pomiarze). W wyniku zastosowania metody PSM uzyskano wysoką zbieżność składu grupy beneficjentów ostatecznych i grupy kontrolnej ze względu na kontrolowane cechy (co obrazuje tabela 1). Będąca miernikiem rozbieżności (obciążenia) wartość bezwzględna standaryzowanej różnicy jedynie w przypadku odsetka stawień się na wezwanie z urzędu przekraczała 2%.

Tab. 1. Dopasowanie grupy kontrolnej do grupy beneficjentów ostatecznych w analizie PSM.

Zmienna	Wartość	Grupa b.o.	Grupa kontrolna	Standaryz. różnica (w %)
Płeć	mężczyzna	51,7%	51,6%	0,3%
Stan cywilny	żonaty	47,6%	47,9%	-0,7%
	separacja	0,2%	0,3%	-0,7%
	wolny	52,0%	51,7%	0,7%
Wykształcenie	brak lub niepełne podstawowe	0,1%	0,1%	0,0%
	podstawowe	15,7%	15,9%	-0,7%
	gimnazjalne	1,2%	1,3%	-0,4%
	średnie zawodowe czteroletnie	2,0%	2,1%	-0,7%
	pomaturalne lub policealne	6,0%	6,3%	-1,2%
	średnie ogólnokształcące	13,5%	13,5%	-0,2%
	średnie zawodowe	24,3%	24,7%	-0,8%
	wyższe (w tym licencjat)	12,2%	11,9%	0,9%
	zasadnicze zawodowe	25,0%	24,1%	1,9%
Niepełnosprawność	tak	2,3%	2,2%	1,3%
Samotne wychowywanie dzieci	tak	6,2%	6,0%	0,8%
Liczba dzieci	(średnia)	0,56	0,56	0,2%
Wiek	18-24	34,3%	34,3%	0,1%
	25-42	49,0%	49,1%	-0,2%
	43-49	11,1%	11,3%	-0,7%
	50-54	4,4%	4,1%	1,3%
	55-64	1,1%	1,2%	-0,1%
Staż pracy ogółem (dni)	(średnia)	2191	2162	1,0%
Wielka grupa zawodowa (KZiS)	0	13,7%	13,8%	-0,3%
	1	0,5%	0,5%	-0,1%
	2	11,3%	11,1%	0,6%
	3	23,1%	23,6%	-1,3%
	4	4,0%	4,1%	-0,4%
	5	11,3%	11,3%	0,0%
	6	2,0%	2,1%	-0,4%
	7	23,6%	23,2%	0,9%
	8	4,5%	4,2%	1,5%
	9	6,0%	6,1%	-0,4%
Liczba wykonywanych zawodów	(średnia)	1,88	1,88	0,1%
Prawo jazdy	posiada	29,0%	29,1%	0,0%
Liczba znanych języków obcych	(średnia)	0,54	0,54	-0,5%

Zmienna	Wartość	Grupa b.o.	Grupa kontrolna	Standaryz. różnica (w %)
Status absolwenta	tak	13,7%	13,9%	-0,4%
Status bez kwalifikacji	tak	9,1%	9,2%	-0,3%
Liczba dni bezrobocia w momencie rozpoczęcia udziału w projekcie	(średnia)	312	314	-0,4%
Liczba otrzymanych propozycji pracy w ciągu roku przed udziałem w projekcie	(średnia)	0,23	0,22	0,1%
Odsetek stawień na wezwanie z urzędu	(średnia)	84,1%	84,9%	-3,2%
Liczba dni przepracowanych w ramach staży absolwenckich, prac interwencyjnych, robót publicznych w ciągu 6 miesięcy przed przystąpieniem do projektu	(średnia)	0,58	0,61	-0,4%
Liczba ukończonych szkoleń w ciągu 2 lat przed udziałem w projekcie	(średnia)	0,03	0,03	-0,2%
Liczba dni ukończonych szkoleń w ciągu 2 lat przed udziałem w projekcie	(średnia)	1,35	1,35	0,0%
Udział w stażu przed rozpoczęciem projektu	tak	6,3%	6,1%	0,6%
Stopień urbanizacji obszaru PUP	(średnia)	58%	58%	0,6%

Źródło: opracowanie własne na podstawie analizy baz danych.

Po dobraniu grupy kontrolnej, na podstawie danych z baz PULS, oszacowano odsetek osób w grupie beneficjentów ostatecznych i w grupie kontrolnej, które posiadały zatrudnienie określoną liczbę miesięcy po zakończeniu udziału w szkoleniu⁷⁴. Różnica między grupą beneficjentów ostatecznych a grupą kontrolną była interpretowana jako efekt uczestnictwa w szkoleniach. W dalszej części artykułu termin „efekt szkoleń” oznaczać będzie właśnie tak rozumiany szacowany wpływ projektów na podjęcie pracy przez beneficjentów ostatecznych.

Wyniki podstawowej analizy

Wyniki uzyskane w ramach oszacowania efektów zatrudnieniowych szkoleń przedstawia poniższa tabela.

Tab. 2. Efekt szkoleń (odsetek pracujących w proc.)

Liczba miesięcy od szkolenia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Beneficjenci ost.	13	28	36	42	45	49	51	54	56	59	60	62	63	64	64	65	65	66
Grupa kontrolna	27	35	40	44	48	52	54	56	58	59	61	62	64	64	65	66	67	67
Efekt szkoleń	-14	-7	-4	-2	-3	-3	-3	-2	-2	0	-1	0	-1	0	-1	-1	-2	-1

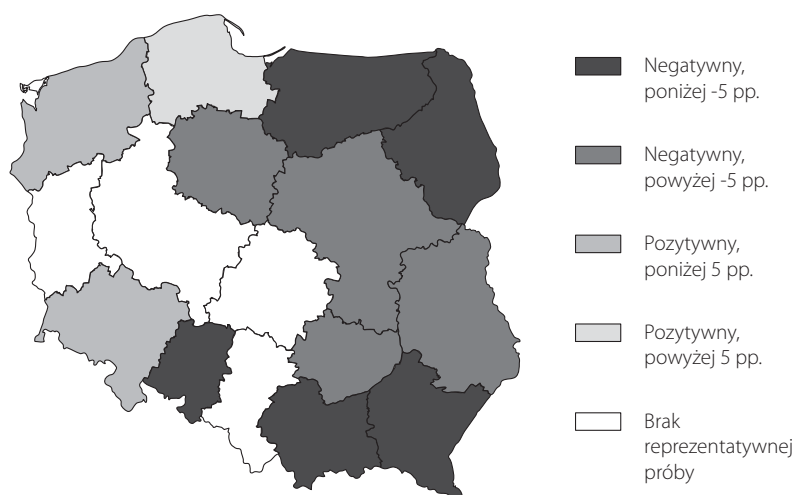
Źródło: opracowanie własne na podstawie analizy PSM.

⁷⁴ W przypadku próby zasadniczej była to faktyczna data ukończenia udziału w szkoleniu, natomiast osobom nieuczestniczącym w szkoleniach współfinansowanych z EFS datę rozpoczęcia i zakończenia udziału przyporządkowano losowo, z puli faktycznych dat rozpoczęcia i zakończenia uczestnictwa, jeszcze przed wyłonieniem spośród nich próby kontrolnej.

Jak widać, w pierwszych miesiącach po zakończeniu szkolenia zanotowano stopniowo malejący efekt ujemny, a następnie, po upływie dziewięciu miesięcy, odsetki osób pracujących w grupie beneficjentów ostatecznych i grupie kontrolnej wyrównały się. Warto zauważyć, że odsetek zatrudnionych w grupie beneficjentów ostatecznych jest bardzo zbliżony do odsetka zatrudnionych w grupie kontrolnej miesiąc wcześniej. Na podstawie tej obserwacji sformułowano wniosek, że przeprowadzone szkolenia nie miały wpływu na szanse podjęcia zatrudnienia z wyjątkiem opóźnienia rozpoczęcia tego zatrudnienia o miesiąc. Zjawisko opóźnienia podjęcia pracy jako ubocznego efektu projektu, przypisywane mniejszej intensywności poszukiwania pracy w trakcie udziału bezrobotnego w szkoleniu, zostało stwierdzone w wielu wcześniejszych badaniach i jest znane w literaturze przedmiotu jako efekt *lock-in* (por. *Study on the Return on ESF Investment in Human Capital, 2010*, s. 74; Lalive, van Ours, Zweimüller, 2002; Fitzenberger, Speckesser, 2005).

Wyniki przeprowadzonej analizy wskazywały również na zróżnicowanie efektu szkoleń między województwami, co przedstawia zamieszczona mapa (Ryc. 1).

Ryc. 1. Oszacowany efekt szkoleń w zależności od województwa



Źródło: opracowanie własne na podstawie analizy PSM

Pogłębienie analiz

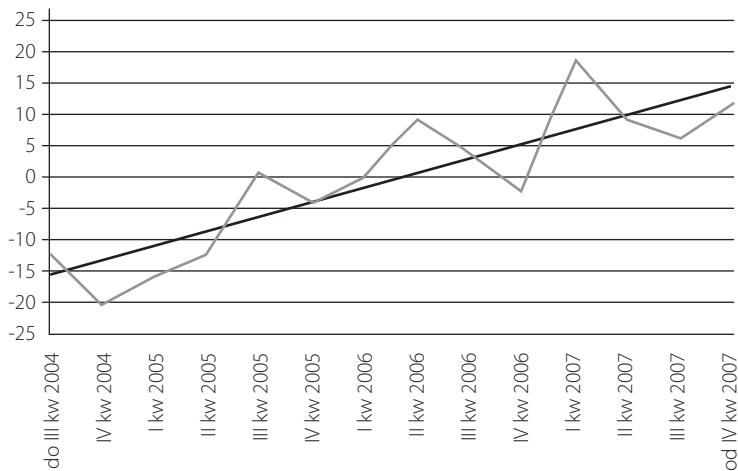
Po uzyskaniu wyżej opisanych wyników przeprowadzono dalsze, pogłębiające analizy, które miały służyć ich wyjaśnieniu i zinterpretowaniu. Jako kluczowe zarysowały się dwa pytania badawcze: po pierwsze, dlaczego, w przeciwieństwie do innych polskich badań, nie stwierdzono pozytywnego efektu szkoleń dla osób bezrobotnych, a po drugie, jakie są przyczyny różnic między województwami. W poszukiwaniu odpowiedzi na te pytania wykorzystano dane pochodzące z następujących źródeł:

- wcześniejszych badań i opracowań dotyczących wsparcia osób bezrobotnych,
- Banku Danych Regionalnych GUS,
- baz danych pozyskanych z Powiatowych Urzędów Pracy,
- dalszych analiz PSM, w ramach których poddano modyfikacjom przyjmowane założenia i algorytm doboru grupy kontrolnej.

Jak już zostało wspomniane, we wcześniejszych badaniach również obserwowano efekt *lock-in*. Wystąpiła jednak wyraźna rozbieżność między wynikami uzyskanymi przez nas a pomiarem efektów dokonanym również przy pomocy metody PSM w ramach opracowania „Zatrudnienie w Polsce 2007” (zob. Bu-

kowski, 2008.). W tym drugim badaniu efekt szkoleń po 18 miesiącach oszacowano na 12,8 pp. Jedną z hipotez upatrywała źródeł rozbieżności w pewnych różnicach metodologicznych. Aby zbadać wpływ tych różnic na wyniki, ponownie przeprowadzono analizę PSM przy założeniach metodologicznych możliwie zbliżonych do tych, które zostały przyjęte w ramach opracowania „Zatrudnienie w Polsce 2007”. W szczególności grupę beneficjentów ostatecznych ograniczono do osób, które korzystały ze wsparcia jedynie w postaci szkoleń współfinansowanych z EFS, a grupę kontrolną do osób, które nie korzystały z żadnego rodzaju aktywnych polityk rynku pracy (niezależnie od ich źródeł finansowania). Ponieważ tego rodzaju operacja nie doprowadziła do zmniejszenia rozbieżności między wynikami badań, a wręcz przeciwnie, do ich zwiększenia, hipoteza została zweryfikowana negatywnie. Inna hipoteza wiązała odmiennosć uzyskanych oszacowań z tym, że analizy na potrzeby opracowania „Zatrudnienie w Polsce 2007” zostały przeprowadzone na grupie obejmującej jedynie uczestników szkoleń rozpoczętych w II kwartale 2006 roku. Po raz kolejny zmodyfikowano zatem założenia analizy PSM, przeprowadzając dobór grupy kontrolnej w ramach poszczególnych kwartałów, co umożliwiło obliczenie efektu szkoleń prowadzonych w różnych okresach realizacji programów perspektywy 2004-2006. Okazało się, że również zgodnie z naszymi analizami w przypadku szkoleń zakończonych w II lub III kwartale 2006 roku wystąpił wyraźny efekt pozytywny, choć jego szacunkowa wielkość była niższa niż podana w opracowaniu „Zatrudnienie w Polsce 2007”. Stwierdzono zatem, że rozbieżności w wynikach badań częściowo mogły wynikać z innego czasowego zakresu analiz. Jeszcze bardziej interesujące ustalenie dokonane na podstawie zmodyfikowanego algorytmu PSM dotyczyło zaobserwowanego trendu stopniowego wzrostu efektu projektów współfinansowanych ze środków EFS: od wyraźnie ujemnego dla szkoleń zakończonych przed III kwartałem 2005 roku, do wyraźnie dodatniego dla szkoleń zakończonych w 2007 roku lub później. Tendencję tę ilustruje zamieszczony wykres (Ryc. 2).

Ryc. 2. Oszacowany efekt szkoleń w zależności od kwartału zakończenia udziału



Źródło: opracowanie własne na podstawie analizy PSM

Zarejestrowany w przybliżeniu zerowy ogólny efekt po 18 miesiącach był zatem wypadkową efektu ujemnego zarejestrowanego w początkowym okresie realizacji programów unijnych oraz efektu dodatniego zarejestrowanego w końcowym okresie realizacji programów unijnych. W znacznej mierze wyjaśniało to, dlaczego pierwotnie uzyskane łączne wyniki sugerowały, że szkolenia nie przyniosły żadnego efektu poza efektem *lock-in*. Jednocześnie jednak pojawiły się dwa kolejne pytania badawcze: po pierwsze, z cze-

go wynika trend wzrostu efektu zatrudnieniowego szkoleń, po drugie, czemu w początkowym okresie zarejestrowany efekt był ujemny.

Dysponując pomiarem efektu szkoleń w rozbiciu na poszczególne województwa oraz na poszczególne kwartały, zestawiono wartość efektu z dostępnymi danymi statystycznymi charakteryzującymi poszczególne województwa lub kwartały. Przyniosło to kolejne ustalenia badawcze. Po pierwsze, nie zaobserwowano klarownej zależności między wielkością zarejestrowanego efektu a ogólnymi wskaźnikami dotyczącymi rynku pracy, takimi jak stopa bezrobocia i zmiana stopy bezrobocia w okresie objętym badaniem. Po drugie, z szacowaną wielkością efektu szkoleń w danym województwie spośród uwzględnionych zmiennych zdecydowanie najsilniej skorelowany był odsetek osób przebywających za granicą w celach zarobkowych (współczynnik korelacji -0,40), zaś wszystkie uwzględnione zmienne (oprócz migracji), charakteryzujące poziom bezrobocia, poziom urbanizacji i wsparcie udzielane bezrobotnym przez PUP łącznie wyjaśniały 56 proc. międzywojewódzkiego zróżnicowania efektu szkoleń. Można zatem powiedzieć, że w pewnym stopniu udało się odpowiedzieć na pytanie, jakie czynniki stoją za zaobserwowanymi różnicami między regionami.

Wyniki analiz dostarczały poparcia dla hipotezy, że zarejestrowany ujemny efekt szkoleń zakończonych w latach 2004-2005 można wiązać z falą masowych zagranicznych migracji zarobkowych w latach 2004-2007 (pięć województw o najbardziej negatywnym efekcie szkoleń to jednocześnie pięć województw o największej skali takich migracji⁷⁵). Aby zrozumieć, w jaki sposób migracje zagraniczne przekładają się na rejestrowany ujemny efekt szkoleń, nieodzowne jest uświadomienie sobie konsekwencji przyjętej metodologii. Sytuacja zawodowa danej osoby była ustalana na podstawie baz danych PULS, które odzwierciedlały wiedzę posiadaną przez Powiatowy Urząd Pracy. W przypadku wyjazdu danej osoby za granicę i jej wykreślenia z rejestru bezrobotnych, w analizach przyjmowane było, że dana osoba znalazła pracę. Można przypuszczać (a za takim przypuszczeniem przemawiają również wyniki badań empirycznych⁷⁶), że osoby uczestniczące w szkoleniach współfinansowanych ze środków EFS cechowała mniejsza skłonność do zagranicznej migracji zarobkowej niż resztę bezrobotnych: klienci PUP biorący udział w szkoleniu raczej nie udawali się po jego odbyciu do innych krajów, a osoby myślące o wyjeździe zagranicznym raczej nie były zainteresowane uczestnictwem w projektach tego typu. W rezultacie w grupie beneficjentów ostatecznych odsetek osób migrujących zarobkowo za granicę był niższy niż w grupie kontrolnej, a ponieważ migracja zagraniczna była klasyfikowana jako podjęcie pracy⁷⁷, sprzyjało to wyższemu odsetkowi zatrudnionych w grupie kontrolnej, czyli ujemnemu rejestrowanemu efektowi szkoleń. Warto zauważyć, że zupełnie inny mechanizm zadziałałby w przypadku ustalenia sytuacji zawodowej na podstawie badań kwestionariuszowych. Z osobami, które wyjechały za granicę i przebywały tam w momencie badania, najprawdopodobniej nie przeprowadzono by wywiadu, czyli przy obliczaniu odsetka pracujących zostałyby one wyłączone z podstawy procentowania. Ponieważ, jak można domniemywać, udział takich osób wśród nieuczestniczących w szkoleniach był większy niż wśród uczestniczących, pociągałoby to za sobą znaczniejsze zaniżenie odsetka pracujących w grupie

⁷⁵ Dane o skali migracji zarobkowych pochodziły z Narodowego Spisu Powszechnego, a zatem dotyczyły 2002 roku. Z powodu braku wiarygodnych danych o skali migracji zarobkowych w latach 2004-2007 wykorzystano dane odnoszące się do sytuacji 3-5 lat wcześniej, przyjmując założenie, że kolejność województw uszeregowanych ze względu na skalę migracji zarobkowej nie zmieniła się zasadniczo w tym okresie. Uzyskana stosunkowo wysoka wartość korelacji z efektem szkoleń przemawia za słusznością takiego założenia.

⁷⁶ W grupie kontrolnej długotrwały wyjazd za granicę (zannotowany jako powód ostatecznego wyrejestrowania) miał miejsce około 35% częściej niż w grupie beneficjentów ostatecznych (por. *Wpływ funduszy Unii Europejskiej na saldo migracji wewnętrznych i zewnętrznych w Polsce*, 2009, s. 29-30.).

⁷⁷ W rejestrze kategoria „wyjazd za granicę na okres co najmniej 30 dni – brak gotowości” stanowiła tylko – zależnie od województwa – 0,5-3% powodów wyrejestrowań. W badaniu założono, że duża część (być może większość) migrantów nie raportowała swojego wyjazdu i została wyrejestrowana na innej podstawie (np. niestawienia się na wezwanie z urzędu). Rodziło to istotny problem w identyfikacji migrantów w analizach.

kontrolnej, a zatem zawyżenie efektów szkoleń. Przyjęta metodologia badania może tu mieć zatem istotny wpływ na uzyskiwane wyniki.

Wpływ migracji zarobkowych, wobec stopniowego zmniejszenia ich skali po 2007 roku, przekłada się również na wzrost rejestrowanego efektu szkoleń na przestrzeni lat. W naszym przekonaniu, migracje nie dostarczają jednak wyczerpującej odpowiedzi na pytanie o przyczyny zaobserwowanej tendencji wzrostowej. Szczegółowe analizy pokazują, że związek skali zagranicznych migracji zarobkowych z wartością rejestrowanego efektu szkoleń nie ma charakteru liniowego – wydaje się raczej, że duża skala migracji w województwie sprzyja zanotowaniu wyraźnie negatywnego efektu wsparcia, ale przy niższej skali o efekcie zatrudnieniowym decydują inne czynniki. Wobec niestwierdzenia klarownej zależności między efektem szkoleń a poziomem lub dynamiką bezrobocia, do zweryfikowania pozostała hipoteza o stopniowym wzroście jakości szkoleń, wynikającym z kumulacji wiedzy i doświadczenia instytucji oraz personelu wdrażającego lub realizującego projekty. W dostępnych danych nie odnaleziono ani empirycznych argumentów przemawiających za tą hipotezą, ani danych pozwalających ją obalić – kluczowym problemem był tu brak wiarygodnych i niezależnych wskaźników jakości szkoleń.

Reasumując, dzięki pogłębieniu analiz udało się w znacznym stopniu odpowiedzieć na sformułowane wyżej pytania badawcze dotyczące efektów szkoleń, choć nie można stwierdzić, że jest to odpowiedź ostateczna i kompletna. Pozostało tu zatem pole do dalszych dociekań, badań empirycznych i analiz.

Ponadto, zdobyte informacje prowadzą do postawienia kolejnych pytań, zmierzających do pełniejszego, bardziej precyzyjnego poznania złożonych zjawisk będących przedmiotem badania. Poza sformułowanymi powyżej, warto wymienić takie zagadnienia badawcze jak:

- Kwestia sezonowości. Jak można zauważyć na Ryc. 2, szacowany efekt szkoleń cechowało znaczne zróżnicowanie w poszczególnych kwartałach roku⁷⁸, przy czym wydaje się, że wyjątkowo niski efekt występował w czwartym kwartale. Czy można te wyniki przypisać sezonowości mającej miejsce na rynku pracy? Czy też rolę odegrały tu inne czynniki?
- Inne wymiary zróżnicowania efektu szkoleń. Przeprowadzone analizy wskazywały na zróżnicowanie międzywojewódzkie oraz zmienność efektu szkoleń w czasie. Okazało się, że efekt szkoleń zbliżony do zerowego jest uśrednieniem ujemnego efektu w niektórych okresach i województwach oraz dodatniego efektu w innych okresach i województwach. Dezagregacja efektu na poziom poszczególnych kategorii uczestników (na przykład wyróżnionych ze względu na płeć, wiek, wykształcenie, miejsce zamieszkania, szczególną sytuację na rynku pracy) lub na poziom rodzaju wsparcia (na przykład wyróżnionych ze względu na tematykę lub formę prowadzenia zajęć) mogłaby ujawnić inne wymiary podobnego zróżnicowania. Potencjalnie mogłoby to być źródłem kolejnych wyników interesujących badawczo, a także użytecznych w kontekście formułowania rekomendacji dotyczących wsparcia osób bezrobotnych.
- Efekt *lock-in* a długość trwania szkoleń. Interesujące byłoby przyjrzenie się długości okresu uczestnictwa w projektach oraz jej wpływu na wystąpienie i długość trwania efektu *lock-in*.
- Zależności między efektem szkoleń a sytuacją na rynku pracy. Wyniki analiz sugerowały, że brak jest prostego, jednoznacznego i silnego powiązania między obserwowanym efektem a poziomem bezrobocia i jego dynamiką. Z teoretycznego punktu widzenia można jednak oczekiwać tu jakichś zależności, co skłania do dalszych badań w tym zakresie.
- Kwestia osób, które przerwały udział w projektach. Osoby takie, stanowiące nieco mniej niż 10% osób rozpoczynających szkolenie, nie weszły ani w skład grupy beneficjentów ostatecznych, ani w skład grupy kontrolnej. Można sformułować pytania, jakie były przyczyny przerwania udziału w projekcie, w jakim stopniu wiązały się one ze znalezieniem pracy, jaki odsetek osób niekończą-

cych szkolenia podjął i utrzymał pracę, ile dni trwał udział tych osób w projekcie i czy udział ten mógł mieć wpływ na podjęcie zatrudnienia. Ma to o tyle znaczenie, że wyłączenie z analiz osób, które znalazły pracę w trakcie uczestnictwa w szkoleniu, może zaniżyć odsetek osób zatrudnionych w grupie beneficjentów ostatecznych, a zatem zaniżyć szacowany efekt projektów.

Powyższe zagadnienia nie zostały podjęte w ramach omawianego badania albo z powodu jego ściśle określonych ram czasowych, które wymusiły zakończenie w pewnym punkcie potencjalnie niemal nieskończonego procesu pogłębiania i poszerzania analiz, albo z powodu braku odpowiednich danych źródłowych. Wyznaczają jednak perspektywy ewentualnych dalszych dociekań badawczych.

Ograniczenia związane z zastosowaną metodą pomiaru efektów

Każda metodologia niesie ze sobą ograniczenia, z których należy zdawać sobie sprawę. Oszacowanie efektów wsparcia z wykorzystaniem grupy kontrolnej, dobranej za pomocą techniki PSM, ma istotne zalety. Należy do nich uniezależnienie tego oszacowania od subiektywnych opinii i deklaracji uczestników szkoleń oraz możliwość kontrolowania jednocześnie bardzo wielu zmiennych mogących wpływać na dalsze losy beneficjentów ostatecznych. Zalety te przesądziły o zastosowaniu techniki PSM w omawianym badaniu – jako najlepszego z potencjalnych rozwiązań. Metoda ta ma naturalnie również swoje ograniczenia i wady. Należy do nich na przykład brak możliwości kontroli efektu zastępowania⁷⁹ i innych efektów pośrednich, wąski zakres uzyskiwanych wyników (ograniczający się do pomiaru średniego wpływu szkoleń na sytuację osób objętych wsparciem) i czasochłonność analizowania danych w podziale na kategorie. Bardziej szczegółowe omówienie zalet, wad i ograniczeń metody PSM czytelnik może znaleźć w opracowaniach jej poświęconych (np.: Bryson, Dorsett, Purdon, 2002; Trzciniński, 2009). W niniejszym artykule skoncentrujemy się natomiast na ograniczeniach związanych z oparciem się przy doborze próby kontrolnej i ustalaniu sytuacji zawodowej badanych osób wyłącznie na bazach danych pochodzących z systemów informatycznych urzędów pracy.

Pierwszym takim, oczywistym ograniczeniem jest to, że dysponuje się jedynie zmiennymi, które są zawarte w otrzymanej bazie danych lub są wynikiem operacji na zmiennych zawartych w bazie. Dla przykładu, ponieważ uzyskane bazy danych nie zawierały informacji o otrzymaniu w ramach projektów EFS dotacji na rozwój działalności gospodarczej, usług pośrednictwa pracy lub doradztwa zawodowego, ani o dacie przerwania przez daną osobę uczestnictwa w szkoleniu, informacje te nie mogły zostać uwzględnione w analizach. Miało to istotne konsekwencje metodologiczne: uniemożliwiło kontrolę otrzymania wyżej wymienionych rodzajów wsparcia przez badane osoby (na przykład w celu wyłączenia osób korzystających z takiego wsparcia z próby kontrolnej) oraz wyodrębnienie kategorii osób, które uczestniczyły tylko w części szkolenia.

Poza samym zakresem zmiennych zawartych w bazach znaczenie ma również jakość tych baz. Osiem przesłanych przez urzędy baz nie mogło zostać włączonych do analizy, ponieważ istniejące w nich braki i błędy uniemożliwiały ustalenie źródła finansowania szkoleń.

Zasób informacji dostępnych w bazach jest silnie związany z ich pochodzeniem – w przypadku omawianego badania były to bazy przekazane przez Powiatowe Urzędy Pracy, które, jak już podkreślano, odzwierciedlały stan wiedzy posiadanej przez urząd. Zawierały one dane związane z kontaktami osób zarejestrowanych w urzędzie. Urząd nie dysponował jednak kompletną informacją o aktualnej sytuacji zawodowej wszystkich osób uwzględnionych w bazie. W związku z tym na potrzeby analiz w wielu przypadkach konieczne było nieco arbitralne, oparte na domniemaniu założenie, jaka jest sytuacja danego członka próby w danym miesiącu. Chyba najbardziej wyrazistym przykładem tego są osoby, które zostały

⁷⁹ W ramach PSM nie można bowiem ustalić, czy podjęcie zatrudnienia przez odbiorców wsparcia nie odbyło się kosztem innych osób nieuczestniczących w projektach i wypartych z miejsc pracy.

wyrejestrowane w związku z niestawianiem się na obowiązkowe wizyty w urzędzie. W badaniu należało rozstrzygnąć, czy osoby te będą kwalifikowane jako osoby, które znalazły pracę (i z tego powodu nie były już zainteresowane odwiedzeniem urzędu), jako osoby, które pozostają bez pracy, czy jako osoby, których sytuacja zawodowa jest nieznana. Ostatecznie uznano, że najmniejszym złem z punktu widzenia dokładności oszacowania efektu szkoleń jest założenie, że osoby te podjęły pracę⁸⁰. Należy jednak zdawać sobie sprawę, że pomiar sytuacji zawodowej przy zastosowanej metodologii jest obarczony pewnym błędem. Do poważnych ograniczeń można ponadto zaliczyć brak danych charakteryzujących miejsce pracy zajmowane przez badaną osobę. Wymienione problemy są związane z pochodzeniem i charakterem wykorzystanych baz, których zasadniczym przeznaczeniem było wspomaganie obsługi osób bezrobotnych i poszukujących pracy, a jedynie wtórnie użyte zostały w ewaluacji efektów projektów unijnych. Pewnie niedostatki baz z punktu widzenia możliwości przeprowadzania analiz nie są tu zatem przywoływane jako zarzut, lecz jako uwarunkowania, które warto mieć na uwadze przy projektowaniu, prowadzeniu i interpretacji wyników badań.

Z oparciem analiz na zastanych, urzędowych bazach danych wiąże się w polskich warunkach prawno-instytucjonalnych jeszcze jedna niedogodność. Czynniki takie, jak niezakończony jeszcze rozwój systemów informatycznych wspomagających administrację publiczną, owocujący rozproszeniem i brakiem kompatybilności baz danych, nie zawsze doskonała współpraca między instytucjami publicznymi oraz restrykcyjne stosowanie przepisów o ochronie danych osobowych, sprawiają, że dostęp ewaluatorów do baz danych jest w Polsce utrudniony w porównaniu do innych krajów Unii Europejskiej⁸¹. W omawianym badaniu przejawiało się to w ten sposób, że z wielu Powiatowych Urzędów Pracy nie udało się pozyskać danych albo z powodu braku woli współdziałania ze strony urzędów, albo z powodu niekompatybilności stosowanych systemów informatycznych⁸², względnie innych problemów technicznych. Na skutek tego w przypadku trzech województw stwierdzono, że próba uczestników szkoleń nie może być uznana za reprezentatywną. Zrezygnowano więc z odrębnego oszacowania efektu szkoleń dla tych trzech województw.

Rozwiązania metodologiczne

Warto zastanowić się w tym miejscu, jakie rozwiązania metodologiczne byłyby pomocne w przezwyciężeniu wyżej opisanych ograniczeń oraz w dalszym zgłębianiu wymienionych zagadnień badawczych. Abstrahując od metodologicznych, czasowych i finansowych ram konkretnego projektu badawczego, można wskazać na kilka takich rozwiązań.

Po pierwsze, zaprezentowane badanie wskazuje na korzyści, które mogą płynąć z modyfikowania założeń stosowanego modelu i algorytmów używanych przy jego stosowaniu. Tego typu operacje wykorzystano do weryfikacji stawianych hipotez oraz do zbadania zmienności efektu szkoleń w czasie. Dalsze modyfikacje w modelu PSM dotyczące określenia zbiorowości, w ramach których dokonywano oszacowania *propensity score* i doboru próby kontrolnej, pozwoliłyby na pomiar efektów szkoleń w różnych kategoriach odbiorców, a zatem poruszenie zagadnienia innych wymiarów różnicowania efektów.

Po drugie, środkiem do wyeliminowania ograniczeń związanych z oparciem pomiaru sytuacji zawodowej na bazach danych byłoby połączenie techniki PSM z badaniem kwestionariuszowym. Dobór próby

⁸⁰ Takie rozwiązanie przyjęto wcześniej w analizie PSM wykonanej w 2006 roku przez firmę PBS DGA na potrzeby ewaluacji ex post komponentów regionalnych programów Phare 2001 SSG RZL oraz Phare 2002 SSG RZL (Konarski, Kotna-rowski, 2007).

⁸¹ Do takich wniosków skłania autora przegląd dokonał ewaluacyjnych poszczególnych krajów zaprezentowanych na spotkaniu *Expert Hearing. MS experiences on using control groups in ESF evaluations*, które odbyło się w Brukseli 25 października 2011 roku.

⁸² W momencie realizacji omawianego badania Powiatowe Urzędy Pracy używały różnych systemów informatycznych do prowadzenia ewidencji osób bezrobotnych. Od tego czasu sytuacja uległa zasadniczej zmianie – obecnie wszystkie PUP używają Systemu Informatycznego Syriusz.

kontrolnej odbywałby się przy wykorzystywaniu techniki *propensity score matching*, a następnie z członkami próby zasadniczej i kontrolnej przeprowadzone zostałyby wywiady kwestionariuszowe, które posłużyłyby do ustalenia ich sytuacji zawodowej. Łączne zastosowanie techniki PSM oraz badania kwestionariuszowego nie jest niczym nowym, ani za granicą, ani na gruncie polskim⁸³. Takie rozwiązanie ma ogromne zalety: pozwala na uzyskanie bardziej rzetelnych danych dotyczących dalszych losów badanych⁸⁴, a ponadto na zebranie wielu dodatkowych, cennych informacji od samego badanego. Proporcjonalne do korzyści są jednak również trudności, ponieważ do przeprowadzenia wywiadów niezbędne byłoby pozyskanie danych osobowych i kontaktowych potencjalnych respondentów. Doświadczenie wskazuje, że panujący w Polsce w 2011 roku ład prawny i instytucjonalny bardzo komplikuje, a w niektórych przypadkach wręcz uniemożliwia pozyskanie danych kontaktowych osób bezrobotnych na potrzeby badań (por. Keler, 2011, s. 21). Pozostaje mieć nadzieję, że ten stan rzeczy ulegnie w przyszłości zmianie.

Po trzecie, w omawianym badaniu bardzo owocna okazała się triangulacja metod badawczych i źródeł danych. Zestawienie wyników analizy PSM z danymi statystycznymi oraz wskaźnikami pochodzącymi z analizy baz danych pozwoliło na testowanie hipotez dotyczących możliwych relacji przyczynowo-skutkowych, a wykorzystanie wcześniejszych opracowań pomogło przy interpretacji uzyskanych wyników i osadzeniu ich w szerszym kontekście. Tą ścieżką można by podążać dalej. Poszerzenie zakresu źródeł danych pozwoliłoby na uwzględnienie w analizach większej liczby zmiennych, na przykład przy zgłębianiu zagadnienia sezonowości efektów. Użyteczne mogłoby być również poszerzenie gamy stosowanych metod i technik badawczych, chociażby o:

- Ilościowe badania kwestionariuszowe – powyżej zarysowano już propozycję integracji techniki PSM oraz badań kwestionariuszowych opartą na wykorzystaniu tej pierwszej do doboru próby. Nie jest to jedyne możliwe połączenie. Wyobraźmy sobie na przykład zebranie dodatkowych, użytecznych informacji w wywiadach kwestionariuszowych z osobami, które przerwały udział w szkoleniu.
- Inne matematyczne techniki szacowania efektu szkoleń. *Propensity score matching* jest tylko jedną z czterech „głównych konfaktycznych metod szacowania efektów” przedstawionych w serwisie Dyrekcji Generalnej ds. Polityki Regionalnej (por. http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/counterfactual_impact_evaluation/index_en.htm, stan na dzień 28 października 2011). Zastosowanie wielu modeli matematycznych i porównanie uzyskanych wyników pozwoliłoby uzyskać oszacowanie zakresu efektu wsparcia wolne od obciążeń konkretnej techniki analitycznej. Takie rozwiązanie z naukowego punktu widzenia zwiększałoby wiarygodność uzyskanych wyników, choć z pragmatycznego punktu widzenia wiązałoby się z pewną wadą: większą złożonością oraz mniejszą klarownością komunikowanych wniosków. W miejsce jednego punktowego oszacowania efektu prezentowane byłoby bowiem kilka punktowych oszacowań uzyskanych przy pomocy różnych technik albo oszacowanie w postaci przedziału (np. „efekt szkoleń wyraża się w zwiększeniu prawdopodobieństwa bycia zatrudnionym o 4-13 punktów procentowych”).
- Wywiady jakościowe – które mogłyby przyczynić się do lepszego zrozumienia mechanizmów obserwowanych zjawisk. Na przykład, pogłębione wywiady z przedstawicielami pracodawców byłyby potencjalnie przydatne w kontekście wyjaśnienia, dlaczego udział w projektach unijnych przyczynia lub nie przyczynia się do zwiększenia prawdopodobieństwa zatrudnienia uczestników.
- Studia przypadków. Przykładowo, studia przypadków osób uczestniczących w szkoleniach, obejmujące między innymi wywiady jakościowe, pomogłyby poznać czynniki wpływające na szanse podjęcia pracy.

⁸³ Przykłady zastosowania: *Evaluation of the New Deal for Lone Parents: Technical report for the quantitative survey* (praca zbiorowa, 2003) oraz Bukowski (2008) w opracowaniu pod jego redakcją kolejność zastosowania technik była odwrotna: najpierw przeprowadzono badanie kwestionariuszowe, a następnie spośród respondentów dobrano techniką PSM próbę kontrolną.

⁸⁴ Bardzo użyteczne z punktu widzenia interpretacji danych zawartych w bazach byłoby ich zestawienie z informacjami pochodzącymi z wywiadów kwestionariuszowych.

Przedstawiony katalog propozycji metodologicznych nie jest oczywiście listą zamkniętą. Zagadnienia ewaluacji efektów Aktywnych Polityk Rynku Pracy pozostawiają spore pole do popisu dla pomysłowości badaczy.

Paweł Penszko – badacz społeczny specjalizujący się w metodach ilościowych. Ukończył studia socjologiczne i slawistyczne na Uniwersytecie Warszawskim. W latach 2006-2010 badacz w firmie PBS DGA, od lutego 2010 roku ekspert ds. badań w firmie PAG Uniconsult. Brał udział w 19 badaniach ewaluacyjnych w charakterze kierownika projektu, badacza, analityka i współautora raportu. W projekcie „Wpływ polityki spójności na poziom i jakość zatrudnienia w Polsce” był odpowiedzialny za badania i analizy ilościowe w ramach podejścia bottom-up, w tym za przeprowadzenie analiz techniką PSM.

Bibliografia:

- Bryson A., R. Dorsett, S. Purdon, 2002, *The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies*, Working Paper Numer 4, Crown, London;
- Bukowski M. (red.), 2008, *Zatrudnienie w Polsce 2007. Bezpieczeństwo na elastycznym rynku pracy*, Departament Analiz Ekonomicznych i Prognoz, Ministerstwo Pracy i Polityki Społecznej, Warszawa;
- *Evaluation of the New Deal for Lone Parents: Technical report for the quantitative survey*, 2003, praca zbiorowa, Department for Work and Pensions;
- Fitzenberger B., Speckesser S., 2005, *Employment Effects of the Provision of Specific Professional Skills and Techniques in Germany*, IZA, Discussion Paper No. 1868, Bonn;
- http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/counterfactual_impact_evaluation/index_en.htm, stan na dzień 28 października 2011;
- Keler K., 2011, *Raport metodologiczny z badań realizowanych w 2010 r. w ramach projektu „Bilans Kapitału Ludzkiego”*, PARP, Warszawa;
- Konarski R., Kotnarowski M., 2007, *Zastosowanie metody propensity score matching w ewaluacji ex-post*, [w:] A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa;
- Lalive R., van Ours J. C., Zweimüller J., 2002, *The Impact of Active Labor Market Programs on the Duration of Unemployment*, Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 41, Zurich;
- *Study on the Return on ESF Investment in Human Capital. Final Report*, 2010, Euréval / Ecorys / Rambøll Management, DG EMPL;
- Trzciński R., 2009, *Wykorzystanie techniki propensity score matching w badaniach ewaluacyjnych*, PARP, Warszawa;
- *Wpływ funduszy Unii Europejskiej na saldo migracji wewnętrznych i zewnętrznych w Polsce*, 2009, MRR, Warszawa;
- *Wpływ polityki spójności na poziom i jakość zatrudnienia w Polsce*, 2010, Ministerstwo Rozwoju Regionalnego, Warszawa;

Szacowanie efektów projektów drogowych. Przykłady zastosowania modelu ustalania i prognozowania efektów z wykorzystaniem wskaźnika dostępności transportowej

Wprowadzenie

Jednym z głównych kierunków ewaluacji inwestycji infrastrukturalnych w transporcie jest badanie dostępności. Dostępność transportowa analizowana jako tzw. dostępność kumulatywna oraz dostępność mierzona za pomocą metody potencjału (tzw. dostępność potencjałowa) dają możliwości ewaluacji efektów poszczególnych inwestycji, zarówno *ex-post* jak i *ex-ante*.

Institut Geografii i Przestrzennego Zagospodarowania PAN we współpracy z Ministerstwem Rozwoju Regionalnego prowadzi od kilku lat badania dostępności transportowej na poziomie powiatów w postaci Wskaźnika Międzygałęziowej Dostępności Transportowej WMDT. Zaprojektowana w 2011 r. w IGIPIZ PAN i stanowiąca rozszerzenie wskaźnika WMDT aplikacja komputerowa OGAM daje możliwości symulacji zmian dostępności potencjałowej drogowej w wyniku realizacji inwestycji infrastrukturalnych na poziomie gminnym. Szacowanie efektów projektów drogowych na obszarze całego kraju, pojedynczych województw, czy też nawet powiatów lub gmin może mieć wielorakie zastosowanie w realizacji polityki transportowej. Może służyć m.in.:

- ustalaniu i ocenie skali zmian dostępności transportowej jakie nastąpiły w wyniku realizacji łącznie wszystkich inwestycji drogowych (efekt netto) oraz ocenie efektywności rozwiązań i poprawności alokacji środków (poziom koncentracji, spójność z docelową siecią transportową),
- prognozowaniu efektów projektów w celu stworzenia rankingu inwestycji i pokazaniu, które przedsięwzięcia są najbardziej efektywne w kontekście zmian dostępności,
- wskazaniu pożądanej strategii inwestowania, czyli w jakiej kolejności budować poszczególne odcinki aby osiągnąć maksymalne efekty synergii w najkrótszym czasie.

Należy jednak pamiętać, że racjonalna ocena skuteczności i efektywności interwencji publicznych w obszarze infrastruktury transportowej nie może uwzględniać jedynie bezpośrednich potrzeb społecznych i gospodarczych (czyli popytu na przewozy osób i towarów), ale również brać pod uwagę szerszy kontekst, w tym zgodność z celami strategicznymi polityki transportowej, rozwoju regionalnego i polityki ochrony środowiska (wsparcie dla rozwiązań mniej kosztownych ekologicznie) oraz możliwości realizacyjne (w tym finansowe).

Kluczowe dla wyboru *ex ante* inwestycji transportowych powinno być poprzedzanie podejmowania decyzji badaniami, które pozwalają na symulację efektów łącznej realizacji planowanych inwestycji w różnych wariantach zmian sytuacji społeczno-ekonomicznej. Fundamentalną wartość takich analiz – w porównaniu do pojedynczych studiów wykonalności – stanowi możliwość uchwycenia wzajemnych zależności pomiędzy inwestycjami, czyli wpływu realizacji jednych przedsięwzięć na realizację drugich. W ten sposób można systemowo zapobiegać powstawaniu inwestycji wzajemnie ze sobą konkurujących i wspierać inwestycje tworzące efekty synergii, o największym znaczeniu z punktu widzenia spójności terytorialnej kraju, spójności z docelową siecią transportową oraz poprawy dostępności przestrzennej w różnych skalach (Wolański, 2010).

Biorąc pod uwagę powyższe, w sytuacji ograniczonych zasobów finansowych stworzenie mechanizmu selekcji inwestycji drogowych charakteryzujących się największą efektywnością, a nie najłatwiejszych do realizacji, powiązanego ze wstępną analizą efektywności na etapie tworzenia takich dokumentów, jak np. listy indykacyjne, wykazy inwestycji w programie budowy dróg krajowych (MRR, 2011b), powinno doprowadzić do wyboru projektów, które gwarantowałyby że:

- beneficjentami projektu będzie wielu podróżujących (biorąc pod uwagę potoki osób i ładunków);
- nastąpi duże skrócenie czasu przejazdu;
- uzyska się znaczną poprawę efektywności ekonomicznej (w szczególności zapotrzebowania na dotacje operacyjne);
- istotnie wpłynie się na poziom bezpieczeństwa.

1. Diagnoza stanu infrastruktury transportowej w Polsce i wyzwania rozwojowe

Analiza procesów zachodzących w okresie 2000-2010 prowadzi do wniosku, że mimo wzmożonego wysiłku, szczególnie w ostatnich latach w wyniku realizacji projektów współfinansowanych ze środków unijnych, w Polsce w dalszym ciągu brak jest spójnego i sprawnie funkcjonującego systemu transportowego, zintegrowanego z systemem europejskim.

Zjawiskiem negatywnym jest wciąż mniejsza dynamika poprawy wskaźników gęstości sieci infrastruktury w relacji do wzrostu PKB i intensywności przewozów (przyrost sieci dróg utwardzonych (o 7,6% w latach 2000-2009) pozostawał w tyle zarówno za wzrostem liczby pojazdów w ruchu (wzrost o 56,1%), jak i za wzrostem PKB (o 40,7%), co w efekcie doprowadziło do sytuacji, w której na 1 km sieci dróg utwardzonych w Polsce przypadało więcej pojazdów niż średnio w UE-27 (odpowiednio 70 i 53). Głównym problemem jest nadal słabe nasycenie kraju siecią autostrad i dróg ekspresowych, dla których wskaźnik w odniesieniu do PKB kształtuje się na poziomie 60,6% średniego w UE-27 i na poziomie 19,1% wskaźnika najwyższego w EU-27 (w Słowenii) oraz utrzymujący się niski poziom bezpieczeństwa ruchu drogowego (wskaźnik liczby ofiar na 10 mld pasm wykonanych łącznie samochodami osobowymi, autobusami i autokarami dalekobieżnymi w 2009 r. kształtował się na poziomie ponad dwukrotnie wyższym od średniego w UE, a gorszy wskaźnik niż Polska odnotowały jedynie Słowacja, Węgry, Bułgaria i Rumunia (Ministerstwo Infrastruktury, 2011).

Powyższe dane pokazują skalę istniejącej w Polsce luki infrastrukturalnej w porównaniu z resztą Unii Europejskiej oraz niedopasowanie krajowego systemu transportowego do intensywności produkcji i wymiany, jak też mobilności mieszkańców.

Na podstawie najnowszych badań można oszacować, że spowodowana niewydolnością systemu transportowego względna wartość strat czasu w transporcie (w stosunku do PKB, tylko w obrębie miast) wzrosła w latach 2001-2009 z 0,4% do 0,9%. Straty te wywołuje zarówno zły stan infrastruktury, jak i zła organizacja systemów transportowych. Inwestycje eliminujące tę niewydolność mogą więc przynieść oszczędności ogólnogospodarcze wynoszące kilkanaście miliardów złotych rocznie (Ministerstwo Infrastruktury, 2011).

Dlatego też długookresowym celem polityki transportowej (do 2030 roku) jest „zwiększenie dostępności terytorialnej Polski poprzez utworzenie zrównoważonego, spójnego, przyjaznego użytkownikom systemu transportowego w wymiarze krajowym, europejskim i globalnym” (KPRM, 2011). Wyzwaniem zatem jest w pierwszej kolejności usunięcie zaległości w rozbudowie i modernizacji zintegrowanego systemu infrastruktury transportowej (por. Ryc. 1)⁸⁵ oraz połączenie ośrodków wzrostu z obszarami peryferyjnymi i włączenie ich w sieć transportu

europejskiego. Konieczna jest również zmiana sposobu organizacji i zarządzania infrastrukturą transportową, poprawa bezpieczeństwa użytkowników ruchu i towarów. Poprawa komunikacji w obszarach miejskich i metropolitalnych oraz ograniczenie negatywnego wpływu transportu na środowisko⁸⁶.

Europejska polityka transportowa coraz większą uwagę poświęca zagadnieniom dotyczącym ograniczeniu negatywnego wpływu transportu na środowisko naturalne, a główne wyzwania stanowią zmiany klimatyczne, ograniczenie emisji gazów cieplarnianych, hałas, zatłoczenie dróg oraz wypadki. Najważniejszym elementem polityki transportowej UE jest zrównoważony rozwój transportu⁸⁷ (por. Komisja Europejska, 2010 i 2011).

Należy podkreślić, że strategia Europa 2020 odchodzi od traktowania rozwoju infrastruktury transportowej jako celu samego w sobie. Niemniej jednak infrastruktura transportowa bezpośrednio i pośrednio wpływa na tworzenie inteligentnego, zrównoważonego rozwoju sprzyjającego włączeniu społecznemu. Nowoczesne połączenia komunikacyjne na obszarach metropolitalnych przyspieszają rozpowszechnianie wiedzy między centrum miasta a przedmieściami i najbliższymi okolicami. Zrównoważony transport sprzyja gospodarce bardziej ekologicznej, efektywniej korzystającej z zasobów. Co równie istotne, sieci transportowe gwarantują spójność terytorialną, dzięki czemu wzrost w większym stopniu sprzyja włączeniu społecznemu (MRR, 2011a).

Bezpośrednio działania z zakresu transportu wskazane są w strategii Europa 2020 w ramach II priorytetu: *rozwój zrównoważony*. Modernizacja transportu służyć ma bowiem przechodzeniu na gospodarkę niskoemisyjną (Komisja Europejska, 2010).

Zgodnie z pakietem legislacyjnym polityki spójności na lata 2014-2020⁸⁸ promowanie zrównoważonego transportu i usuwanie niedoborów przepustowości w działaniu najważniejszych infrastruktur sieciowych (cel tematyczny 7) będzie realizowane m.in. poprzez:

- wspieranie multimodalnego jednolitego europejskiego obszaru transportu poprzez inwestowanie w transeuropejską sieć transportową (TEN-T);
- zwiększanie mobilności regionalnej poprzez łączenie węzłów drugorzędnych i trzeciorzędnych z infrastrukturą TEN-T;
- rozwój przyjaznych dla środowiska i niskoemisyjnych systemów transportu miejskiego oraz promowanie mobilności miejskiej zgodnej z zasadami zrównoważonego rozwoju;
- rozwój kompleksowego, wysokiej jakości interoperacyjnego systemu kolejowego.

Ocena ex post efektów polityki spójności w Polsce w okresie 2004-2006 podkreśla, że dalsza rozbudowa infrastruktury wymaga podejścia sieciowego i systemowego oraz powinna opierać się na solidnych

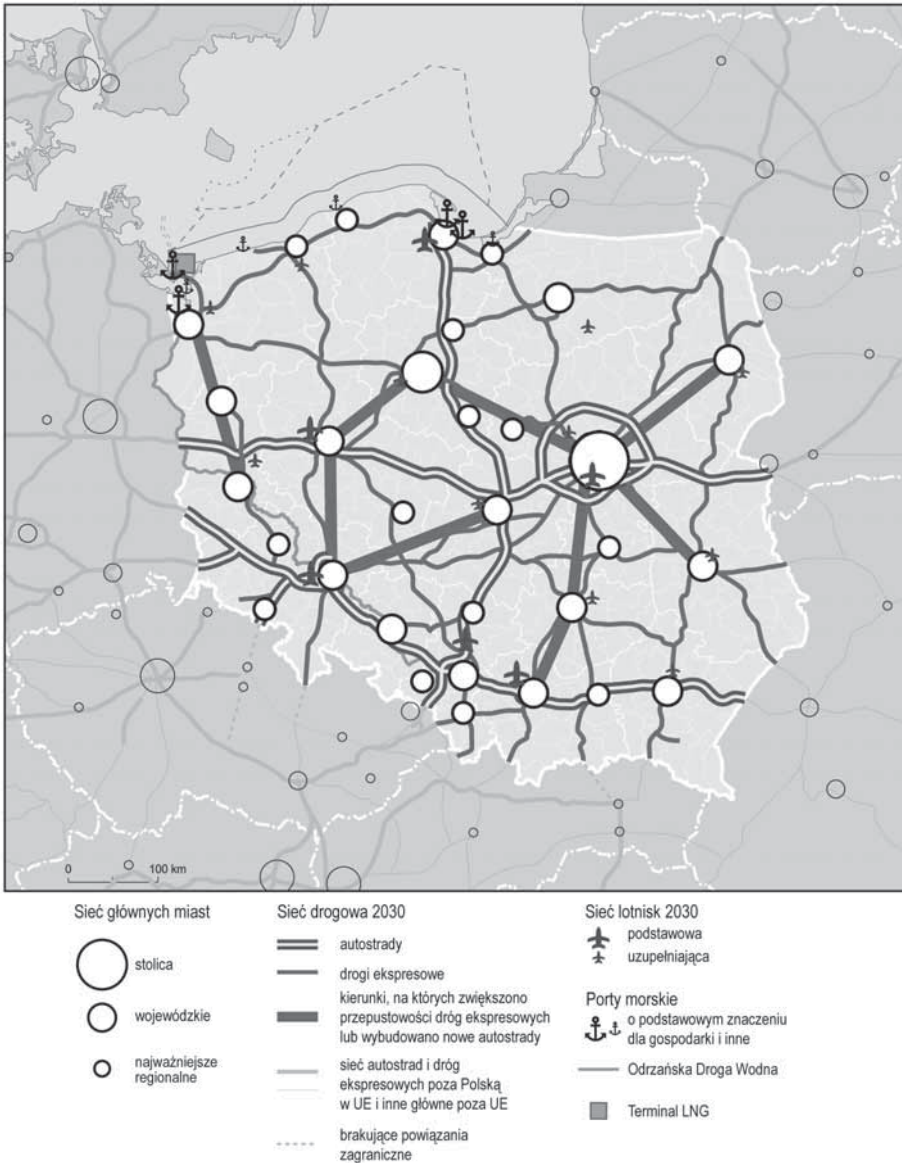
⁸⁶ Szczegółowe cele dotyczące rozwoju transportu (na poziomie kraju i poszczególnych regionów) w perspektywie do roku 2020 sformułowano w *Strategii Rozwoju Kraju 2020* (projekt), *Strategii Rozwoju Transportu* (projekt) i *Krajowej Strategii Rozwoju Regionalnego 2010-2020: Regiony, Miasta, Obszary wiejskie*.

⁸⁷ *Zrównoważony rozwój transportu* – jest to taki rozwój transportu, który: a) zapewnia dostępność celów komunikacyjnych w sposób bezpieczny, niezagrażający zdrowiu ludzi i środowisku w sposób równy dla obecnej i następnych generacji; b) pozwala funkcjonować efektywnie, oferować możliwość wyboru środka transportowego i podtrzymać gospodarkę oraz rozwój regionalny; c) ogranicza emisje i odpady w ramach możliwości zaabsorbowania ich przez ziemię, zużywa odnawialne zasoby w ilościach możliwych do ich odtworzenia, zużywa nieodnawialne zasoby w ilościach możliwych do ich zastąpienia przez odnawialne substytuty, przy minimalizowaniu zajęcia terenu i hałasu (definicja ze *Strategii Rozwoju Transportu do 2020* – projekt).

⁸⁸ 6 października 2011 r. Komisja Europejska przedstawiła pakiet legislacyjny dla polityki spójności na lata 2014-2020, składający się z następujących projektów rozporządzeń:

1. rozporządzenie podstawowe, ustanawiające wspólne przepisy dotyczące Europejskiego Funduszu Rozwoju Regionalnego (EFRR), Europejskiego Funduszu Społecznego (EFS), Funduszu Spójności, Europejskiego Funduszu Rolnego na rzecz Rozwoju Obszarów Wiejskich (EFRROW) oraz Europejskiego Funduszu Morskiego i Rybołówstwa (EFRM)
2. rozporządzenie dot. Europejskiego Funduszu Rozwoju Regionalnego (EFRR),
3. rozporządzenie dot. Europejskiego Funduszu Społecznego (EFS),
4. rozporządzenie dot. Funduszu Spójności,
5. rozporządzenie dot. Europejskiej Współpracy Terytorialnej (EWT),
6. rozporządzenie dot. Europejskiego Ugrupowania Współpracy Terytorialnej (EUWT).

Ryc. 1. Wizja kształtowania podstawowych elementów sieci drogowej 2013*: autostrady i drogi ekspresowe, lotniska, porty morskie



* realizacja poszczególnych inwestycji będzie zależeć od polityki transportowej rządu

Źródło: Koncepcja Przestrzennego Zagospodarowania Kraju 2030, Ministerstwo rozwoju Regionalnego, Warszawa 2011.

podstawach merytorycznych (większy udział ekspertów w podejmowaniu ważnych decyzji w tym zakresie, badania preferencji społecznej). Istotnym warunkiem optymalnego zaspokajania potrzeb społeczno-gospodarczych jest minimalizacja wpływu czynników pozagospodarczych i pozarozwojowych na zakres i kolejność realizowanych inwestycji (MRR, 2010a, s. 87-88).

Prognozowanie efektów projektów drogowych z wykorzystaniem wskaźnika dostępności transportowej wskazuje na potrzebę priorytetyzacji inwestycji transportowych, tak by w pierwszej kolejności realizować te fragmenty sieci, które są jednocześnie odcinkami dróg wylotowych z najważniejszych metropolii kraju oraz miast wojewódzkich, obwodnicami miast powiatowych, a także odcinkami na trasach łączących

najważniejsze aglomeracje kraju. Odcinki te odgrywają kluczową rolę w poziomie dostępności. Należy zatem zidentyfikować najważniejsze trasy (autostrady i drogi ekspresowe), które nie mogą podlegać ewentualnym dalszym redukcjom budżetowym⁸⁹.

Potrzebna jest ponadto, w dalszej kolejności budowa pozostałych dróg ekspresowych. Ich realizacja powinna następować etapami. Najważniejsze są odcinki odprowadzające ruch z miast powyżej 100 tys. mieszk., dalej z miast pow. 50 tys. mieszk., a dopiero w dalszej kolejności odcinki zamykające sieć (połączenia z istniejącą siecią dróg ekspresowych i autostrad) oraz odcinki przygraniczne (z tych ostatnich na pierwszym planie powinny się znaleźć odcinki o największym natężeniu ruchu, np. budowa drogi ekspresowej S61 w kierunku granicy polsko-litewskiej w Budzisku).

2. Metody badania dostępności transportowej – diagnoza stanu dostępności

Dostępność transportową można badać za pomocą zestawu metod (por. Komornicki, Śleszyński, Rosik, Pomianowski, 2010). Najczęściej stosuje się jedno z czterech podstawowych podejść metodycznych:

- Dostępność rozumiana jako gęstość sieci infrastrukturalnej i/lub sieci transportu publicznego, w odniesieniu do jednostek powierzchni lub zaludnienia (podejście tradycyjne).
- Dostępność kumulatywna – wyrażona odległością czasową (minuty), fizyczną lub ekonomiczną do określonych ośrodków-celów podróży (najczęściej węzłów sieci); prezentowana najczęściej za pomocą map izochronowych oraz zestawień ukazujących kumulacje potencjału (demograficznego lub ekonomicznego) w obrębie określonej izochrony.
- Dostępność dzienna, obrazująca możliwość podróży powrotnej do określonego celu (celów) podróży np. w przeciągu jednego dnia; uwzględniająca jednocześnie możliwości infrastruktury oraz organizację przewozów w transporcie publicznym.
- Dostępność potencjałowa obrazująca za pomocą syntetycznego wskaźnika sumę relacji między wszystkimi elementami (ośrodkami, regionami) badanego zbioru, przy czym każda relacja uwzględnia zarówno czas przejazdu między ośrodkami, a także znaczenia (atrakcyjności) tych ośrodków w systemie (potencjał demograficzny, ekonomiczny lub inny).

W niniejszym opracowaniu nacisk został położony na dostępność potencjałową, aczkolwiek na wstępie pokazano również możliwości, jakie niesie wykorzystanie dostępności kumulatywnej.

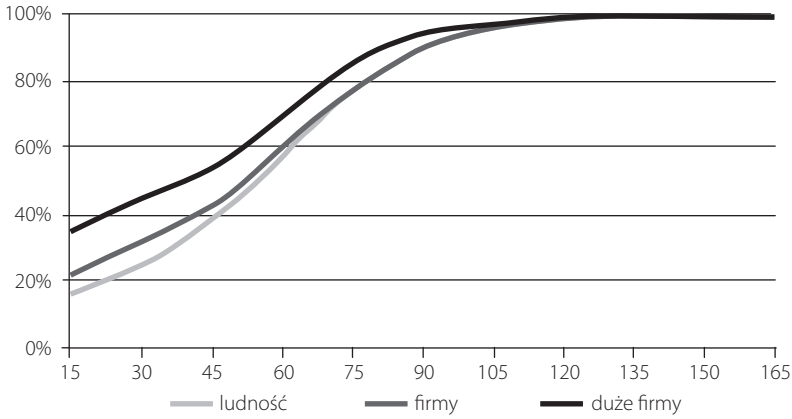
2.1. Dostępność kumulatywna (*cumulative accessibility*)

W przypadku dostępności kumulatywnej badamy, ile danego rodzaju celów podróży znajduje się w obrębie danej izochrony (linii łączącej punkty o jednakowej odległości od źródła podróży). Zakłada się, że atrakcyjność tych celów jest taka sama niezależnie od tego, czy są one położone relatywnie blisko, czy też bardziej oddalone. Przykładem zastosowania tej metody jest diagnoza skumulowanego odsetka ludności i firm Polski Wschodniej znajdujących się w zasięgu poszczególnych izochron dojazdu do pięciu ośrodków wojewódzkich, tj. Białegostoku, Lublina, Rzeszowa, Kielc oraz Olsztyna.

⁸⁹ Jako takie wymienić należy:

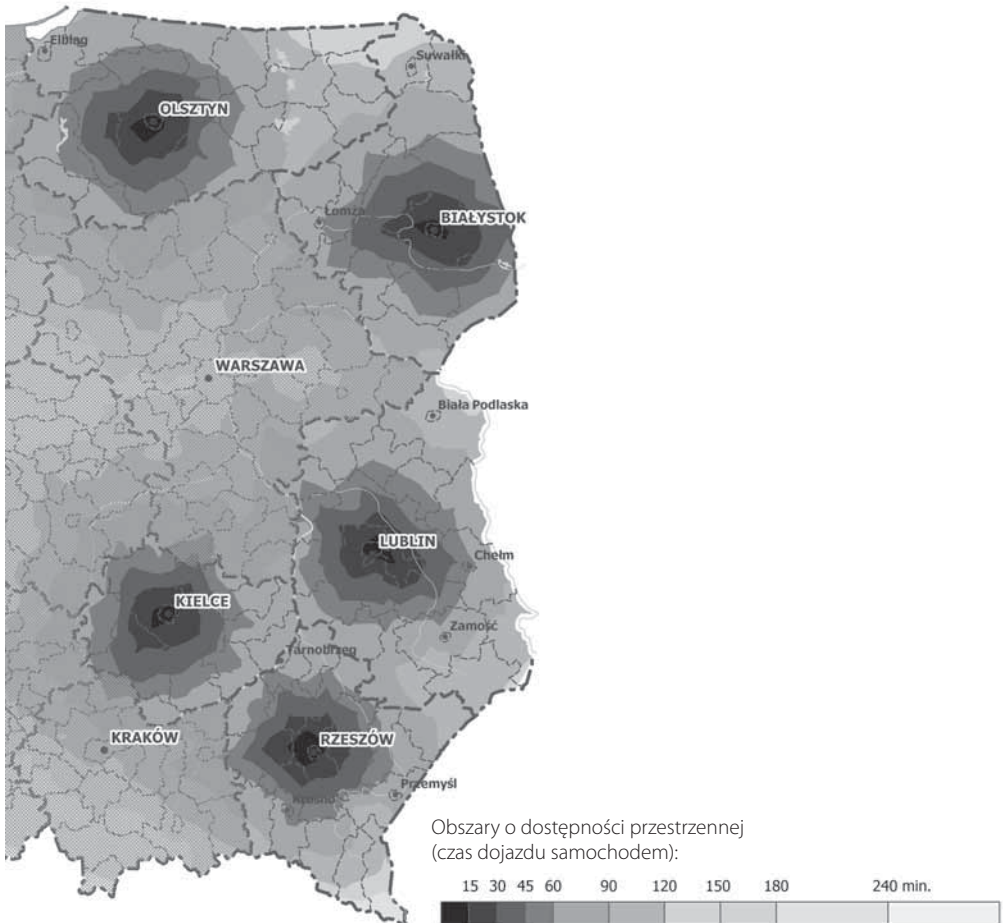
- ukończenie budowy warszawskiego węzła drogowego z pełną obwodnicą i wschodnim wyjściem A2 (do Siedlec) oraz autostrad A1 i A4 na całym ich przebiegu przez terytorium Polski,
- ukończenie drogi ekspresowej S7 w pełnym przebiegu od Gdańska do Rabki,
- ukończenie drogi ekspresowej S5 od Grudziądza przez Poznań do Wrocławia;
- ukończenie drogi ekspresowej S17 Warszawa-Lublin (Piaski),
- ukończenie drogi ekspresowej S8 od Wrocławia do Łodzi i od Piotrkowa Trybunalskiego przez Warszawę do Białegostoku.

Ryc. 2. Skumulowany odsetek ludności i firm Polski Wschodniej według czasu dojazdu do ośrodków wojewódzkich



Źródło: T. Komornicki, P. Rosik, M. Stępnik, 2011a, *Dostępność transportowa w Polsce Wschodniej*.

Ryc. 3. Przestrzenna dostępność ośrodków wojewódzkich Polski Wschodniej



2.2. Dostępność potencjałowa (*potential accessibility*)

Do celów ewaluacyjnych, a przede wszystkim w celu badania wpływu poszczególnych inwestycji na zmiany dostępności, najlepiej służy tzw. dostępność potencjałowa. Dostępność potencjałowa mówi o dostępności do wszystkich możliwych celów podróży. Do zbioru celów podróży może należeć np. cały obszar Polski – badania prowadzone przez IGIPIZ PAN lub obszar całej Europy – badania ESPON (Spiekermann, Schürmann, 2007). W przestrzeni Polski można uogólniać, że podstawowymi celami podróży są miejscowości węzłowe, w których znajdują się siedziby powiatu (przy wskaźniku WMDT), lub siedziby gminy – najnowsza aplikacja OGAM (Rosik, Pomianowski, Stępiak, Komornicki, Śleszyński, 2011). Przy obliczaniu wskaźnika dostępności poszczególnych miejscowości węzłowych bierze się pod uwagę szybkość dotarcia do wszystkich pozostałych miast. Atrakcyjność tychże miast węzłowych (celów podróży) maleje wraz z wydłużaniem się czasu podróży. Innymi słowy zakłada się, że wraz z wydłużaniem się czasu podróży atrakcyjność celu podróży maleje, ponieważ uczestnik ruchu jest bardziej skłonny do podróżowania na krótsze niż dłuższe odległości.

W niniejszym opracowaniu wykorzystano dwie możliwości badania dostępności potencjałowej przestrzeni Polski. Pierwszy to wskaźnik WMDT, czyli Wskaźnik Międzygałęziowej Dostępności Transportowej, rozwinięty w IGIPIZ PAN w latach 2008-2010 (szerzej Komornicki, Śleszyński, Rosik, Pomianowski, 2010). Drugi to wskaźnik WDDT, czyli Wskaźnik Drogowej Dostępności Transportowej⁹⁰ obliczany w ramach nowej aplikacji zbudowanej w IGIPIZ PAN.

Istnieje parę różnic między wyżej wymienionymi wskaźnikami. Do najważniejszej należy z pewnością to, że pierwszy uwzględnia badania dostępności potencjałowej na poziomie powiatów, drugi natomiast „schodzi” poziom niżej, tj. pozwala na badanie dostępności na poziomie gminnym. Kolejną różnicą jest to, że pierwszy ma naturę multimodalną (uwzględnia oprócz transportu drogowego również kolejowy, wodny-śródlądowy oraz lotniczy), a drugi ma jak dotąd naturę unimodalną (tylko transport drogowy). W przyszłości na poziomie gminnym istnieje również możliwość uwzględnienia sieci transportu publicznego. Jest to jednak dość skomplikowane metodologicznie.

2.3. Dostępność potencjałowa powiatów według wskaźnika WMDT

W pierwszym wymienionym wyżej wskaźniku WMDT do opisanego mas celów podróży (założono, że celami podróży są siedziby powiatów lub grup powiatów w układach metropolitalnych) wykorzystano 12 zmiennych społeczno-gospodarczych, takich jak: liczba ludności, liczba zameldowań i wymeldowań na pobyt stały, liczba mieszkańców z wykształceniem wyższym, liczba pracujących w usługach, liczba pracujących w przemyśle i budownictwie, liczba pracujących w przemyśle i budownictwie w zakładach powyżej 9 osób, liczba gospodarstw rolnych produkujących na rynek, liczba podmiotów gospodarczych o liczbie pracujących 250 i więcej, produkcja sprzedana przemysłu, korzystający z noclegów ogółem, liczba studentów oraz ranga administracyjna. Zmienne te zważono przypisując wagi w zależności od ich roli w generowaniu pracy przewozowej w transporcie pasażerskim i towarowym według subiektywnego kryterium. Taką metodologię zastosowano przy konstrukcji wskaźnika WMDT.

Warto jednak zaznaczyć, że metoda dostępności potencjałowej nie definiuje, czym ma być tzw. masa stanowiąca cel podróży. Wybór tzw. masy zależy głównie od motywacji podróży. Dla podróżujących w celu dokonania cotygodniowych zakupów uwzględnioną masą powinny być centra handlowe, dla podróżujących do pracy – miejsca pracy, dla podróżujących do szpitali – np. liczba łóżek szpitalnych. Przy konstrukcji wskaźnika WMDT przyjęto natomiast subiektywnie, że wyżej wymienione 12 zmiennych ważone w różnych proporcjach stanowi o atrakcyjności danego celu podróży (powiatu lub grupy powiatów).

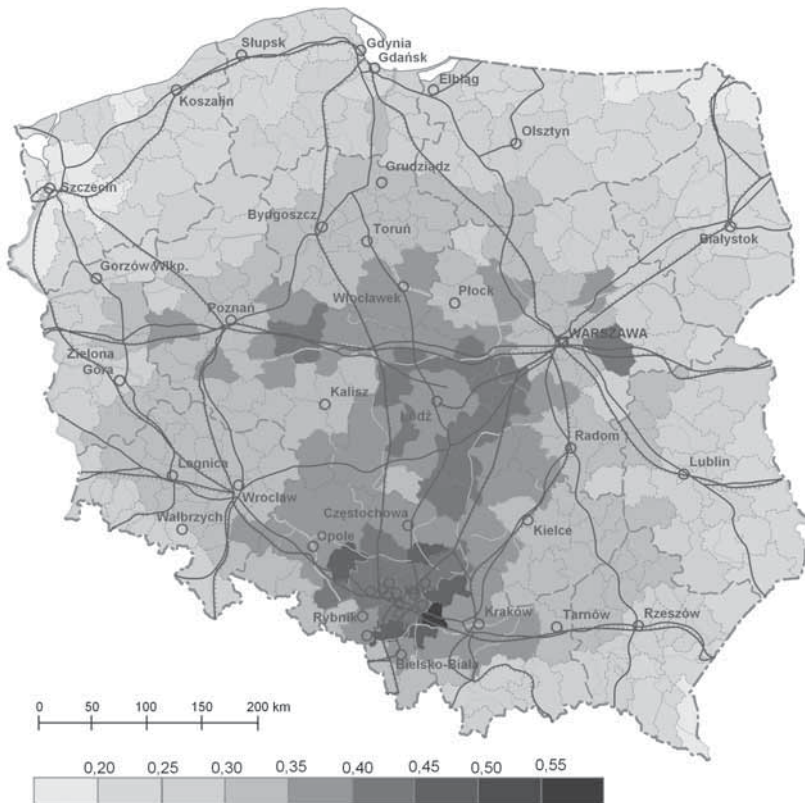
⁹⁰ Nazwa wskaźnika powinna brzmieć Wskaźnik Drogowej Dostępności Potencjałowej, jednak główną ideą było porównanie wskaźnika do istniejącego wcześniej WMDT i z tego względu podobna nazwa, tj. WDDT, a nie WDDP. (szerzej [w:] Rosik, Pomianowski, Stępiak, Komornicki, Śleszyński, 2011)

Przy wykorzystaniu metody dostępności potencjałowej skonstruowano wskaźnik syntetyczny WMDT charakteryzujący atrakcyjność poszczególnych powiatów jako celów podróży i uwzględniający jednocześnie cztery gałęzie transportu, tj. transport drogowy, kolejowy, wodny-śródlądowy i lotniczy. Tak skonstruowany wskaźnik określa stan dostępności transportowej powiatu. Wskaźnik nie ma jednostek (jest to cecha dostępności potencjałowej). Można jedynie powiedzieć, że jednostki o wyższej wartości wskaźnika charakteryzują wyższą dostępność. Przy konstrukcji wskaźnika dostępności potencjałowej wykorzystano następujący wzór:

$$A_i = \sum_{j, j \neq i} \frac{M_j}{t_{ij}} \quad (1)$$

gdzie: A_i – dostępność powiatu lub grupy powiatów i , M_j – masa powiatu lub grupy powiatów j , obliczona na podstawie 12 zmiennych (niezależnie dla transportu pasażerskiego i towarowego), t_{ij} – czas przejazdu między powiatem lub grupą powiatów i a powiatem lub grupą powiatów j . W ten sposób oblicza się wskaźnik dla poszczególnych gałęzi, a następnie przy konstruowaniu wskaźnika multimodalnego uznaje się, że udziały poszczególnych gałęzi transportu są ważone ich rolą w pracy przewozowej w transporcie pasażerskim i towarowym na poziomie wojewódzkim⁹¹.

Ryc. 4. Wskaźnik WMDT w ruchu pasażerskim w 2008 r.



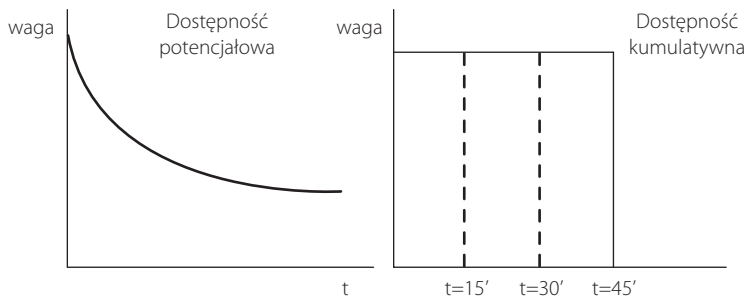
Źródło: T. Komornicki, P. Śleszyński, P. Rosik, W. Pomianowski, 2010, *Dostępność przestrzenna jako przesłanka kształtowania polskiej polityki transportowej*, Biuletyn KPZK 241, Komitet Przestrzennego Zagospodarowania Kraju PAN, s. 167, Warszawa.

Założono zatem, zgodnie z metodologią modelu potencjału, że wraz z wydłużaniem się czasu podróży, atrakcyjność powiatów maleje. Czasy podróży zostały obliczone dla transportu drogowego za pomocą modelu prędkości ruchu opracowanego w IGIPIZ PAN. Model ten uwzględnia różnice w prędkości podróży na sieci dróg krajowych i wojewódzkich wynikające z kategorii drogi, szerokości jezdni, ukształtowania powierzchni terenu, liczby ludności zamieszkującej w buforze poszczególnych odcinków sieci, a także udziału terenu zabudowanego w długości odcinków sieci. Dla każdej kategorii oraz zmiennej warunkującej prędkość zastosowano inne parametry w funkcjach logitowych obliczając w ten sposób spadek prędkości pojazdu następujący w wyniku wystąpienia ograniczeń w prędkości⁹². Dla transportu kolejowego zmienną warunkującą prędkość we wskaźniku WMDT były prędkości techniczne uzyskiwane na sieci kolejowej. Czasy podróży między dowolną parą powiatów były obliczane w transporcie drogowym i kolejowym za pomocą metody odnajdywania najkrótszych ścieżek przejazdu między dowolną parą miejscowości węzłowych. Wskaźnik WMDT, będący wskaźnikiem multimodalnym uwzględnia również inne środki transportu, takie jak transport lotniczy lub transport wodny śródlądowy. Ich rola w ogólnej pracy przewozowej wykonywanej na terytorium Polski jest jednak relatywnie niewielka. Najwyższą dostępnością w Polsce charakteryzują się obszary położone na Górnym Śląsku oraz w centralnej Polsce. Nieco niższa dostępność cechuje tereny wzdłuż popularnej tzw. gierkówki, autostrady A2 do Poznania oraz A4 do Wrocławia (w kierunku zachodnim) i Krakowa (w kierunku wschodnim). Najniższa dostępność potencjałowa cechuje Suwalszczyznę i Bieszczady. Biorąc pod uwagę województwa najgorszą dostępność w skali kraju posiada, relatywnie dobrze dostępne w skali europejskiej województwo zachodniopomorskie.

2.4. Funkcja oporu przestrzeni oraz ogólny wzór na dostępność potencjałową

W teorii różnice między dostępnością potencjałową a kumulatywną (przy założeniu dla dostępności kumulatywnej maksymalnej izochrony wynoszącej 45 minut oraz początkowych izochron wytyczonych na 15 i 30 minut) przedstawiono na rycinie 5.

Ryc. 5. Waga masy celu podróży w dostępności potencjałowej i kumulatywnej wraz z wydłużającym się czasem podróży (dla $t=0$ atrakcyjność wynosi 100%)



Źródło: opracowanie własne.

Dla dostępności kumulatywnej (wykres z prawej) cel podróży jest tak samo atrakcyjny dla podróżującego niezależnie od odległości między źródłem podróży, skąd rozpoczyna się podróż, a jej celem. Zakłada się jednak zawsze, że badanie ma ograniczony wymiar przestrzenny i badamy cele podróży znajdujące się w obrębie konkretnej izochrony. Na rycinie 5 jest to izochrona 45 minut. Wszystkie cele podróży w obrębie tej izochrony są jednakowo istotne dla podróżującego, ale już te położone w odległości czasowej powyżej 45 minut zakłada się, że nie są ujęte w analizie dostępności.

⁹² Szerzej o modelu prędkości ruchu [w:] T. Komornicki, P. Śleszyński, P. Rosik, W. Pomianowski, 2010 oraz jego najnowszej wersji [w:] P. Rosik, W. Pomianowski, M. Stępnia, T. Komornicki, P. Śleszyński, 2011.

Z kolei przy dostępności potencjalowej atrakcyjność celu podróży (mierzona np. liczbą ludności albo dochodem) maleje wraz ze zwiększaniem się czasu podróży. O ile cel podróży znajdujący się niedaleko miejsca rozpoczęcia podróży (źródła podróży) charakteryzuje się wagą równą około 100%, o tyle każdy kolejny kwadrans wydłużający podróż skutkuje zmniejszeniem atrakcyjności celu podróży poprzez odpowiednie obniżenie jego masy np. liczby ludności albo dochodu (por. ryc. 5, wykres z lewej oraz ryc. 6). Szybkość spadku atrakcyjności celu podróży jest określona przez tzw. funkcję oporu przestrzeni. Jako funkcję oporu przestrzeni wykorzystuje się przykładowo funkcje potęgowe lub wykładnicze. We wskaźniku WMDT zastosowano funkcję $1/t$ charakteryzująca się szybkim spadkiem atrakcyjności celu podróży wraz z wydłużającym się czasem podróży. Z kolei we wskaźniku WDDT (opisanym w dalszej części opracowania) wykorzystuje się funkcje wykładnicze (funkcja eksponencjalna). Eksponencjalna funkcja oporu przestrzeni jest w ogólnym wzorze dostępności potencjalowej określona jako:

$$f_{dd} = \exp(-\beta t_{ij}) \quad (2)$$

gdzie:

f_{dd} – funkcja oporu przestrzeni,

t_{ij} – czas podróży/przewozy między rejonem komunikacyjnym (np. powiatem lub gminą) i i j ,

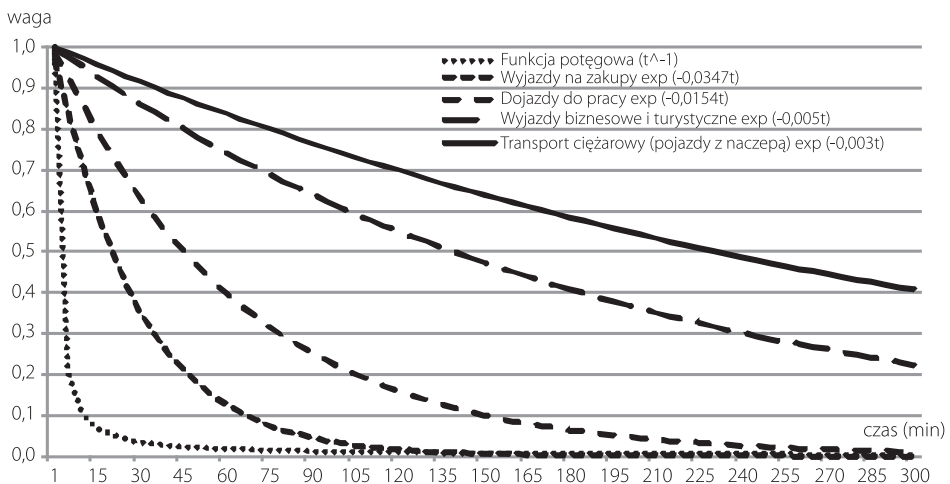
β – parametr *beta* (przykładowe wysokości parametru w zależności od motywacji podróży w tab. 1).

Tab. 1. Przeciętny czas podróży w zależności od motywacji podróży a parametr *beta* w funkcji oporu przestrzeni

Motywacja podróży	Przeciętny czas (odległość) podróży			Przykładowa wysokość parametru <i>beta</i>
	Warszawskie Badanie Ruchu [min]	Krajowy Model Ruchu [km]	Model OGAM 2011 (min)	
Wyjazdy na zakupy	28-30	-	20	0,0347
Dojazdy do pracy	38	56	45	0,0154
Wyjazdy służbowe i turystyczne	-	132-135	140	0,0050
Transport ciężarowy (pojazdy z przyczepami)	-	230	230	0,0030

Źródło: opracowanie własne na podstawie *Krajowy Model Ruchu*, 2008, opracowany przez Politechnikę Warszawską na zlecenie GDDKiA i *Warszawskie Badanie Ruchu*, 2005, Biuro Planowania Rozwoju Warszawy na zlecenie Biura Naczelnego Architekta Miasta st. Warszawy.

Ryc. 6. Funkcja oporu przestrzeni dla czterech motywacji podróży (funkcja wykładnicza) oraz dla funkcji potęgowej (t^{-1})



Źródło: P. Rosik, W. Pomianowski, M. Stępnik, T. Komornicki, P. Śleszyński, 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich*.

Przy uwzględnieniu potencjału własnego oraz funkcji wykładniczej jako funkcji oporu przestrzeni ogólny wzór na dostępność potencjałową w rejonie komunikacyjnym i (np. gminie lub powiecie i) przyjmuje postać:

$$A_i = M_i \exp(-\beta c_{ii}) + \sum_j M_j \exp(-\beta c_{ij}) \quad (3)$$

gdzie:

A_i – dostępność potencjałowa rejonu komunikacyjnego i ,

M_i – masa własna rejonu komunikacyjnego i ,

c_{ii} – czasy podróży wewnętrznej w rejonie komunikacyjnym i ,

M_j – masa rejonu komunikacyjnego j ,

c_{ij} – czasy podróży między rejonami komunikacyjnymi i oraz j ,

β – parametr *beta*.

2.5. Dostępność potencjałowa gmin (w ramach aplikacji OGAM)

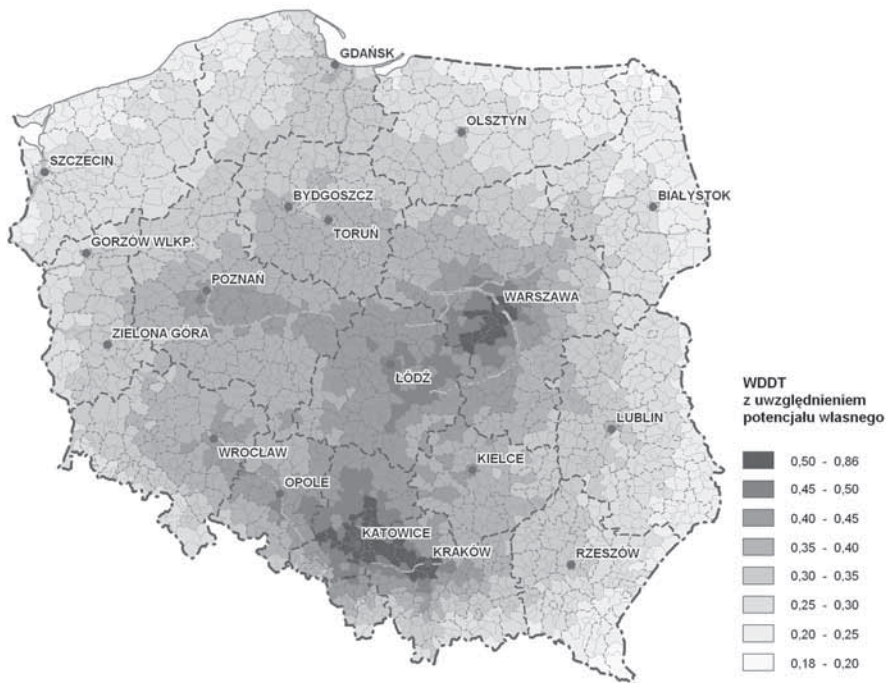
Istotną cechą opracowanego w 2011 r. w IGiPZ PAN narzędzia w postaci aplikacji komputerowej OGAM jest jego pełna kompatybilność z opisanym w poprzednim rozdziale i obecnie wykorzystywanym przez Ministerstwo Rozwoju Regionalnego Wskaźnikiem Międzygałęziowej Dostępności Transportowej (WMDT). Aplikacja OGAM jest uszczegółowieniem wcześniejszych narzędzi. Daje ona możliwości większej szczegółowości terytorialnej badania, a także pozwala na większą elastyczność w zakresie wykorzystywanych danych oraz formuł ich przetwarzania. Przy zastosowaniu podstawowych założeń obliczeniowych uzyskany wynik (na poziomie krajowym i regionalnym) jest niemalże tożsamy z wynikiem otrzymanym dla gałęzi transportu drogowego w dotychczas wykorzystywanej aplikacji i wskaźnika WMDT (ryc. 7)⁹³. Otrzymany w ten sposób Wskaźnik Drogowej Dostępności Transportowej (WDDT) stanowi zatem jedynie rozwinięcie idei wskaźnika dostępności potencjałowej przy nieco odmiennych założeniach badawczych oraz bazowaniu na wzorze ogólnym na dostępność potencjałową (3).

Wskaźnik WDDT na poziomie gmin przyjmuje podobny rozkład jak wskaźnik WMDT na poziomie powiatów. Najwyższą dostępnością w Polsce charakteryzują się obszary położone na Górnym Śląsku oraz w obszarze metropolitalnym Warszawy. Nieco niższa dostępność cechuje tereny wzdłuż popularnej tzw. gierkówki, autostrady A2 do Poznania oraz A4 do Wrocławia (w kierunku zachodnim) i Krakowa (w kierunku wschodnim). Wyższa dostępność rejonów gęsto zaludnionych, w tych miast wynika z uwzględnienia w analizie tzw. potencjału własnego tych miast, który określa potencjał celów podróży znajdujących się w tej samej gminie co źródło podróży.

Aplikacja OGAM daje możliwości dowolnego sterowania parametrami modelu prędkości ruchu, funkcji oporu przestrzeni (funkcja mówiąca o tym, jak szybko spada atrakcyjność celu podróży wraz ze zwiększaniem się czasu podróży), a także wyboru dowolnej masy docelowej (np. liczba miejsc pracy, liczba łóżek szpitalnych lub liczba studentów). W celu możliwości porównań między wynikami symulacji dla różnych funkcji oporu przestrzeni założono, że masą docelową jest po prostu liczba ludności. Tym samym wszystkie dalsze mapy prezentowane z wykorzystaniem wskaźnika WDDT można utożsamiać z tzw. **potencjałem demograficznym**. Zastosowanie różnych **funkcji oporu przestrzeni** daje możliwość zdiagnozowania drogowej dostępności potencjałowej dla wyjazdów krótkich (np. wyjazdy na zakupy), średnich (np. dojazdy do pracy) lub długich (np. wyjazdy służbowe i turystyczne). Wartości wynikowych wskaźników nie można porównywać między poszczególnymi typami wyjazdów. Kolejne mapy służą prezentacji różnic w rozkładzie dostępności na terytorium Polski dla każdego z typów podróży z osobna.

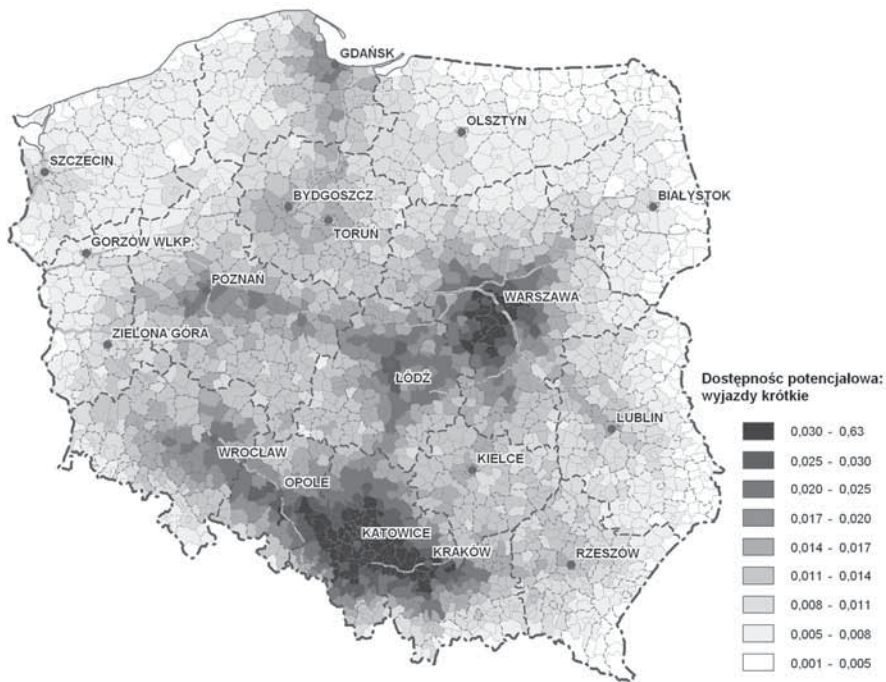
⁹³ Różnice między wynikami otrzymanymi w WMDT a tymi z WDDT dotyczą faktu, że nastąpiła dezagregacja (z poziomu powiatowego na poziom gminny), zrezygnowano z agregacji w obszarach metropolitalnych oraz faktu, że wskaźnik dotyczy transportu drogowego i nie jest wskaźnikiem multimodalnym. Zespół IGiPZ PAN nie wyklucza zmiany tej nazwy w najbliższej przyszłości w celu dokładniejszego oddania specyfiki nowego wskaźnika (lub dokładniej zespołu wskaźników).

Ryc. 7. Wskaźnik Drogowej Dostępności Transportowej (z uwzględnieniem tzw. potencjału własnego)



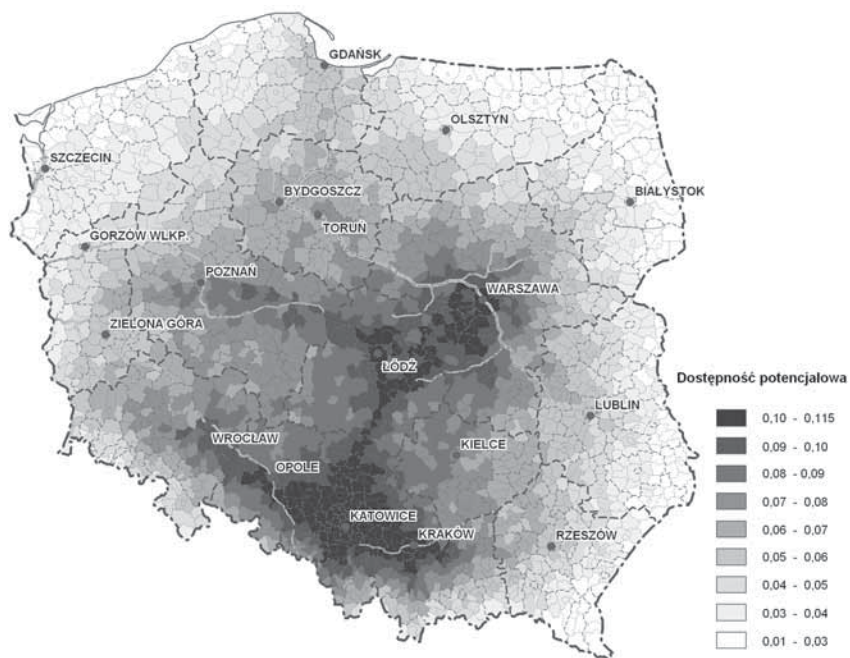
Źródło: P. Rosik, W. Pomianowski, M. Stępnia, T. Komornicki, P. Śleszyński, 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich.*

Ryc. 8. Dostępność potencjałowa dla wyjazdów krótkich (np. wyjazdy na zakupy)



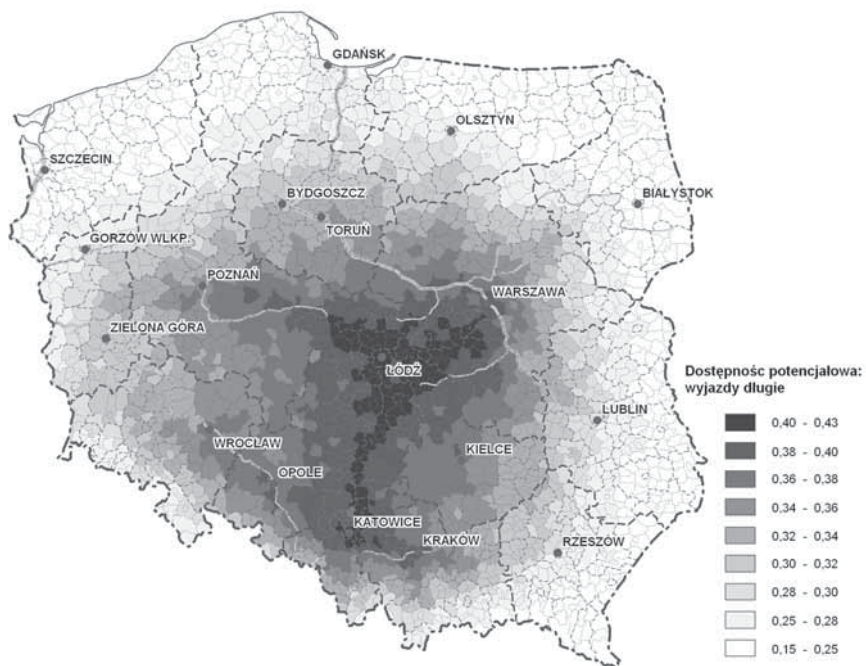
Źródło: P. Rosik, W. Pomianowski, M. Stępnia, T. Komornicki, P. Śleszyński, 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich.*

Ryc. 9. Dostępność potencjalowa dla wyjazdów średnich (np. dojazdów do pracy)



Źródło: P. Rosik, W. Pomianowski, M. Stępnik, T. Komornicki, P. Śleszyński, 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich.*

Ryc. 10. Dostępność potencjalowa dla wyjazdów długich (np. wyjazdy służbowe i turystyczne)



Źródło: P. Rosik, W. Pomianowski, M. Stępnik, T. Komornicki, P. Śleszyński, 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich.*

Rozkład dostępności dla dojazdów średniego zasięgu, np. dojazdów do pracy jest najbardziej zbliżony do rozkładu dostępności uzyskanego przy wskaźniku WDDT. Natomiast wyjazdy długie, do których należą przykładowo wyjazdy biznesowe i turystyczne cechuje koncentryczny rozkład dostępności.

3. Wykorzystanie analiz dostępności transportowej do badania efektów inwestycji na sieci drogowej (ewaluacja *ex-post* i *ex-ante*)

Dostępność mierzona za pomocą izochron, dostępność kumulatywna, a także, a może w szczególności, dostępność potencjałowa były przez IGIPZ PAN wielokrotnie wykorzystywane zarówno przy ewaluacji *ex-post* jak i *ex-ante* inwestycji transportowych.

W przypadku dostępności potencjałowej oszacowanie wpływu inwestycji na zmiany dostępności na poziomie gmin, powiatów, województw oraz całego kraju w wyniku realizacji poszczególnych inwestycji następuje poprzez zmianę kategorii poszczególnych odcinków sieci transportowych. W przypadku modernizacji istniejących dróg użytkownik aplikacji ma możliwość zmiany kategorii drogi, np. z drogi krajowej na drogę ekspresową, jeżeli jest ona poprowadzona „po starym śladzie” (model sam wylicza wówczas zmianę prędkości na zaznaczonym odcinku) lub z kategorii „planowanej” na „istniejącą” jeżeli autostrada lub droga ekspresowa jest prowadzona „po nowym śladzie” (nowy odcinek drogi wówczas bierze udział w obliczaniu najkrótszych ścieżek przejazdu i tym samym w utworzeniu nowej macierzy czasów przejazdu pomiędzy wszystkimi analizowanymi gminami bądź powiatami, co wpływa na wysokość wskaźnika). Każda symulacja polega na zbadaniu procentowej zmiany dostępności jednostek pomiędzy dwoma stanami: stanem bazowym – przy założeniu, że budowana droga nie istnieje lub modernizowana droga charakteryzuje się parametrami przed modernizacją oraz stanem końcowym, po oddaniu nowej lub modernizowanej drogi do użytkowania. Symulacje są możliwe zarówno dla jednego odcinka sieci, jak i wielu odcinków, a nawet wszystkich odcinków realizowanych z danego rodzaju funduszu lub w ramach danej perspektywy finansowej.

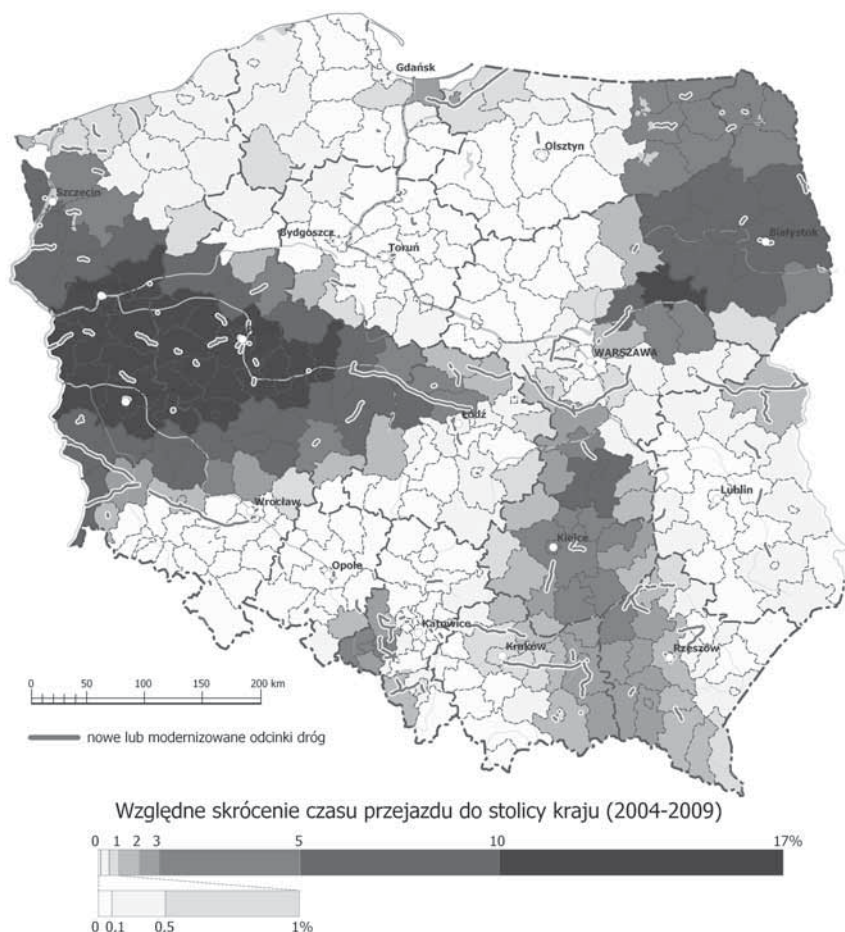
Jak już wspomniano, wyniki symulacji (np. procentowa zmiana dostępności) dają możliwość analizy porównawczej zmian dostępności na poziomie gminy (tylko WDDT), powiatu (tylko WMDT), województwa lub kraju w wyniku realizacji poszczególnych inwestycji. Jest zatem możliwa priorytetyzacja przedsięwzięć transportowych w zależności od ich wpływu na zmiany dostępności na poszczególnych poziomach analizy.

3.1. Ewaluacja *ex post* NPR 2004-2006

Skrócenie czasu podróży do stolicy

W opracowaniu *Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach polityki spójności na wzrost konkurencyjności regionów (MRR, 2010b)* dokonano ewaluacji inwestycji transportowych zrealizowanych w perspektywie finansowej 2004-2006. Po opracowaniu dwóch macierzy czasów przejazdu (przed i po realizacji wybranych inwestycji transportowych) dla celów ewaluacyjnych zaistniała możliwość porównania czasu przejazdu do dowolnego celu podróży, np. miasta przed i po realizacji tychże inwestycji, co daje ciekawe możliwości, szczególnie w zakresie wizualizacji wyników (ryc. 11).

Ryc. 11. Skrócenie czasu przejazdu do Warszawy w wyniku realizacji projektów w ramach perspektywy finansowej 2004-2006 (%)



Źródło: MRR, 2010b, *Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach polityki spójności na wzrost konkurencyjności regionów (w ramach ewaluacji ex post NPR 2004-2006)*.

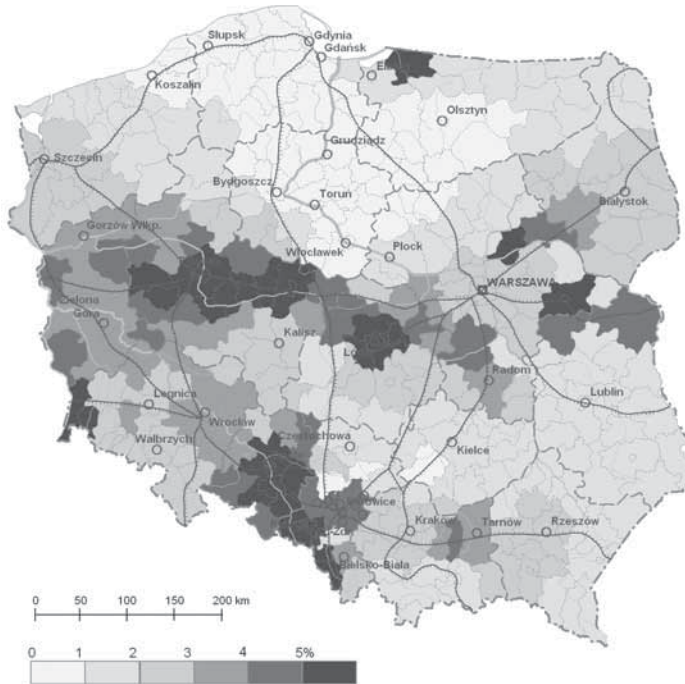
Widać wyraźnie, że w zakresie dojazdów do Warszawy efekty inwestycji drogowych realizowanych w perspektywie finansowej 2004-2006 układają się w formie trzech wachlarzy, które są efektem głównie trzech inwestycji: autostrady A2 na odcinku Konin–Stryków, drogi ekspresowej S7 na odcinku Białobrzegi–Jedlińsk oraz drogi ekspresowej S8 na odcinku Radzimin–Wyszaków.

Wyniki te pozwalają wnioskować, że lokalizacja inwestycji ma istotne znaczenie z punktu widzenia maksymalizacji efektów. Np. inwestycja zrealizowana w obszarze aglomeracji może przynosić wymierną korzyść niektórym obszarom peryferyjnym. Z drugiej strony, działania w samych obszarach peryferyjnych mogą mieć ograniczony efekt, gdyż nie wiążą ich z obszarami lepiej rozwiniętymi. Dlatego też w przypadku dróg krajowych należy rozpoczynać inwestycje na drogach dojazdowych do największych miast i dalej koncentrować się na eliminacji „wąskich gardeł” komunikacyjnych (poprzez budowę wielu tras jednocześnie). W przypadku inwestycji kolejowych na poziomie krajowym – konieczna jest koncentracja na kilku najważniejszych szlakach, z założeniem ich kompleksowej realizacji na całej długości. Przyjęcie innej strategii inwestycyjnej może spowodować, że efekty projektów nie będą zauważalne, a kierunek przesunięć modalnych może być odwrotny od zakładanego (MRR, 2011c).

Sumaryczna analiza efektów wszystkich inwestycji w ramach perspektywy finansowej 2004-2006

W ramach tej samej ewaluacji został wykorzystany wskaźnik WMDT m.in. do zbadania zmian jakie nastąpiły w dostępności potencjałowej powiatów w wyniku realizacji inwestycji infrastrukturalnych w ramach perspektywy finansowej 2004-2006 (były to inwestycje realizowane przede wszystkim w ramach Funduszu Spójności oraz SPOT) (ryc. 12).

Ryc. 12. Zmiana WMDT w związku z realizacją inwestycji w perspektywie finansowej 2004-2006



Źródło: MRR, 2010b, *Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach polityki spójności na wzrost konkurencyjności regionów (w ramach ewaluacji ex post NPR 2004-2006)*.

W wyniku analizy otrzymano następujące wnioski. Zdecydowanie najbardziej znacząca poprawa dostępności związana z inwestycjami wspartymi ze środków Unii Europejskiej miała miejsce w pasie od granicy niemieckiej poprzez niemal całe województwo lubuskie, następnie obszar od Poznania do Warszawy oraz dalej po granicę z Białorusią. Drugi zwarty obszar znacznych przyrostów wskaźnika WMDT znajduje się w rejonie Opolszczyzny i zachodniej części konurbacji górnośląskiej. Ponadto poprawa miała miejsce na niektórych obszarach województwa dolnośląskiego, w rejonie Tarnowa, na północno-wschodnim Mazowszu oraz częściowo na Podlasiu, a także w powiecie braniewskim. O takim rozkładzie przesądziły konkretne inwestycje, przede wszystkim realizowane w ramach Funduszu Spójności (m.in. fragmenty autostrad A2 i A4, droga ekspresowa S8 między Radzyminem a Wyszkowem, modernizacja dróg krajowych nr 2 (Siedlce–Terеспol) i nr 4 (między Krakowem a Tarnowem), a także modernizacja linii kolejowej (Warszawa–Terеспol).

Powyższe wyniki pokazują, że koncentracja wsparcia na inwestycjach mających największe znaczenie z punktu widzenia spójności terytorialnej kraju i regionów oraz poprawy dostępności przestrzennej (czyli inwestycjach w ciągach korytarzy transportowych oraz przebudowa tras wylotowych z dużych miast), skorelowanych z potrzebami i mobilnością mieszkańców, daje znaczącą poprawę dostępności transportowej (osiągnięcie „masy krytycznej”). Jednocześnie widoczne jest słabe oddziaływanie inwestycji rozproszonych (MRR, 2010a).

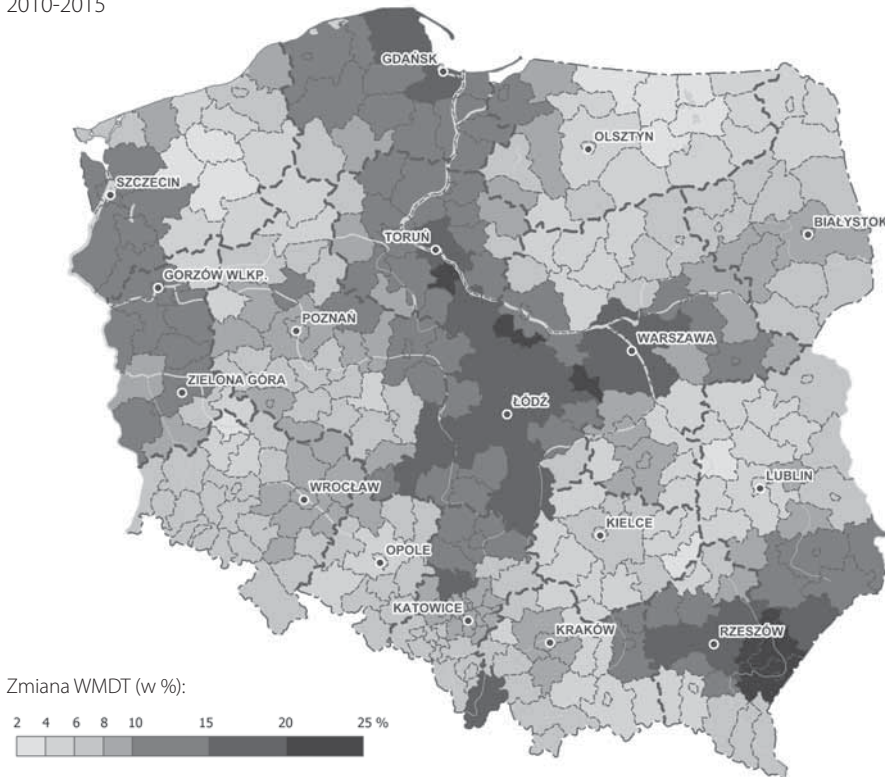
3.2. Ewaluacja *ex-ante*

Sumaryczna analiza efektów wszystkich planowanych inwestycji drogowych do 2015 roku

W przypadku ewaluacji *ex-ante* warto zwrócić uwagę na ekspertyzę wykonaną przez zespół IGiPZ PAN dotyczącą m.in. prawdopodobnych zmian dostępności drogowej w wyniku inwestycji, jakie będą miały miejsce na sieci dróg ekspresowych i autostrad w latach 2010-2015 oraz 2015-2020. Ewaluacja bazuje na założeniach wykorzystanych przy wskaźniku WMDT, również z podziałem na podsięci.

Najwyższe zmiany dotyczą relatywnie słabo w 2010 r. dostępnych powiatów wschodniego Podkarpacia położonych wzdłuż autostrady A4. Rycina pokazuje, jak ważny jest wschodni odcinek autostrady A4 nie tylko dla ruchu tranzytowego w kierunku Ukrainy, ale również dla mieszkańców okolic Przemyśla. Poprawa dostępności nastąpi na zwartym obszarze w Polsce południowo-wschodniej. Również dla relatywnie słabo dostępnego w skali kraju województwa pomorskiego budowa autostrady A1 od Nowych Marz koło Grudziądza w kierunku centralnej Polski (Łódź) oraz dalej, aż do granicy z Czechami daje ogromny wzrost dostępności województwa (szczególnie obszarów położonych wzdłuż Wisły). Oddanie do użytku odcinka od Grudziądza do Torunia w 2011 roku niewątpliwie już polepszyło sytuację woj. pomorskiego. Dostępność południowo-zachodnich obszarów aglomeracji warszawskiej, aż do granicy z województwem łódzkim wzrosła do 2015 r. dzięki nowemu połączeniu autostradowemu poprzez autostradę A2 Stryków–Konotopa z Łodzią i przede wszystkim z Warszawą. W województwie łódzkim z kolei wzrost dostępności jest w miarę równo rozłożony na obszarze całego województwa. Wynika to z faktu realizacji wielu inwestycji, takich jak wspomniana A4 oraz

Ryc. 13. Zmiana WMDT (podsięć drogową) w związku z realizacją inwestycji w perspektywie finansowej 2010-2015



Źródło: T. Komornicki, P. Rosik, M. Stępnia, W. Pomianowski, 2011b, *Symulacja Wskaźnika Międzygałęzowej Dostępności Transportowej (WMDT) na terytorium Polski do roku 2020.*

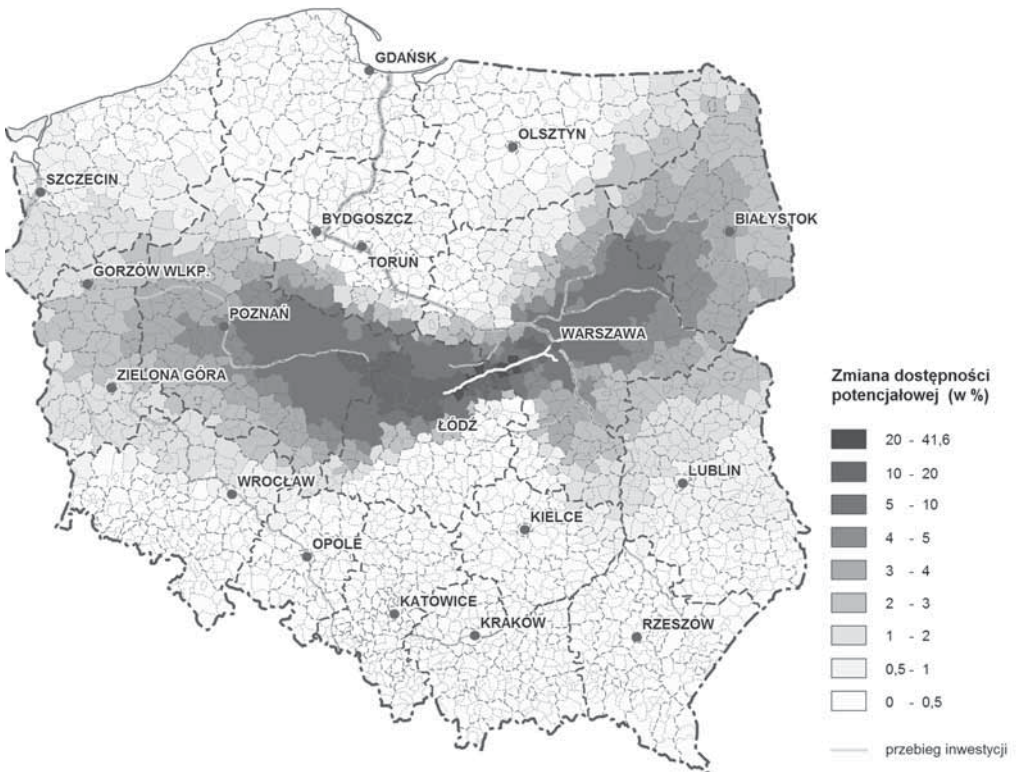
A1, ale również droga ekspresowa S8 realizowana do 2015 roku na całej długości swojego przebiegu przez województwo łódzkie. Ponadto województwo łódzkie z uwagi na swoje centralne położenie jest po części beneficjentem wielu innych inwestycji położonych na terenie innych regionów. W Polsce Zachodniej wyraźnie widoczny jest efekt wzrostu dostępności wzdłuż drogi ekspresowej S3. Jednocześnie inwestycje realizowane obecnie i w najbliższej przyszłości nie przyniosą praktycznie żadnych efektów w poprawie dostępności w takich regionach jak: województwo świętokrzyskie, Karpaty, wschodnia część warmińsko-mazurskiego, północna Lubelszczyzna oraz Pomorze Środkowe.

Symulacje efektów pojedynczych inwestycji drogowych realizowanych w perspektywie finansowej 2007-2013

Nowa aplikacja o charakterze otwartym OGAM zbudowana w IGIPZ PAN w 2011 roku pozwala analizować zmiany dostępności potencjałowej w wyniku realizacji poszczególnych inwestycji drogowych na poziomie gminnym (ale również wojewódzkim lub krajowym). Symulacje zmian potencjałowej dostępności drogowej są również możliwe przy wskaźniku WMDT. Nie są one jednak wówczas tak dokładne ponieważ dotyczą poziomu powiatów a nie gmin, jak przy wskaźniku WDDT w aplikacji OGAM.

W niniejszym opracowaniu wybrano dwie takie inwestycje, których ukończenie planowane jest w latach 2007-2013⁹⁴. Należą do nich m.in. fragment autostrady A2 Stryków-Konotopa wraz z odcinkami dróg ekspre-

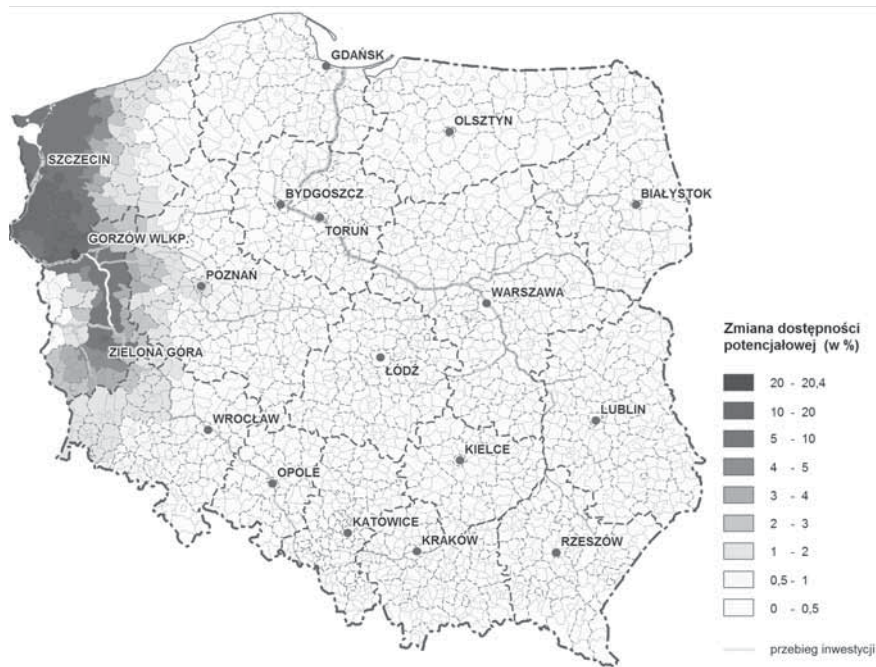
Ryc. 14. A2 Stryków-Konotopa + droga ekspresowa S8 Konotopa-Marki i S2 Konotopa-Puławska



Źródło: P. Rosik, W. Pomianowski, M. Stępnik, T. Komornicki, P. Śleszyński, 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich*.

⁹⁴ Zmiany dostępności w wyniku realizacji tych inwestycji są relatywnie wyższe niż przy zastosowaniu wskaźnika WMDT, co wynika z innej funkcji oporu przestrzeni.

Ryc. 15. S3 Gorzów Wlkp. – Sulechów. Zmiana dostępności wewnętrznej krajowej



Źródło: P. Rosik, W. Pomianowski, M. Stępiak, T. Komornicki, P. Śleszyński, 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich*.

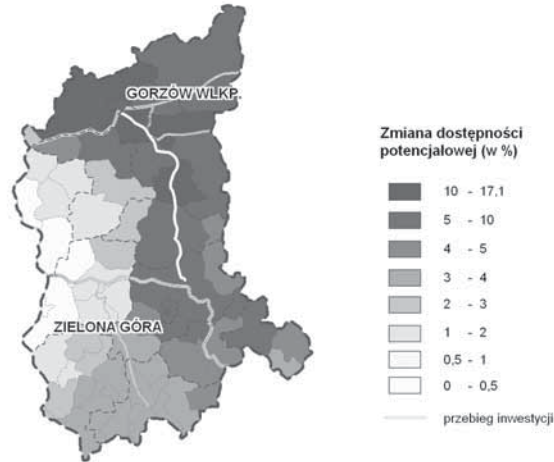
sowych S8 Konotopa–Marki i S2 Konotopa–Puławska oraz fragment drogi ekspresowej S3 Gorzów Wlkp.–Sulechów.

Autostrada A2 łącząca Łódź (Stryków) z Warszawą (Konotopa) daje duże zmiany dostępności w skali kraju na rozległych obszarach tworzących dwa wachlarze: zachodni i wschodni. Zachodni wachlarz zmian rozciąga się od Szczecina i południowych obszarów województwa lubuskiego, przez województwo wielkopolskie oraz południowe fragmenty kujawsko-pomorskiego i północny Dolny Śląsk, aż do zachodnich i północnych fragmentów województwa łódzkiego. Wschodni wachlarz rozpościera się na całe województwo podlaskie, a także północne obszary województwa lubelskiego i wschodnie warmińsko-mazurskiego. Głównymi beneficjentami inwestycji są gminy zlokalizowane na północy województwa łódzkiego, środkowego Mazowsza oraz wschodniej Wielkopolski.

Północny odcinek drogi ekspresowej S3 łączący Szczecin z Gorzowem Wielkopolskim został oddany na całym swoim przebiegu do użytku w grudniu 2010 roku. Analizowany odcinek łączący Gorzów Wlkp. z Sulechowem ma ogromne znaczenie przede wszystkim dla mieszkańców województw zachodniopomorskiego i lubuskiego.

Aplikacja OGAM umożliwia również analizę zmian dostępności w układzie wewnątrz regionalnym (przy kontrafaktycznym założeniu, że każde województwo jest spójną całością i nie ma możliwości wyjazdu poza jego granicę). Przykładową symulację efektu drogi ekspresowej S3 Gorzów Wielkopolski–Sulechów na zmianę dostępności wewnętrznej województwa lubuskiego przedstawiono na ryc. 16.

Ryc. 16. S3 Gorzów Wlkp. – Sulechów. Zmiana dostępności wewnętrznej wojewódzkiej w województwie lubuskim



Źródło: P. Rosik, W. Pomianowski, M. Stępiński, T. Komornicki, P. Śleszyński, 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich*.

4. Wskazanie innych przykładowych wskaźników, które powinny być brane pod uwagę przy planowaniu inwestycji drogowych

Dostępność nie jest jedynym wskaźnikiem mogącym znaleźć zastosowanie w ewaluacji projektów transportowych (w tym drogowych). W *Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach polityki spójności na wzrost konkurencyjności regionów (MRR, 2010b)* autorzy wykorzystali m.in. dwa inne wskaźniki, tj. wskaźnik wąskich gardeł transportowych, nazwany WWGT oraz wskaźnik wypadkowości i kolizyjności (WWiK). Wskaźniki te, podobnie jak wskaźniki dostępności potencjalowej, mogą być wykorzystywane przy ewaluacji *ex-post* oczywiście jedynie wówczas gdy odpowiednie instytucje (Generalna Dyrekcja Dróg Krajowych i Autostrad, Komenda Główna Policji) udostępnią dane dotyczące natężenia ruchu (pomiar ruchu na sieci dróg krajowych i wojewódzkich wykonywany co 5 lat) oraz wypadkowości i kolizyjności (dane w ujęciu rocznym). W odróżnieniu jednak od wskaźnika dostępności potencjalowej stosowanie w ewaluacji wskaźników WWGT oraz WWiK uniemożliwia dokonanie oceny efektów poszczególnych inwestycji oraz poszczególnych wariantów czy strategii. Możliwe jest jedynie porównanie *ex post* zmiany obu wskaźników.

4.1. Eliminacja wąskich gardeł i kongestii

Pierwszy ze wskaźników pokazuje, jak duże znaczenie ma eliminacja wąskich gardeł transportowych (*bottlenecks*), która polega na likwidacji zatorów w ruchu, jakie pojawiają się na danym odcinku sieci o niskiej przepustowości w warunkach kongestii. Rozwiązaniem problemu są inwestycje charakteryzujące się dużym wzrostem przepustowości. Wskaźnik Wąskich Gardel Transportowych (WWGT), powstaje poprzez podzielenie obliczeniowego natężenia ruchu na odcinku drogi przez przepustowość tego odcinka⁹⁵. W ramach

⁹⁵ Warto wskazać, że w projektach związanych z bezpieczeństwem ruchu należy liczyć się z ograniczeniem przepustowości. Mowa tu o projektach, których istotą jest przebudowa lub wzmocnienie jednojezdniowej drogi (spełnienie warunku dostosowania nawierzchni drogi do nacisku 11,5 t/oś), gdzie często dodatkowo występują spowolnienia w postaci np. linii ciągłych lub wysepek. Tym samym występuje sprzeczność w realizacji zadania likwidacji wąskich gardeł oraz zadania poprawy bezpieczeństwa ruchu.

obliczeniowego natężenia ruchu uwzględniono średniodobowe natężenie ruchu pojazdów osobowych, pojazdów ciężarowych lekkich (dostawczych) oraz pojazdów ciężarowych ciężkich (z przyczepami i bez przyczep), autobusów i ciągników z 2005 roku. W dalszej kolejności niezbędne stało się obliczenie przepustowości na poszczególnych kategoriach dróg zamiejskich w zależności od ich przekroju. Wskaźnik powstaje według wzoru:

$$WWGT_i = \frac{Q_i}{P_i} \quad (4)$$

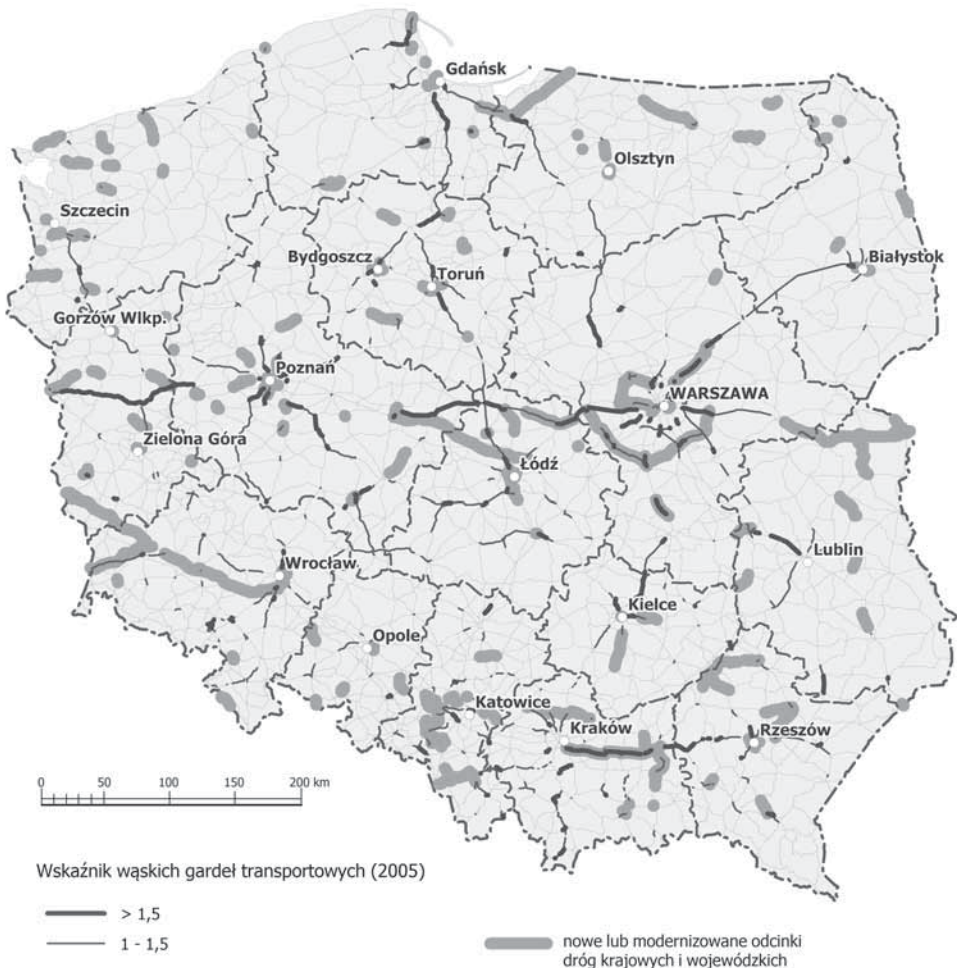
gdzie:

$WWGT_i$ – wskaźnik wąskich gardeł transportowych na odcinku i ,

Q_i – obliczeniowe natężenie ruchu na odcinku i ,

P_i – przepustowość na odcinku i .

Ryc. 17. Wskaźnik Wąskich Gardel Transportowych (WWGT) na sieci zamiejskich dróg krajowych i wojewódzkich oraz inwestycje drogowe realizowane na tej sieci w ramach perspektywy finansowej 2004-2006



Źródło: Ministerstwo Rozwoju Regionalnego, 2010(b), *Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach polityki spójności na wzrost konkurencyjności regionów.*

Założono, że wąskie gardła transportowe powstają na tych odcinkach sieci, na których obliczeniowe natężenie ruchu jest wyższe od przepustowości. Tym samym wąskim gardłem można nazwać odcinek, na którym wskaźnik WWGT przyjmuje wartości wyższe od jedności. Dla wskaźnika WWGT wyższego od 1,5 można natomiast mówić o krytycznym wąskim gardle. Na rycinie 17 zobrazowano wąskie gardła na sieci dróg krajowych i wojewódzkich w kontekście lokalizacji nowych lub modernizowanych odcinków sieci w ramach perspektywy finansowania 2004-2006.

Na podstawie wyników wskaźnika WWGT można wnioskować, że wśród zrealizowanych projektów znalazło się zdecydowanie za mało inwestycji podnoszących w sposób wyraźny przepustowość (niewielki udział dróg dwujezdniowych). Charakterystyczna jest niewielka liczba inwestycji w obszarach metropolitalnych, w szczególności na drogach dojazdowych do dużych miast, gdzie znajduje się najwięcej wąskich gardeł, w tym również tych krytycznych, w których obliczeniowe natężenie ruchu znacznie przewyższa przepustowość.

4.2. Poprawa poziomu bezpieczeństwa ruchu

Poprawa bezpieczeństwa ruchu jest jednym z najważniejszych kryteriów oceny inwestycji infrastrukturalnych w transporcie. Budowa i utrzymanie bezpiecznej infrastruktury drogowej jest jednym z podstawowych (obok warunków topograficznych, zagospodarowania urbanistycznego, warunków pogodowych, parku samochodowego, populacji i indywidualnych cech uczestników ruchu) czynników wpływających na bezpieczeństwo ruchu. Mając na uwadze, że jednostkowy koszt osoby rannej jest ponad czterokrotnie mniejszy niż jednostkowy koszt osoby ofiary śmiertelnej (Instytut Badawczy Dróg i Mostów, 2008), natomiast skutki wypadku (udział rannych lub zabitych) są wielokrotnie gorsze niż skutki kolizji, skonstruowano autorski ważony Wskaźnik Wypadkowości i Kolizyjności (WWiK) obliczony według danych Komendy Głównej Policji, według wzoru:

$$W_{\text{Li}} = \frac{4 \times W_i + 8 \times Z_i + 2 \times R_i + K_i}{L_i} \quad (5)$$

gdzie:

W_{Li} – makro wskaźnik wypadkowości i kolizyjności w roku i (2004 i 2009),

W_i – liczba wypadków w powiecie w roku i (2004 i 2009),

Z_i – liczba zabitych w powiecie w roku i (2004 i 2009),

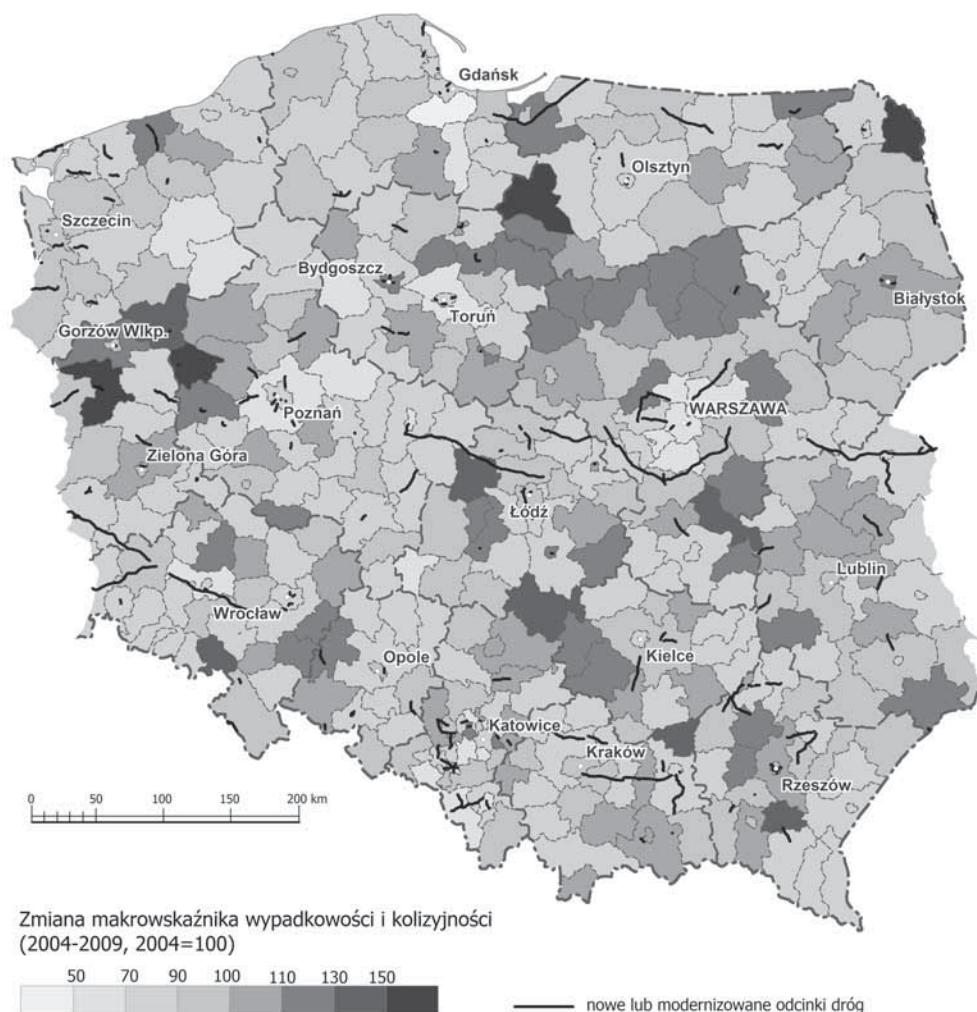
R_i – liczba rannych w powiecie w roku i (2004 i 2009),

K_i – liczba kolizji w powiecie w roku i (2004 i 2009),

L_i – liczba mieszkańców powiatu w roku i (w tys.) (2004 i 2009).

Charakterystyczny jest fakt znacznego spadku wskaźnika WWiK na terenach większości miast grodzkich, przede wszystkim w dużych miastach i aglomeracjach (z wyjątkiem Krakowa) – por. ryc. 18. Obszary te nadal dominują w liczbach bezwzględnych kolizji i wypadków, jednak sytuacja w zakresie poprawy bezpieczeństwa na drogach znacznie się poprawiła. Z pewnością jedną z przyczyn są inwestycje infrastrukturalne, zarówno te drogowe, jak i w transporcie publicznym. Jednak o braku korelacji między nakładami na drogi w ramach perspektywy finansowej 2004-2006 a wypadkowością i kolizyjnością w powiatach mogą świadczyć również te powiaty, w których zaobserwowano wzrost wskaźnika WWiK (pogorszenie poziomu bezpieczeństwa ruchu). Grupy powiatów, w których nastąpił ponadto wzrost wskaźnika WWiK to północna część województwa lubuskiego, północna część województwa mazowieckiego, a także inne grupy powiatów położonych na granicy poszczególnych województw. Wzrost wskaźnika następował zatem raczej

Ryc. 18. Zmiana Wskaźnika Wypadkowości i Kolizyjności (WWiK) a inwestycje na sieci dróg krajowych i wojewódzkich w perspektywie finansowej 2004-2006



Źródło: Ministerstwo Rozwoju Regionalnego, 2010(b), *Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach polityki spójności na wzrost konkurencyjności regionów*

na terenach położonych w dalszej odległości od aglomeracji, zazwyczaj relatywnie uboższych. Możliwym wytłumaczeniem zwiększonej wypadkowości i kolizyjności na tych terenach jest raczej wzrost motoryzacji przy jednoczesnym braku (jak w aglomeracjach) poważnych korków ulicznych skutkujących niższą prędkością ruchu, a co za tym idzie mniejszą wypadkowością.

Można wnioskować, że spadek wypadkowości i kolizyjności jest zawsze wynikiem budowy dwujezdniowych dróg bezkolizyjnych. Natomiast w przypadku inwestycji o mniejszym zasięgu efekt w skali całego kraju jest widoczny zapewne dopiero w dłuższej perspektywie czasowej. Pozytywnie należy oceniać szczególnie te inwestycje, przy których nastąpiło oddzielenie ruchu samochodowego od pieszego i rowerowego (MRR, 2011c).

5. Podsumowanie

Większość ograniczeń narzędzia badania dostępności transportowej zostało przewyżczonych w momencie powstania aplikacji OGAM, która ma charakter otwarty i pozwala na analizę dostępności potencjałowej na dowolnej sieci transportowej oraz porównania dowolnych momentów czasu, co umożliwia zarówno analizę *ex-post* jak i *ex-ante* dowolnej sieci transportowej niezależnie, czy to w skali województwa, kraju, czy może przy założeniu wykonania odpowiedniego podkładu sieci, również w skali europejskiej. Jedyną trudnością jest odpowiednie przygotowanie sieci oraz dostępność danych dotyczących np. czasu przejazdu w przypadku sieci transportu publicznego na poziomie gminnym.

Dzięki przygotowanej sieci drogowej Polski na poziomie gminnym aplikacja komputerowa OGAM umożliwia potencjalnemu użytkownikowi wykonywanie symulacji zmian drogowej potencjałowej dostępności transportowej z punktu widzenia zarówno kraju, jak i regionu w układach wojewódzkich na poziomie gminnym. Staje się zatem możliwa ocena dowolnych inwestycji na sieci transportowej z punktu widzenia ich wpływu na zmiany dostępności gmin w kontekście dostępności krajowej oraz dostępności wewnątrz regionalnej.

Należy wymienić również istotne rezultaty metodyczne w postaci znacznie większego spektrum możliwości wyboru danych wejściowych do symulacji w różnych wariantach (krajowy, regionalny) w ramach aplikacji. Można tu wymienić m.in. możliwość dowolnego parametryzowania modelu prędkości ruchu, importu dowolnej sieci transportowej, importu dowolnej kolumny mas, dowolnego parametryzowania funkcji oporu przestrzeni oraz włączenia parametryzowania tzw. potencjału własnego.

Powyższe możliwości, jakie daje aplikacja, przekładają się na duże spektrum możliwych zastosowań w przyszłości oraz dalszych badań. Są to m.in. wykorzystanie metody ilorazu potencjału (np. na rynku pracy, w zakresie dostępu do szkół wyższych lub szpitali) lub porównywanie wyników dostępności przy różnych założeniach wejściowych (np. przy różnych funkcjach oporu przestrzeni lub różnych prędkościach podróży wewnątrz regionalnych). Ponadto, podobnie jak w przypadku poprzedniej wersji aplikacji istnieje możliwość wykonywania symulacji zmian dostępności w wyniku realizacji inwestycji infrastrukturalnych, z tą różnicą, że generowane wyniki są znacznie dokładniejsze, ponieważ przedstawiane na poziomie gminnym, a nie powiatowym.

Potencjalne możliwości rozbudowy aplikacji są bardzo duże i dotyczą przede wszystkim uwzględnienia: przepływów transportowych (wówczas aplikacja mogłaby służyć przy ewaluacji nie tylko jako model potencjału ale również grawitacji), kosztów transportu, jako elementu oporu przestrzeni (do tej pory tym elementem jest czas podróży) oraz środków transportu (w tym przede wszystkim na poziomie gminnym sieci transportu publicznego uwzględniającej możliwości przesiadki między transportem autobusowym a kolejowym, a także lotniczym). Kolejnym wyzwaniem mogą być również dalsze prace nad motywacjami podróży (np. poprzez rozróżnienie dojazdów do pracy – w postaci badania dostępności do miejsc pracy, a także dojazdów do tzw. *services of general interest* w postaci dojazdów do szpitali, szkół, celów turystycznych itd.).

Kwestią otwartą są możliwości uwzględnienia przy szacowaniu *ex-ante* efektów projektów drogowych uciążliwości związanych z przeprowadzaniem inwestycji (np. z wydłużeniem czasu przejazdu w wyniku remontów), zwłaszcza w przypadku modernizacji istniejących dróg i linii kolejowych – dla autostrad prowadzonych nowym śladem te uciążliwości pozostają z reguły marginalne (MRR, 2001b). W niektórych przypadkach może mieć ona zauważalny, czasowy wpływ na wskaźnik dostępności transportowej oraz generować straty czasu na tyle istotne, że korzystniejszy byłby wariant bardziej kosztownej, lecz mniej uciążliwej strategii inwestowania. Bilans kosztów i korzyści w tym zakresie powinien być uwzględniany na etapie analizy *ex-ante* dotyczącej wyboru wariantu inwestycji, zaś rzeczywiste oddziaływanie – powinno być monitorowane w szerszej skali (np. na poziomie programu).

Piotr Rosik – adiunkt w Instytucie Geografii i Przestrzennego Zagospodarowania Polskiej Akademii Nauk. Z wykształcenia doktor nauk ekonomicznych. Prowadził wykłady i ćwiczenia na Uniwersytecie Ekonomicznym w Poznaniu oraz Politechnice Poznańskiej. Specjalizuje się w geografii transportu, w tym przede

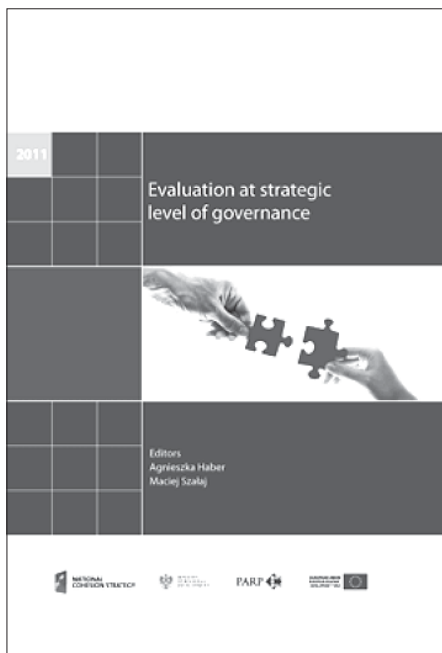
wszystkim w tematyce dostępności i mobilności. Autor i współautor kilku książek, kilkudziesięciu artykułów, a także kilkunastu raportów projektowych i ekspertyz.

Elżbieta Opalka – ukończyła studia magisterskie na Wydziale Geografii i Studiów Regionalnych na Uniwersytecie Warszawskim. Od 2001 roku pracuje w strukturach administracji państwowej, a od 2004 r. jest zatrudniona w Wydziale Ewaluacji (Krajowej Jednostce Oceny) w Departamencie Koordynacji Polityki Strukturalnej Ministerstwa Rozwoju Regionalnego, gdzie odpowiada za realizację badań ewaluacyjnych dot. efektów polityki spójności w rozwoju infrastruktury transportowej, środowiskowej i społecznej. Przewodniczy pracom grupy sterującej ewaluacją w jednym z najistotniejszych obszarów wsparcia polityki spójności – *rozwój i modernizacja infrastruktury*. Uczestnikami prac grupy są przedstawiciele ministerstw i administracji regionalnej. Jest współautorką kilku publikacji i poradników metodologicznych dotyczących ewaluacji interwencji finansowanych z funduszy europejskich oraz prowadzi wykłady na uczelniach wyższych.

Bibliografia

- Biuro Naczelnego Architekta Miasta st. Warszawy, 2005, *Warszawskie Badanie Ruchu*, Biuro Planowania Rozwoju Warszawy.
- GDDKiA, 2008, *Krajowy Model Ruchu*, opracowany przez Politechnikę Warszawską na zlecenie.
- Instytut Badawczy Dróg i Mostów, 2008, *Instrukcja oceny efektywności ekonomicznej przedsięwzięć drogowych i mostowych dla dróg wojewódzkich*, Warszawa.
- Komisja Europejska, 2010, *Europa 2020 – Strategia na rzecz inteligentnego i zrównoważonego rozwoju sprzyjającemu włączeniu społecznemu*.
- Komisja Europejska, 2011 *Zrównoważona przyszłość transportu: w kierunku zintegrowanego, zaawansowanego technologicznie i przyjaznego użytkownikowi systemu*.
- Komornicki T., Rosik P., Stępnia M., 2011(a), *Dostępność transportowa w Polsce Wschodniej*, ekspertyza na zlecenie Ministerstwa Rozwoju Regionalnego.
- Komornicki T., Rosik P., Stępnia M., Pomianowski W., 2011(b), *Symulacja Wskaźnika Międzygałęziowej Dostępności Transportowej (WMDT) na terytorium Polski do roku 2020*, ekspertyza na zlecenie Ministerstwa Rozwoju Regionalnego.
- Komornicki T., Śleszyński P. (red.), 2009, *Studia nad lokalizacją regionalnych portów lotniczych na Mazowszu*. Prace Geograficzne, 220, IGiPZ PAN, Warszawa.
- Komornicki T., Śleszyński P., Rosik P., Pomianowski W., 2010, *Dostępność przestrzenna jako przesłanka kształtowania polskiej polityki transportowej*, Biuletyn KPZK 241, Komitet Przestrzennego Zagospodarowania Kraju PAN, Warszawa.
- Komornicki T., Śleszyński P., Siłka P., Stępnia M., 2008, *Wariantowa analiza dostępności w transporcie lądowym*, [w:] *Ekspertyza do Koncepcji Przestrzennego Zagospodarowania Kraju*, t. II, na zlecenie Ministerstwa Rozwoju Regionalnego, Warszawa.
- KPRM, 2011, *Długookresowa Strategia Rozwoju Kraju 2030 (projekt)*, Warszawa.
- Ministerstwo Infrastruktury, 2011, *Strategia Rozwoju Transportu do 2020 (z perspektywą do 2030 roku)* – projekt.
- Ministerstwo Rozwoju Regionalnego, 2010(a), *Ewaluacja ex post NPR 2004-2006*, Warszawa.
- Ministerstwo Rozwoju Regionalnego, 2010(b), *Ocena wpływu inwestycji infrastruktury transportowej realizowanych w ramach polityki spójności na wzrost konkurencyjności regionów (w ramach ewaluacji ex post NPR 2004-2006)*, opracowanie wykonane na zlecenie MRR przez zespół badawczy IGiPZ PAN.
- Ministerstwo Rozwoju Regionalnego, 2011(a), *Evidence-based Cohesion Policy and its role in achieving Europe 2020 objectives. Background report*, Warszawa.
- Ministerstwo Rozwoju Regionalnego, 2011(b), *Raport Polska 2011. Gospodarka – Społeczeństwo – Regiony (projekt)*, Warszawa.
- Ministerstwo Rozwoju Regionalnego, 2011(c), *Wnioski z realizacji Polityki Spójności w świetle wyników prac analityczno-ewaluacyjnych okresu 2004-2006*, Warszawa.
- Ministerstwo Rozwoju Regionalnego, 2011(d); *Koncepcja Przestrzennego Zagospodarowania Kraju 2030. Streszczenie dokumentu rządowego*, Warszawa.
- Rosik P., Pomianowski W., Stępnia M., Komornicki T., Śleszyński P., 2011, *Narzędzie ewaluacyjno-badawcze dostępności transportowej gmin w podukładach wojewódzkich*, Raport końcowy (raport wykonany w ramach IV konkursu dotacji MRR).
- Spiekermann K., Schürmann C., 2007, *Update of selected potential accessibility indicators. Final report*, Spiekermann & Wegener, Urban and Regional Research (S&W), RRG Spatial Planning and Geoinformation.
- Wolański M., 2010, *Rozwój infrastruktury transportowej w latach 2007-2010 w kontekście dotychczasowej realizacji Strategii Rozwoju Kraju 2007-2015 oraz kluczowych strategii sektorowych*, ekspertyza wykonana na zlecenie Ministerstwa Rozwoju Regionalnego.

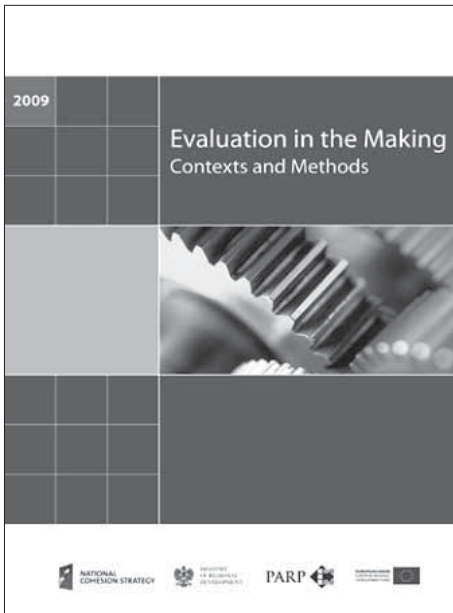
W ramach serii Ewaluacja ukazały się następujące publikacje:



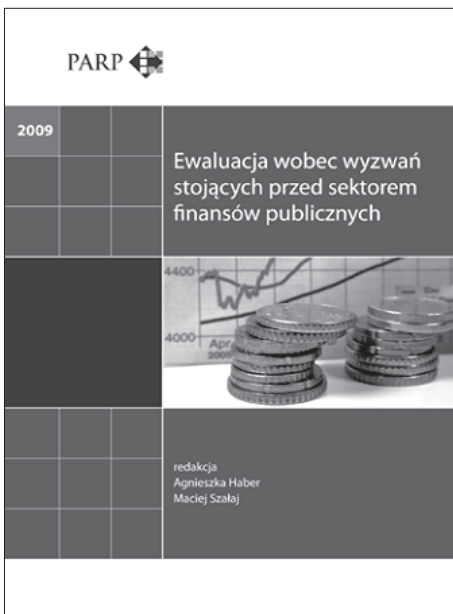
<http://www.parp.gov.pl/index/more/22858>



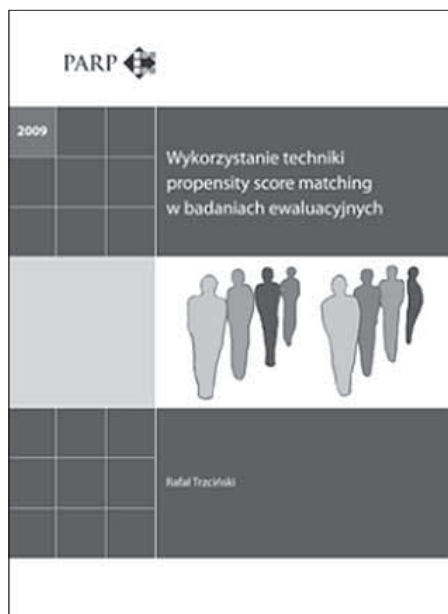
<http://www.parp.gov.pl/index/more/19489>



<http://www.parp.gov.pl/index/more/14819>



<http://www.parp.gov.pl/index/more/12416>



<http://www.parp.gov.pl/index/more/13335>



<http://www.parp.gov.pl/index/more/9658>



<http://www.parp.gov.pl/index/more/2046>



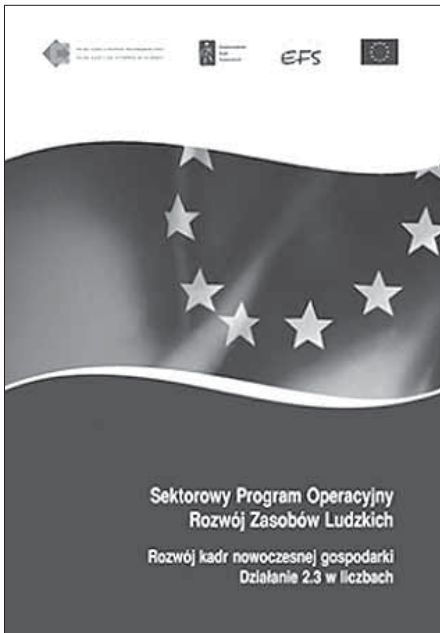
<http://www.parp.gov.pl/index/more/24238>



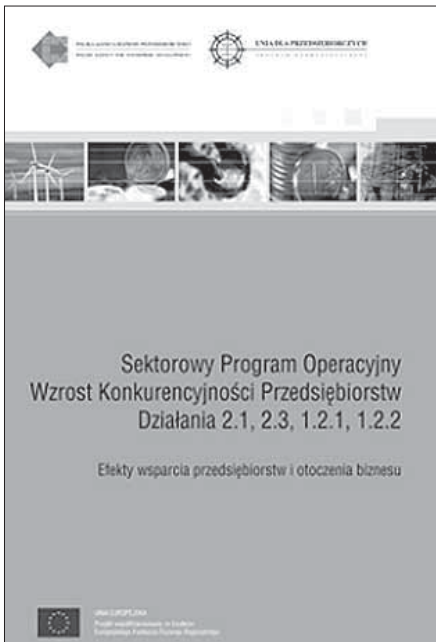
<http://www.parp.gov.pl/index/more/19735>



<http://www.parp.gov.pl/index/more/9850>



<http://www.parp.gov.pl/index/more/5474>



<http://www.parp.gov.pl/index/more/5475>

