

*Dominika Wojtowicz**, *Michał Wolański***, *Łukasz Widła-Domaradzki****

OCENA RZECZYWISTYCH EFEKTÓW PROJEKTÓW WSPÓLFINANSOWANYCH Z FUNDUSZY UNIJNYCH NA PRZYKŁADZIE PROJEKTU PRZEBUDOWY SKRZYŻOWAŃ

Niniejszy artykuł ma celu przedstawienie możliwości adaptacji założeń metodologii *impact evaluation* (ewaluacji wpływu) do oceny interwencji publicznych w infrastrukturę transportu drogowego. Artykuł składa się z dwóch części. W pierwszej przedstawione zostały główne założenia stosowania metody ewaluacji wpływu pozwalającej na oszacowanie rzeczywistych efektów – efektów netto danej interwencji – wraz ze wskazaniem przykładów badań ewaluacyjnych wykorzystujących tę metodę. Druga część artykułu zawiera prezentację wyników badań empirycznych mających na celu sprawdzenie możliwości zastosowania metodologii, na której opiera się ewaluacja wpływu, na przykładzie konkretnego projektu dotyczącego inwestycji w infrastrukturę drogową¹. Artykuł kończy krytyczna analiza metody, a więc ocena wiarygodności i użyteczności mierzenia efektów netto w odniesieniu do projektów inwestycyjnych z zakresu infrastruktury drogowej.

Doskonalenie narzędzi ewaluacyjnych stosowanych przy ocenie inwestycji w dziedzinie infrastruktury nabiera szczególnej wagi w kontekście toczącej się dyskusji na temat przyszłości polityki spójności po 2013 r. Przed rozpoczęciem projektowania założeń finansowania – wielkości budżetu oraz strategicznych kierunków wsparcia – Komisja Europejska potrzebuje rzetelnych informacji na temat efektów realizacji dotychczasowej polityki. Rozwój metodologii ewaluacji wpływu i ocena efektów netto może okazać się cenną pomocą w generowaniu argumentów do dyskusji na temat kształtu polityki spójności w zakresie infrastruktury transportu drogowego przyszłości.

Impact evaluation (ang. ewaluacja wpływu)², zakładająca szacowanie efektu netto danego projektu/programu, jest jednym z bardziej zaawansowanych, sku-

* Akademia Leona Koźmińskiego.

** Szkoła Główna Handlowa.

*** Tandem Analityczny S.C.

¹ Projekt badawczy współfinansowany ze środków UE oraz budżetu państwa w ramach konkursu Ministerstwa Rozwoju Regionalnego, realizowany przez Akademię Leona Koźmińskiego.

² *Impact evaluation* tłumaczona w literaturze przedmiotu jako ewaluacja wpływu (ocena efektów wpływu) lub ocena efektów netto programów. Wg klasyfikacji J. Lennie i n. (2006) jest to ewaluacja skoncentrowana na poszerzonej, długoterminowej ocenie wpływu lub rezultatów, zarówno zamierzonych, jak i niezamierzonych programu bądź interwencji. Zazwyczaj jest realizowana w pewnym odstępie czasu od zakończenia programu lub interwencji (za: Bruska 2009, s. 30). Autorzy proponują tłumaczenie *impact evaluation* jako ewaluacja wpływu, w odróżnieniu od ewaluacji oddziaływania, w której ocenia się szersze oddziaływanie danego projektu/programu na otoczenie.

tecznych narzędzi ewaluacji skutków inwestycji finansowanych ze środków publicznych, gdyż pomaga ocenić jej rzeczywisty efekt, oddzielając oddziaływanie czynników zewnętrznych. Dotychczas metodologia ta nie była stosowana w sektorze infrastrukturalnym ze względu na trudności związane z wyznaczeniem wiarygodnych współczynników sukcesu (np. w przypadku budowy nowych dróg), zebraniem danych (zwłaszcza historycznych) czy wreszcie wyborem tzw. sytuacji kontryfaktycznych. Przez to ostatnie pojęcie rozumie się jednostki lub populacje możliwie podobne do danej, które nie zostały poddane interwencji. Wzajemne ich porównanie pozwala określić zmiany, niebędące wynikiem ewaluowanych działań.

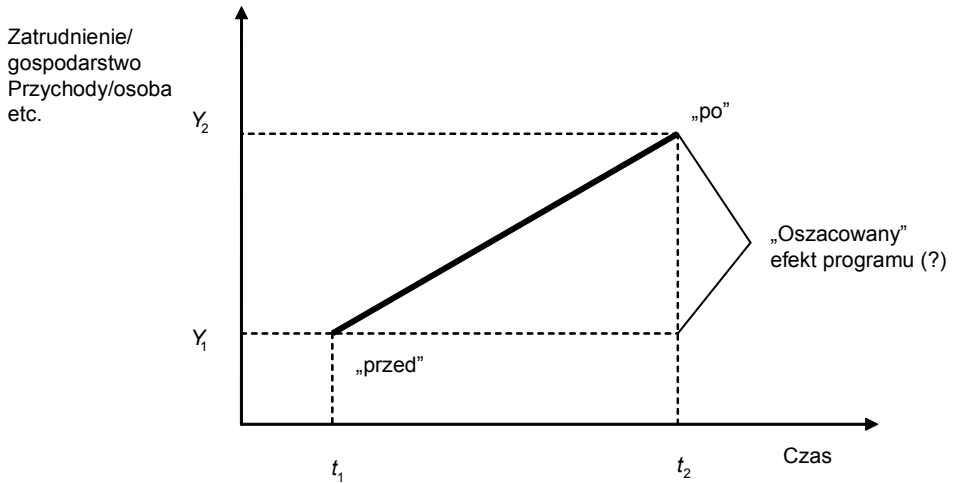
O trudnościach w praktycznym zastosowaniu tejże metody może również świadczyć fakt, że pomimo uznania zasadności jej założeń przez KE, która podkreślała, iż „gdy mówimy, że dane efekty zostały wyprodukowane lub spowodowane przez program, oznacza to, że gdyby program nie miał miejsca lub miałyby miejsce w innej formie lub innym stopniu, efekty te nie wystąpiłyby lub nie wystąpiłyby w tym samym stopniu. Oznacza to, że ważne jest, by mieć precyzyjny obraz tego, co stałoby się bez programu” (Komisja Europejska... 1997, s. 39). W dość długiej praktyce ewaluacyjnej odnaleźć można jedynie sporadyczne przypadki jej zastosowania. Należy więc zadać sobie pytanie, co sprawia, że ewaluacja wpływu nie jest powszechnie stosowana do oceny interwencji publicznych.

Dlaczego warto stosować ewaluację wpływu?

Jednym z kluczowych zadań ewaluacji jest ustalenie związku przyczynowego między podjętą interwencją a obserwowanymi wynikami. Ustalenie przyczynowego wpływu interwencji wymaga:

- zmierzenia efektu całkowitego interwencji, jako ogólnej zmiany w zakresie zdefiniowanym przez cel (zmiany na poziomie przyjętych wskaźników),
- oddzielenia zmiany niezależnej od podjętej interwencji od zmiany, którą można przypisać wpływowi interwencji (efekt netto) (Górniak 2007).

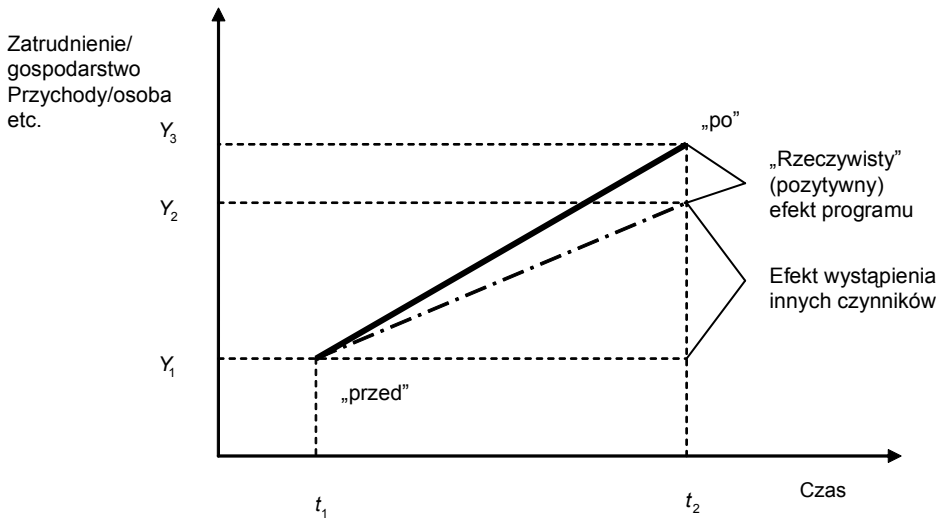
Istotność zastosowania metody porównania stanów kontryfaktycznych jako jedynego narzędzia gwarantującego rzetelne oszacowanie rzeczywistego efektu netto danego programu/projektu tłumaczy w swoich raportach z badań dotyczących skuteczności interwencji podejmowanych w ramach programu SAPARD J. Michałek (Michałek i in. 2008). Najczęściej wykorzystywaną w badaniach ewaluacyjnych metodę badania efektów, polegającą na zmierzeniu wybranych wskaźników przed realizacją danej interwencji po jej wykonaniu, nazywa on „oceną naiwną” (*naive evaluation*), gdyż nie jest ona w stanie uwzględnić innych czynników (np. koniunktury, zmian w przepisach prawnych) oddziałujących na zjawisko/beneficjentów, których dotyczy projekt/program. I tak na przykład zbadanie wzrostu zatrudnienia czy dochodów *per capita* beneficjentów programu zakładającego wsparcie producentów sektora rolnego nie tłumaczy, w jakim stopniu zmiany te wynikały właśnie z faktu otrzymania dotacji, a nie były wynikiem innych, niezależnych od programu czynników.



Ryc. 1. Przykład „naiwnej” ewaluacji efektów programu

Źródło: Michałek 2008.

Zastosowanie metody analizy sytuacji kontrfaktycznych pozwala na udzielenie odpowiedzi na najważniejsze z punktu widzenia oceny rzeczywistych efektów projektu/programu pytanie: „Jaka byłaby sytuacja beneficjentów, gdyby nie mieli oni możliwości skorzystania ze wsparcia w ramach programu?” (Michałek i in. 2008). Dzięki niej możemy zauważyć wystąpienie pozytywnego lub negatywnego efektu netto danej interwencji, co pokazuje poniższy wykres.



Ryc. 2. Istotność uwzględnienia w badaniach jednostek niebiorących udziału w ewaluowanym programie/projekcie (porównania sytuacji kontrfaktycznej do sytuacji badanej)

Źródło: Michałek 2008.

Dotychczasowe doświadczenie ewaluacji inwestycji w infrastrukturę transportową

Dotychczasowe badania wpływu inwestycji transportowych przeprowadzane na zlecenie Komisji Europejskiej koncentrowały się na modelowaniu ekonomicznym. Wart przytoczenia jest wciąż rozwijany model SASI (*Socio-Economic and Spatial Impacts of Trans-European Transport Networks*). W modelu na strukturę sieci trzech gałęzi transportu, tj. drogowego, kolejowego i lotniczego, nałożono przestrzenne relacje transportowe (podróże biznesowe oraz prywatne). Model, składający się z kilku submodeli, w jednym z nich, tj. submodelu „Regionalny PKB”, prognozuje wielkość PKB generowaną w każdym regionie w trzech sektorach – rolnictwie, przemyśle i usługach. Prognozy oparte są na funkcji produkcji, w której dostępność regionalna (szacowana w innym submodelu) jest jedną ze zmiennych objaśniających. Model zakłada cztery scenariusze: referencyjny (bez inwestycji, inwestycje zakończone do 2007 r.), maksymalny drogowy (realizowane będą wszystkie postulowane inwestycje drogowe), maksymalny kolejowy oraz zrównoważony. Wyniki dla Polski przewidują stopę wzrostu PKB wyższą w przypadku scenariusza zrównoważonego o ok. 0,8 punktu procentowego w 2031 r. niż w przypadku scenariusza referencyjnego w tym samym roku. Skala oddziaływań na stopę wzrostu PKB na poziomie kraju pokazuje, jak trudny do uchwycenia będzie wpływ netto inwestycji³.

W sektorze transportowym ocena efektu netto jest bardzo pożądana ze względu na jego dużą zależność od innych działów gospodarki, co wydatnie zmniejsza efekt brutto w przypadku wielu współczynników. Przykładem może być tutaj chociażby łączna liczba wypadków drogowych i ich ofiar, które w ostatnim roku przed przeprowadzeniem opisywanego badania – mimo wielu działań prewencyjnych, w tym wprowadzenia całorocznego obowiązku jazdy na światłach – wzrosły. Jest jednak prawdopodobne, że wzrost ten nastąpił wskutek zwiększenia liczby samochodów, podjęte działania prewencyjne zaś charakteryzują się pozytywnym efektem netto (tj. bez nich obserwowany wzrost wypadkowości byłby jeszcze większy). Niemniej przeprowadzona kwerenda literatury oraz raportów z europejskich i amerykańskich badań ewaluacyjnych wskazuje, że w **przypadku przedsięwzięć infrastrukturalnych pomiar efektu netto nie był dotychczas stosowany**.

Pewną analogią do efektu netto może być monitorowanie wybranych współczynników nie tylko w wartościach bezwzględnych, lecz także względem prognoz referencyjnych. Tego typu podejście stosowane jest w pewnym zakresie również w Polsce, czego przykładem mogą być chociażby projekty zakupu i modernizacji taboru realizowane przez PKP Przewozy Regionalne (Wolański i in. 2008). Brak jest przy tym jednak prób weryfikacji *ex post* prognoz referencyj-

³ Model SASI jest wciąż rozwijany, m.in. w projekcie IASON (*Integrated Appraisal of Spatial Economic and Network effects of transport investments and policies*) oraz w pracach Europejskiej Sieci Obserwacyjnej Rozwoju Terytorialnego i Spójności Terytorialnej (ESPON).

nych, bazujących przecież na wielu innych założeniach makroekonomicznych, dokonanych *ex ante*, a zatem na podstawie dużo mniejszej wiedzy.

Również w przypadku analizowanych badań amerykańskich (*Detailed Information...* 2008) operuje się jedynie zakładanymi i faktycznie osiągniętymi wielkościami wybranych współczynników, nie wyodrębniając przy tym efektów badanej interwencji oraz oddziaływania innych czynników – a zatem zmian, do których doszłoby także bez interwencji.

Podobne systemy wskaźników stosuje się również w Niemczech, chociaż to autorzy analizowanej ewaluacji niemieckiej zauważają, że: „w zasadzie skutki działań polityki transportowej powinny być oceniane poprzez porównanie sytuacji, w której dane działanie wprowadzamy (przypadek «with»), i w której go nie wprowadzamy (przypadek «without»). W przypadku stosowanych tabel wskaźników opisywany jest tymczasem jedynie przypadek «with»” (Ridder i in. 2003, s. 36).

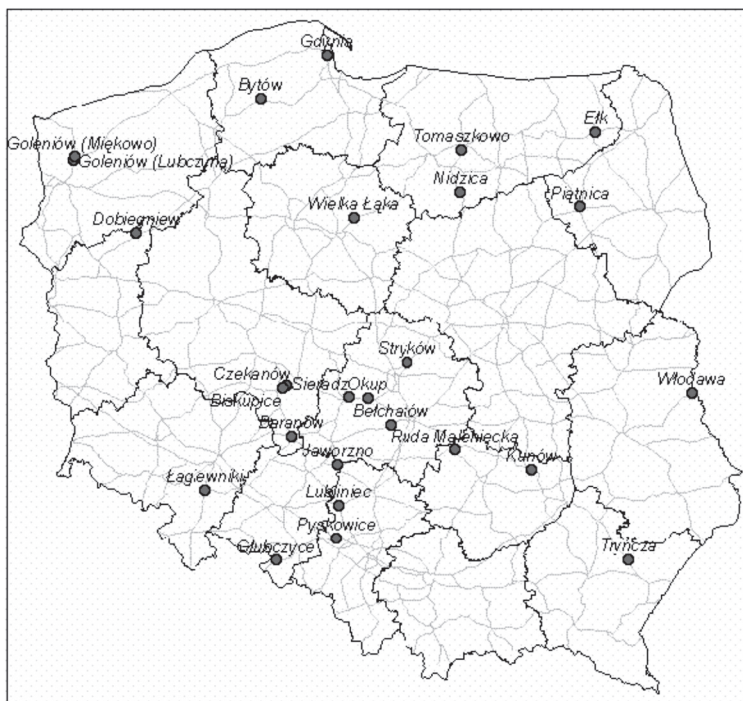
Przyczyn braku porównania *ex post* przypadku „with” i przypadku „without” można dopatrywać się m.in. w dużej skali podejmowanych działań, co w niektórych sytuacjach uniemożliwia określenie stanu kontrfaktycznego (np. jeśli w danym okresie modernizujemy wszystkie systemy komunikacji miejskiej w kraju). Problemem jest również wyznaczenie jednostek obrazujących stan kontrfaktyczny – użytkowników bądź elementów infrastruktury, a wreszcie – stworzenie odpowiednio trafnych wskaźników oddziaływania i rezultatu oraz prawidłowe określenie relacji przyczynowo-skutkowych pomiędzy nimi.

Nie ulega jednak wątpliwości, że kwestia oddzielenia wpływu interwencji od innych czynników w przypadku transportu jest zasadna, o czym świadczy fakt, iż większość wskaźników oddziaływania oraz rezultatu, monitorowanych w ramach programów i projektów infrastrukturalnych, jest podatna na wiele czynników w skali makro. Prostym, lecz dobitnym przykładem mogą być tutaj zmiany PKB. Wpływają one istotnie na mobilność społeczeństwa, przez co z jednej strony mają wpływ na wielkość przewozów, a z drugiej – również ryzyko wypadku, rosnące przy zatłoczeniu dróg, ale malejące dzięki zakupom bezpieczniejszych samochodów. Wszystko to oddziałuje na typowe wskaźniki projektów i programów infrastrukturalnych, takie jak wielkość przewozów czy liczba wypadków.

Wdrożenie metodologii ewaluacji wpływu na przykładzie projektu *Przebudowa szczególnie niebezpiecznych skrzyżowań na sieci dróg krajowych*

Charakterystyka projektu

Podstawowym celem realizowanego przez GDDKiA projektu *Przebudowa szczególnie niebezpiecznych skrzyżowań na sieci dróg krajowych* było podniesienie poziomu bezpieczeństwa ruchu drogowego, a w szczególności ograniczenie liczby wypadków na każdym z zaznaczonych na poniższej mapie skrzyżowań.



Ryc. 3. Lokalizacja skrzyżowań uwzględnionych w projekcie *Przebudowa szczególnie niebezpiecznych skrzyżowań na sieci dróg krajowych*

Źródło: materiały zamieszczone na stronie <http://www.gddkia.gov.pl>.

W wyniku realizacji projektu oczekiwano (w zależności od typu skrzyżowania) wystąpienia pozytywnych efektów dotyczących:

- zapewnienia bezpiecznego przejazdu użytkownikom krzyżujących się dróg,
- ograniczenia prędkości i uspokojenia ruchu w obszarze skrzyżowania,
- wzrostu czytelności układu skrzyżowania,
- maksymalnego rozdzielenia i ograniczenia liczby punktów kolizji na skrzyżowaniu,
- poprawy przepustowości i płynności ruchu na skrzyżowaniu,
- zapewnienia odpowiedniej dostrzegalności i percepcji układu drogowego,
- zapewnienia właściwych (bezpiecznych) warunków ruchu pieszych w obrębie skrzyżowania,
- dostosowania rozwiązań geometrycznych do obowiązujących wymagań technicznych,
- przywrócenia właściwego stanu technicznego infrastruktury drogowej⁴.

⁴ Opracowane na podstawie materiałów zamieszczonych na stronie <http://www.gddkia.gov.pl>.

Wybór metody analizy danych – dobór jednostek referencyjnych

Zazwyczaj do badania efektu netto w ewaluacjach korzysta się z metody *Propensity Score Matching* (PSM). W przypadku niniejszego badania zachodziła uzasadniona obawa, że PSM nie da rzetelnych wyników bądź w ogóle nie da się jej zastosować ze względu na słabą jakość dopasowania. Obawa ta wynikała z faktu, że baza danych jednostek niepoddanych interwencji powinna być na tyle obszerna, by można było dobrać do jednostek ewaluowanych odpowiedniej jakości jednostki bliźniacze.

W toku konceptualizacji metody do badania zakładano, że w razie zaistnienia sytuacji uniemożliwiającej wykorzystanie w zadowalający sposób PSM będą poszukiwane inne metody, które pozwolą na dobranie jednostek referencyjnych do jednostek ocenianych. Na tym etapie przygotowań zakładano nie tylko możliwość wykorzystania metody szeregów czasowych (TS), lecz także, zważywszy na niewielką liczbę ewaluowanych obiektów, ręczny dobór skrzyżowań bliźniaczych. Oczywiście ostatnią z wymienionych metod posłużono by się w ostateczności, gdyby metody PSM oraz TS nie dały zadowalających rezultatów.

Obawy co do rzetelności zastosowanych metod budziła także jakość bazy danych, jak również konieczność nielosowego doboru skrzyżowań referencyjnych. Dobór nielosowy był konieczny, gdyż skrzyżowania wybrane do przebudowy zostały określone jako „szczególnie niebezpieczne”. W związku z tym nie można było wziąć pod uwagę całej populacji skrzyżowań, ale jedynie wybrać spośród nich te, które – zdaniem badaczy – są szczególnie niebezpieczne.

Dla każdego rekordu – a zatem skrzyżowania – określono:

- jego dokładną lokalizację wraz z pikietażem oraz informacją o położeniu w terenie zabudowanym lub niezabudowanym;
- liczbę wlotów;
- kategorię dróg na poszczególnych wlotach;
- kształt skrzyżowania i jego ewentualne wyposażenie w sygnalizację świetlną;
- dane o łącznej wypadkowości w latach 2001–2005 (a zatem przed rozpoczęcia projektu) oraz w roku 2007 (po jego realizacji).

Należy przy tym zwrócić uwagę, że:

- otrzymana baza danych zawierała tylko wypadki – czyli te zdarzenia, w których co najmniej jedna osoba została ranna i w przypadku których obowiązkiem jest interwencja policji; brak było jakichkolwiek danych o kolizjach;
- jakość danych dotyczących wypadkowości jest średnia ze względu na błędny opis części zdarzeń (w szczególności w bazie wykryto sprzeczności, takie jak np. sformułowanie, iż do niektórych wypadków doszło na „odcinku między skrzyżowaniami”, a jednocześnie zaznaczono je na pikietażu wskazującym skrzyżowanie);
- rekord w bazie stanowił jeden wypadek, a zatem przypisywanie wypadków do skrzyżowania następowało ręcznie (nie należy tego mylić z ręcznym doбором skrzyżowań bliźniaczych, rozważanym jako alternatywna wobec PSM meto-

da) i niemożliwe było określenie wszystkich skrzyżowań szczególnie niebezpiecznych ani ich w pełni usystematyzowana analiza;

- w studium wykonalności projektu zakładano przebudowę 26 z góry wybranych „czarnych punktów” – nie analizując jako wariantu projektu innych skrzyżowań; w ten sposób beneficjent nie tylko nie udowodnił, dlaczego taki zestaw skrzyżowań jest najtrafniejszy, lecz także utrudnił określenie stanu kontrfaktycznego (co jest możliwe np. w projektach EFS dzięki analizie jednostek odrzuconych na etapie aplikacji);
- nie dysponowano danymi dotyczącymi ruchu na tych wlotach skrzyżowań, które stanowiły drogi powiatowe lub drogi kategorii niższych⁵.

W początkowej fazie prowadzonych badań zdecydowano się na klasyczny sposób doboru skrzyżowań referencyjnych, czyli na model regresji logistycznej. W modelu uwzględnione zostały następujące zmienne: liczba wlotów skrzyżowania, ruch na wszystkich wlotach skrzyżowania, rodzaj terenu (zabudowany, niezabudowany), rodzaj skrzyżowania (jednopoziomowe, dwupoziomowe, rondo) oraz liczba uszkodzonych w latach 2001–2005 (wypadki ogółem, zabici, ranni).

Okazało się, że taki sposób doboru daje bardzo dobre rezultaty. Ostatecznie otrzymano dopasowanie skrzyżowań referencyjnych do wszystkich skrzyżowań ewaluowanych na poziomie 98,1%. **W związku z tak wysokim poziomem dopasowania zarzucono inne pomysły na dobór skrzyżowań referencyjnych, pozostawszy przy metodzie PSM.**

Metodę tę wykorzystano zarówno w przypadku oceny całego projektu, jak i zaimplementowano do ewaluacji poszczególnych skrzyżowań wchodzących w skład projektu.

Jak zaznaczono powyżej, przy szacowaniu skłonności do partycypacji wykorzystano model regresji logistycznej. Na podstawie rozkładu prawdopodobieństw sukcesu zmiennej zależnej oszacowano bliskości poszczególnych skrzyżowań i ostateczny dobór skrzyżowań referencyjnych.

Wyniki analizy efektu netto interwencji publicznej w zakresie bezpieczeństwa ruchu drogowego

Przebieg zastosowania metody można podzielić na trzy etapy. Pierwszy polegał na uzupełnieniu braków danych dotyczących ruchu na wlotach skrzyżowania, gdy wlot ów stanowiła droga o kategorii niższej niż wojewódzka. Jeśli chodzi o średni ruch na wlotach skrzyżowania, to ostateczna baza w wielu miejscach była niekompletna. Podzielono więc wloty skrzyżowania na kategorie (droga wojewódzka, powiatowa, gminna, lokalna, wewnętrzna) i oszacowano te wartości, których brakowało. Okazało się, że w żadnej ze składowych baz danych nie zostały zawarte dokładniejsze informacje inne niż wypadkowość. Innymi słowy,

⁵ Dodatkowo w projekcie uwzględniono tylko 24 z 26 przebudowanych skrzyżowań, ze względu na brak dokładnych danych o pikietażu jednego z nich oraz specyficzny charakter skrzyżowania ul. Morskiej z Obwodnicą Trójmiasta w Gdyni, która mogłaby negatywnie wpłynąć na rzetelność otrzymanych wyników.

członkowie zespołu badawczego skoncentrowali się na wypadkowości nie tyle ze względów merytorycznych, ile raczej z powodu niedostępności innych danych.

W drugim etapie nastąpił dobór skrzyżowań referencyjnych, opisany w poprzedniej części artykułu, w trzecim zaś – przeprowadzenie właściwej analizy tak przygotowanych danych.

Analizę danych przeprowadzano w sposób stopniowy. Najpierw wyliczono średnie roczne wartości wypadkowości w latach 2001–2005 dla skrzyżowań referencyjnych i ewaluowanych. W przypadku skrzyżowań, które uwzględniono w projekcie, był to okres przed przebudową. Następnie wyliczono wypadkowość w roku 2007 (czyli już po przebudowie 24 skrzyżowań, objętych projektem). W dalszej kolejności oszacowano proporcje odsetka wypadkowości w latach 2001–2005 w stosunku do wypadkowości z roku 2007. Otrzymane proporcje od siebie odjęto, w wyniku czego otrzymano cztery składowe całościowego efektu netto. Były to: efekt netto dotyczący wypadkowości, liczby rannych oraz liczby zabitych.

Na tym etapie zrezygnowano z prostszej metody polegającej na odjęciu od siebie różnic wynikających ze wzrostu/spadku wypadkowości w latach 2001–2005 i roku 2007. Ponieważ dobór skrzyżowań referencyjnych na wstępnym etapie został dokonany w sposób celowy, w związku z tym nie było możliwości sprawdzenia, jakie są średnie wartości błędów oszacowania wypadkowości i liczby osób poszkodowanych. Oznaczało to *de facto*, że możliwy obserwowalny spadek z trzech do dwóch wypadków między średnią z lat 2001–2005 a rokiem 2007 nie implikuje koniecznie spadku o jedność, lecz spadek o 66,7%. Dopiero porównując obie frakcje (dla skrzyżowań referencyjnych i skrzyżowań, które zostały poddane interwencji), możemy określić ostateczny efekt netto. Powyższa metoda opiera się na założeniu, że o ile konkretnych wartości wypadkowości nie da się oszacować (ze względu na dobór próby), o tyle proporcje wypadkowości powinny dobrze oddawać stan faktyczny – zarówno w przypadku skrzyżowań referencyjnych, jak i przebudowanych.

Na ostatnim etapie podjęto próbę zredukowania efektu netto do dwóch tylko wskaźników: efektu netto liczby wypadków i efektu netto ich najważniejszych skutków, a zatem liczby poszkodowanych w wypadkach. Co prawda istniała teoretyczna możliwość zintegrowania tych dwóch wskaźników w jeden, lecz powstały w ten sposób wskaźnik byłby trudny do zinterpretowania, dlatego też zespół badaczy zaproponował pozostanie przy dwóch miarach efektu netto.

W trakcie prac pojawił się dylemat wyboru znaku dla ostatecznej prezentacji efektu netto. W literaturze przyjęło się, że ujemny efekt netto oznacza pogorszenie wyniku, dodatni zaś – że osiągnięto sukces. Problem polega na tym, że efekty netto w niniejszej pracy są same z siebie zdefiniowane negatywnie: zmniejszenie wypadkowości i zmniejszenie liczby ofiar to dodatnie efekty netto, natomiast zwiększenie wypadkowości oznaczałoby efekt netto ujemny. Rodzi się w związku z tym pokusa, by ujemnym efektem netto oznaczać zmniejszenie wypadkowości, dodatnim zaś – zwiększenie. Choć jest to rozwiązanie ze wszech miar intuicyjne, kłóci się jednak z ogólnie przyjętymi definicjami, w których ujemny efekt netto oznacza porażkę. W poniższych wyliczeniach zastosowano taką właśnie no-

tację: za dodatni efekt netto uznano sukces, czyli zmniejszenie wypadkowości/liczby poszkodowanych, za ujemny zaś – porażkę, czyli zwiększenie wypadkowości/liczby ofiar lub ich zmniejszenie, ale słabsze na skrzyżowaniach ewaluowanych w porównaniu z referencyjnymi.

Zgodnie z zaproponowaną metodologią dla 24 ewaluowanych skrzyżowań oraz dla skrzyżowań referencyjnych otrzymano wyniki zamieszczone w poniższej tabeli.

Tab. 1. Wyniki analiz dla 24 ewaluowanych skrzyżowań oraz dla skrzyżowań referencyjnych

	Wypadki		
	średnia wypadków rocznie	średnia liczba zabitych rocznie	średnia liczba rannych rocznie
	2001–2005		
Ewaluowane	1,03	0,22	1,93
Referencyjne	0,81	0,33	1,29
	2007		
Ewaluowane	0,83	0,00	1,33
Referencyjne	0,46	0,08	1,00

Na podstawie przedstawionych wartości wyliczono względny wzrost/spadek wypadkowości, zakładając, że podda się on interpretacji empirycznej. Posługując się wzorem na przyrost względny ($\frac{W_{2007} - W_{2001-2005}}{W_{2001-2005}} \cdot 100$), wyliczono dane do kolejnej tabeli 2:

Tab. 2. Względny przyrost/spadek wypadkowości na badanych skrzyżowaniach

	Średnia wypadków rocznie	Średnia liczba zabitych rocznie	Średnia liczba rannych rocznie
	względny przyrost poziomu wypadkowości		
Ewaluowane	-19,4%	-100%	-31,1%
Referencyjne	-43,2%	-75,8%	-22,5%

Po uwzględnieniu zaokrągleń otrzymujemy ostatecznie składowe efekty netto, wyliczone ze wzoru $Wp_r - Wp_e$, gdzie Wp_r oznacza względny przyrost poziomu wypadkowości dla skrzyżowań referencyjnych, Wp_e zaś – dla skrzyżowań ewaluowanych. Wyniki przekształcenia zawarto w tabeli 3.

Tab. 3. Wartości składowych efektu netto

	Składowe efekty netto (1)		
	średnia wypadków rocznie	średnia liczba zabitych rocznie	średnia liczba rannych rocznie
Efekt netto	-23,8%	24,2%	8,6%

Empiryczna interpretacja wyników wydaje się prosta. Wartość $-23,8\%$ oznacza, że w porównaniu ze skrzyżowaniami z grupy kontrolnej poprawa sytuacji (zmniejszenie wypadków) na skrzyżowaniach ewaluowanych była słabsza – nastąpiło zmniejszenie liczby wypadków w przeliczeniu na jedno skrzyżowanie: w grupie ewaluowanej o $19,4\%$, a w grupie kontrolnej o $43,2\%$. Pozytywny kierunek zmian jest zauważalny w obu przypadkach, jednak przy skrzyżowaniach ewaluowanych jest on słabszy.

Jednocześnie warto zwrócić uwagę na dodatnie wartości składowych efektów netto w przypadku szacowania osób poszkodowanych, zwłaszcza jeśli chodzi o szacowaną liczbę osób zabitych. Można więc stwierdzić, że choć wypadkowość na ewaluowanych skrzyżowaniach spadła w mniejszym stopniu niż na referencyjnych, liczba osób poszkodowanych w wypadkach spadła wyraźniej na skrzyżowaniach ewaluowanych niż referencyjnych. Co więcej, wynik ten nie jest spowodowany ogólnym wzrostem bezpieczeństwa samochodów czy koniecznością zapinania pasów, gdyż wówczas, jak można przypuszczać, podobne zjawisko obserwowalibyśmy w przypadku skrzyżowań referencyjnych i ostatecznie efekt netto byłby niższy (lub nie wystąpiłby w ogóle).

Jeżeli weźmiemy pod uwagę, że przedstawione liczby odzwierciedlają względny procent wzrostu/spadku danego zjawiska, który jest wynikiem interwencji, możemy procentowe wyniki efektu netto odnieść do statystyk zawartych w tabeli 1. Otrzymujemy wówczas efekt netto wyrażony w jednostkach (liczbie wypadków oraz liczbie osób rannych i zabitych), które zostały „zaoszczędzone” w wyniku interwencji. Składowy efekt netto mnożymy przez wartość wskaźnika uzyskaną dla ewaluowanego skrzyżowania w latach 2001–2005 ($EN_a \cdot \varphi(2001;2005)_a$), gdzie EN_a oznacza składowy efekt netto dla wskaźnika α (np. wypadkowości), $\varphi(2001;2005)_a$ zaś – wartość danego wskaźnika (np. wypadkowości) w latach 2001–2005 dla ewaluowanych skrzyżowań. Wyniki przedstawiono w tabeli 4.

Tab. 4. Wartości składowych efektu netto

	Składowe efekty netto (2)		
	średnia wypadków rocznie	średnia liczba zabitych rocznie	średnia liczba rannych rocznie
Efekt netto	$-24\% \cdot 1,03 = \mathbf{-0,245}$ wypadku rocznie	$24\% \cdot 0,22 = \mathbf{0,053}$ osoby zabitej rocznie	$9\% \cdot 1,93 = \mathbf{0,166}$ osoby rannej rocznie

Tutaj też, jak w przypadku tabeli 3, ujemny efekt netto oznacza słabszą poprawę wyniku, dodatni zaś – wyraźne polepszenie. Przekształcenie to ma tę zaletę, że pozwala interpretować nie względne wzrosty/spadki wypadkowości i odsetki osób poszkodowanych, lecz analizować je w jednostkach mierzalnych: pojedynczych wypadków i osób poszkodowanych. Z powyższego wynika, że w wyniku przebudowy zmniejszono wypadkowość jedynie mniej więcej o jeden wypadek na cztery lata, podczas gdy liczbę ofiar śmiertelnych zmniejszono do poziomu jednej w ciągu 20 lat ($0,053$ ofiary rocznie $\cdot 20$ lat ≈ 1 ofiara śmiertelna).

Przekształcenie to pozwala również na zredukowanie liczby składowych efektów netto do dwóch tylko wskaźników: wzrostu/spadku wypadkowości *per se* i wzrostu/spadku liczby osób poszkodowanych. Wyniki zamieszczono w tabeli 5:

Tab. 5. Wartości składowych efektu netto (w jednostkach)

Składowe efekty netto (3a)	
	średnia wypadków rocznie
Efekt netto (w jednostkach)	-0,245 wypadku rocznie
	wskaźnik poszkodowanych w wypadkach (w przeliczeniu na osoby ranne)
	0,539 osoby poszkodowanej rocznie

Tak jak wspomniano powyżej, zdecydowano się nie łączyć efektu netto związanego z wypadkowością i efektu netto związanego z liczbą osób poszkodowanych. Wziąwszy jednak pod uwagę dane zamieszczone w *Niebieskiej Księdze* (2008, s. 65), można założyć, że koszt zewnętrzny wypadku przypadający na jedną osobę zabita jest w przybliżeniu równy kosztowi zewnętrznemu siedmiu osób rannych, i na tej podstawie wyliczyć syntetyczny wskaźnik kosztów zewnętrznych szkód na osobach. Zgodnie z tym przeliczeniem otrzymujemy dla średniej wypadkowości wynik tożsamy z poprzednio prezentowanym, pozostałe dwa składowe efekty netto zaś zostały przedstawione w sposób zagregowany. Wynik 0,539 oznacza, że dzięki przebudowie udało się uniknąć ponad dwóch osób rannych w wypadkach w skali dwóch lat.

Oczywiście, te same dane można przedstawić także w przeliczeniu na osoby zabite, co pokazuje tabela 6.

Tab. 6. Wartości składowych efektu netto (w przeliczeniu na osoby zabite)

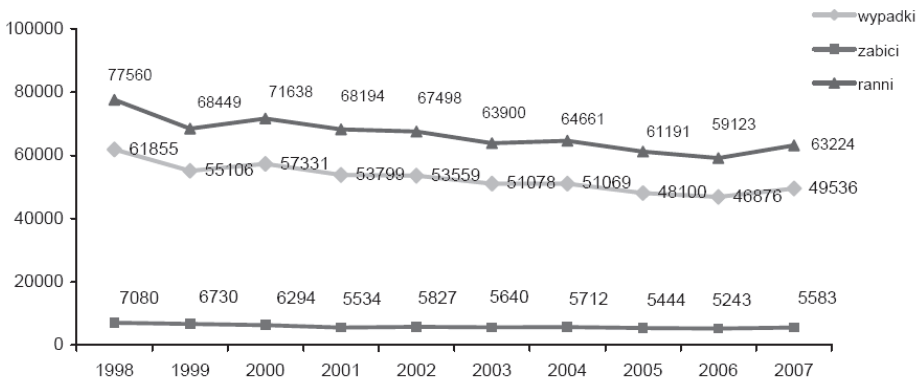
Składowe efekty netto (3b)	
	średnia wypadków rocznie
Efekt netto (w jednostkach)	-0,245 wypadku rocznie
	wskaźnik poszkodowanych w wypadkach (w przeliczeniu na osoby zabite)
	0,077 osoby poszkodowanej rocznie

Wynik ten również otrzymano na podstawie opisanego powyżej przelicznika 1:7.

Interpretacja otrzymanych wyników nie nastęrcza większych trudności. Najistotniejszy jest fakt, że zastosowana metodologia pozwoliła na dokładny i wiarygodny dobór skrzyżowań podobnych, wielkość efektu netto zaś istotnie różniła się od efektu brutto, co uzasadnia konieczność szacowania tego pierwszego.

Natomiast z punktu widzenia ewaluacyjnego – oceny efektywności i skuteczności analizowanego projektu SPOT – bardzo ciekawy jest fakt, że zaobserwowa-

ny efekt brutto również miał wartość ujemną, czyli że także na skrzyżowaniach referencyjnych zaobserwowano istotny spadek wypadkowości, w szczególności zaś spadek liczby zabitych. Wartości te są sprzeczne z ogólnymi trendami panującymi w naszym kraju, w którym liczba ofiar wypadków drogowych od 2000 roku pozostaje praktycznie na stałym poziomie, a liczba rannych jedynie nieznacznie spada (nie więcej niż o 10%) – zob. wykres poniżej.



Ryc. 4. Wypadkowość w Polsce w latach 1998–2007

Źródło: Symon 2008, s. 4.

Przyczyn takiego stanu rzeczy można upatrywać w kilku czynnikach.

- Po pierwsze – część dobranych w niniejszym projekcie alternatywnych „czarnych punktów” mogła mieć charakter efemeryczny – były to miejsca, gdzie kilka wypadków zdarzyło się bez istotnej przyczyny (co upodobniło je do „czarnych punktów”) i podobnie bez większej interwencji – przestały się zdarzać. Tutaj jeszcze raz należy podkreślić, że problemem był fakt, iż beneficjent w dokumentacji projektowej nie uzasadnił wyboru przebudowywanych skrzyżowań spośród większej liczby rzeczywistych „czarnych punktów”, mogących łatwo stanowić skrzyżowania referencyjne. Problem ten można jednak również rozwiązać przez bardziej szczegółowe analizy na etapie późniejszych badań ewaluacyjnych (w skrajnym przypadku – przez wizje lokalne wytypowanych wstępnie skrzyżowań referencyjnych).
- Po drugie – skrzyżowania referencyjne również mogły zostać poddane innej interwencji, np. budowie obwodnicy danej miejscowości, budowie sygnalizacji świetlnej, budowie chodnika, oświetlenia czy postawieniu fotoradaru. Do tej uwagi odnosi się też postulat analiz eksperckich umieszczony w poprzednim punkcie, aczkolwiek należy się wówczas zastanawiać również nad rzeczywistą zasadnością interwencji współfinansowanej z środków unijnych – być może istniejące problemy udałoby się rozwiązać na szczeblu lokalnym czy krajowym (choć z drugiej strony unijna polityka transportowa zwraca uwagę na istotny problem wąskich gardeł).

- Po trzecie – mogło się okazać, że kierowcy w ogóle zaczęli uważać w punktach, w których niegdyś zdarzało się dużo wypadków, i „czarne punkty” podlegają również pewnym naturalnym przesunięciom.
- Po czwarte – może się również okazać, że część istniejących „czarnych punktów” została wybrana nie tylko w sposób niesformalizowany, lecz także błędnie, gdyż wśród przebudowywanych skrzyżowań znajdowały się takie, na których wypadkowość nie była duża (być może miały one ograniczoną przepustowość, jednak zwiększanie płynności ruchu nie było celem tego projektu). Może o tym świadczyć fakt, że według bazy KGP w latach 2001–2005 na prawie połowie (11) z przebudowanych skrzyżowań zdarzyły się co najwyżej dwa wypadki⁶.
- Po piąte – analizowany okres po przebudowie był krótki i chociaż dotyczył łącznie 24 lub 23 obiektów (a zatem odpowiadał prawdopodobieństwu wystąpienia wypadku na jednym skrzyżowaniu w ciągu ćwierci wieku), to jednak wciąż możliwe jest zniekształcenie wyników ze względu na rzadkość wybranych zdarzeń.

Należy zauważyć, że zaproponowany model szacuje prawdopodobieństwo wybrania skrzyżowania do projektu tylko na podstawie dość ograniczonego zestawu zmiennych niezależnych, opisanych we wcześniejszej części tekstu. Nie bierze jednak pod uwagę zmiennych, które mogły istotnie wpłynąć na ostateczną wypadkowość na ewaluowanych skrzyżowaniach, a których wpływ jest często niemierzalny bądź w ogóle ukryty. Mamy tu na myśli przede wszystkim kwestię przyzwyczajania kierowców oraz rozmaite zdarzenia o charakterze intencjonalnym, które nie figurują w żadnej bazie danych.

Jeśli chodzi o kwestię przyzwyczajania, to nie można wykluczyć, że ponieważ do projektu przebudowy wybrane zostały skrzyżowania określone jako „najbardziej niebezpieczne”, to kierowcy, którzy korzystali z tych skrzyżowań, przestali je za takowe uważać po przebudowie. Skrzyżowanie, które zostaje przebudowane, może w świadomości użytkowników dróg zmienić swój status ze „szczególnie niebezpiecznego” na „bezpieczne” – inaczej po co w ogóle je przebudowywać? Zbadanie zmian świadomościowych prowadzących być może do brawury bądź przynajmniej do wygaszenia szczególnej ostrożności jest zadaniem niezwykle trudnym, jeśli nie w ogóle niemożliwym. Na poziomie konstruowania modelu efektu netto dla projektów transportowych nie jesteśmy w stanie oszacować wpływu tego rodzaju zmiennych ukrytych.

Druga kategoria zmiennych, które nie poddają się modelowaniu, dotyczy uwzględnienia tych wszystkich zdarzeń, które są nierejestrowane w bazach danych, a które mogą mieć bezpośrednie przełożenie na wypadkowość. Mowa tu na przykład o znakach dotyczących ograniczenia prędkości oraz znakach „czarny punkt”. Niestety, nie wiemy, czy w przypadku przebudowanych skrzyżowań

⁶ Należy tutaj podkreślić niedoskonałość istniejącej bazy – na potrzeby studiów wykonalności opracowano bardziej szczegółowe dane, wraz z liczbą kolizji, jednakże brak było porównywalnych danych dla skrzyżowań referencyjnych. Niedoskonałość tę może ograniczyć wprowadzenie Systemu Ewidencji Wypadków i Kolizji SEWIK.

z ich bezpośredniego otoczenia nie zdjęto rygorystycznych oznaczeń dotyczących ograniczenia prędkości lub czy po przebudowie nie zdecydowano się na usunięcie oznaczenia miejsca szczególnie niebezpiecznego.

Oczywiste jest, że zarówno zmienne słabo mierzalne, jak i ukryte mogą mieć większy wpływ na skrzyżowania przebudowane niż na skrzyżowania referencyjne właśnie z tego powodu, iż te pierwsze uległy przeobrażeniom, a drugie nie. Są to ograniczenia modelu, z których trzeba zdawać sobie sprawę, a które jednak nie sprawiają, że otrzymane wartości efektu netto nie dadzą się empirycznie oszacować. Otrzymane i opisane powyżej wartości składowych efektów netto zostały opracowane na podstawie pewnych dobrze określonych i łatwych do zmierzania zmiennych, dlatego też można mówić o sukcesach lub porażkach ocenianych projektów.

Zaproponowany model można również zastosować do oceny efektu netto poszczególnych, pojedynczych skrzyżowań, które wzięły udział w przebudowie. Poniżej (tabela 7 i 8) przedstawiono ostateczne wyniki dla dwóch ewaluowanych skrzyżowań, wyliczone zgodnie z podanym powyżej algorytmem⁷:

Tab. 7. Składowe całościowego efektu netto interwencji (skrzyżowanie 1)

	Składowe całościowego efektu netto interwencji (skrzyżowanie 1)		
	średnia wypadków rocznie	średnia liczba zabitych rocznie	średnia liczba rannych rocznie
Efekt netto	$67,3\% \cdot 0,8 = \mathbf{0,539}$ wypadku rocznie	$53,4\% \cdot 0,4 = \mathbf{0,213}$ osoby zabitej rocznie	$73,9\% \cdot 1,4 = \mathbf{1,035}$ osoby rannej rocznie

Tab. 8. Składowe całościowego efektu netto interwencji (skrzyżowanie 2)

	Składowe całościowego efektu netto interwencji (skrzyżowanie 2)		
	średnia wypadków rocznie	średnia liczba zabitych rocznie	średnia liczba rannych rocznie
Efekt netto	$-307,7\% \cdot 0,8 = \mathbf{-2,462}$ wypadku rocznie	$-232,3\% \cdot 0,2 = \mathbf{-0,465}$ osoby zabitej rocznie	$-211,8\% \cdot 1,4 = \mathbf{-2,996}$ osoby rannej rocznie

W powyższym przykładzie mieliśmy do czynienia z sytuacją, gdy dla pierwszego skrzyżowania nie zaobserwowano żadnego wypadku w 2007 roku, wypadkowość w latach 2001–2005 była zaś umiarkowana. W przypadku drugiego skrzyżowania doszło do sytuacji zgoła odwrotnej: podczas gdy statystyki dla lat 2001–2005 utrzymywały się na niewielkim poziomie (zbliżonym zresztą – jak widać – do statystyk dla skrzyżowania 1), dane dla roku 2007 były zatrważają-

⁷ W przypadku szacowania efektu netto dla pojedynczych skrzyżowań wykorzystano metodę kernel PSM, bazującą na wyliczeniu bliskości dla każdego skrzyżowania z osobną bliskości wszystkich skrzyżowań z grupy kontrolnej. Oznacza to *de facto*, że wyników dla poszczególnych, konkretnych skrzyżowań nie porównujemy z jednym skrzyżowaniem-bliźniakiem, ale ze wszystkimi skrzyżowaniami referencyjnymi, które przyjmują w stosunku do danego skrzyżowania różną wagę. Szerzej na temat tej metody: Konarski, Kornatowski 2007.

co wysokie. Stąd też bardzo wysokie wartości ujemne efektu netto w przypadku skrzyżowania 2.

Jak widać, szacowanie efektu netto na poziomie pojedynczych skrzyżowań jest możliwe, jednak niewielka liczba danych (chodzi tu głównie o niewielką liczbę danych zbieranych już po interwencji) może wpływać na rzetelność oszacowania, a co za tym idzie, może dawać w wyniku dość duże rozbieżności efektów netto dla poszczególnych skrzyżowań. Można jednak żywić przekonanie, że wraz z upływem czasu składowe efekty netto dla poszczególnych skrzyżowań będą zarówno mniej „rozstrzelone”, jak i rzetelniejsze.

Wnioski dotyczące zastosowanych metod statystycznych

W toku przygotowywania projektu opracowania uniwersalnej metody służącej do oszacowania efektu netto dla projektów infrastrukturalnych napotkano liczne braki, które uniemożliwiły w tym konkretnym przypadku analizę innych składowych efektów netto (np. tych z zakresu oddziaływania interwencji na środowisko – poziom hałasu, emisje zanieczyszczeń itp.). Wziąwszy pod uwagę, że badacze dysponowali wyłącznie bazą skonstruowaną na potrzeby projektu na podstawie kilku innych zbiorów danych, skupiono się na najbardziej oczywistej mierze efektu netto projektu przebudowywania skrzyżowań, czyli na wypadkowości i powiązanej z nią liczbie osób poszkodowanych.

Uniwersalność zaproponowanej metody umożliwiała jednak badanie innych składowych efektów netto, takich jak efekt netto ograniczenia uciążliwości akustycznej czy wpływu środowiskowego – pod warunkiem rzetelnego gromadzenia, a następnie udostępnienia ewaluatorom danych z tego zakresu.

Co więcej, dostęp do danych związanych z innymi aspektami efektów netto pozwoli w przyszłości na lepsze dopasowanie skrzyżowań referencyjnych. Można by określać inne aspekty podobieństwa skrzyżowań przebudowanych do skrzyżowań referencyjnych, czyli mówiąc językiem statystyki, poprawić (i tak już wysokie) dopasowanie jednostek badanych i jednostek kontrfaktycznych, stosując szerszy wachlarz zmiennych niezależnych.

Oprócz tego zaproponowany model umożliwia ocenę nawet pojedynczych skrzyżowań (czy też, szerzej: jednostek poddanych interwencji). Do tego celu niezbędny jednak jest szerszy horyzont czasowy badania.

Konstruując wskaźniki efektu netto dla wypadkowości, nie można pominąć specyfiki zdarzenia, jakim jest wypadek. W przypadku innych badań ewaluacyjnych mamy bowiem do czynienia z sytuacją, gdy jednostka poddana interwencji stale doświadcza korzyści związanej z udziałem w projekcie (np. osoba biorąca udział w szkoleniu znajduje pracę), bądź też wynik ma charakter ciągły i jest mierzalny na skali (np. szacowanie wpływu realizowanych projektów na fluktuacje związane z bezrobociem w danym regionie). W przypadku wypadkowości analizuje się zaś pojedyncze i nieprzewidywalne zdarzenie losowe. Mówiąc prostym językiem: w pewnym konkretnym miejscu wypadki się zdarzają. Częstotliwość tego zdarzenia jest dość niska, w związku z czym do rzeczywistego oszacowania

długofalowego efektu netto należałoby zebrać dane z kilku kolejnych lat i dopiero poddać je analizie według założeń zaproponowanej metody.

Uwaga ta dotyczy przede wszystkim oceny pojedynczych skrzyżowań, w mniejszym stopniu zaś oceny całego projektu.

Przygotowany model, bazujący na metodzie PSM, będącej z kolei pochodną analizy regresji logistycznej, nadaje się do przewidywania prawdopodobieństwa sukcesu (wzięcia udziału w programie) również w przypadku innych rodzajów projektów infrastrukturalnych, choć nie wszystkich. Należy przy tym pamiętać, że metoda ta ma również swoje ograniczenia i w pewnych warunkach dopasowanie jednostek referencyjnych może być obciążone nieakceptowanym poziomem błędu, bądź też analiza może nie dać w ogóle rzetelnych (czy konkluzywnych) wyników. Stałoby się tak w dwóch przypadkach:

- gdyby baza danych, z której dobierane są jednostki referencyjne, była zbyt uboga w rekordy;
- gdyby podobieństwo jednostek poddanych interwencji oraz referencyjnych określane miało być na podstawie zmiennych słabej jakości.

W przygotowanym modelu zadbane o to, by liczba przypadków, z których wybierane były jednostki referencyjne, była dość obszerna, ponadto w poszukiwaniu skrzyżowań referencyjnych badacze posłużyli się szeroką gamą różnorodnych zmiennych, określających nie tylko specyfikę położenia danego skrzyżowania i jego budowę, lecz także przepustowość i liczbę zaobserwowanych wypadków w okresie do 2005 roku.

W razie stosowania zaproponowanego modelu do innych badań ewaluacyjnych należałoby zadbać o wystarczającą liczbę przypadków, do których dobierane byłyby jednostki kontrfaktyczne, liczbę i jakość zmiennych, na których opisane byłyby jednostki, oraz, jeśli to możliwe, wzięcie pod uwagę szerokiego horyzontu czasowego: zarówno jeśli chodzi o okres przed interwencją, jak i po jej realizacji.

Krytyczna analiza metody mierzenia efektów netto w odniesieniu do inwestycji z zakresu infrastruktury drogowej

W literaturze przedmiotu krytyka szacowania efektów netto opiera się na dwóch głównych nurtach. Z jednej strony w wątpliwość podawana jest słuszność założenia, iż badacze, stosując podejście zaczerpnięte z nauk ścisłych, są w stanie określić jasną relację przyczynowo-skutkową między nakładami inwestycyjnymi a wybranymi efektami. Wynika to z przekonania, że interwencje publiczne najczęściej realizowane są w bardzo złożonym otoczeniu – wielowymiarowym systemie społeczno-ekonomicznym, w którym relacje są wielokierunkowe, a liczba czynników oddziałujących na dane zjawisko jest duża i zmienna w czasie. Ponadto same inwestycje mają charakter unikatowy i niezmiernie trudno jest odnaleźć jednostki, z którymi mogłyby zostać porównane. Możliwości obiektywnego, jednoznacznego stwierdzenia kierunku i siły relacji przyczynowo-skutkowych, które są wypadkową wielu procesów, praktycznie nie ma. Rónocześnie

wskazuje się słabość i brak praktycznej użyteczności (a nawet potencjalną szkodliwość) szacowania „efektów netto” w zakresie oceny interwencji publicznych. Metoda ta pomija wiele istotnych aspektów, które powinny stanowić podstawę oceny skuteczności i użyteczności danej inwestycji, tj. określenie efektów nieprzewidzianych, ubocznych, dodanych, czy identyfikację przyczyn takiego a nie innego stanu rzeczy (odpowiedź na pytanie, dlaczego dana inwestycja się powiodła lub nie). Podkreśla się, że nie wszystkie ujemne efekty netto muszą oznaczać porażkę danej interwencji, gdyż pozytywne wyniki mogą przejawiać się w innych obszarach, których nie wzięto pod uwagę w badaniu.

Do zasadności wyżej przytoczonych argumentów krytycznych pomoże ustosunkować się analiza wiarygodności i szerszej użyteczności zastosowanej metody do mierzenia efektów netto w odniesieniu do infrastruktury drogowej. Analizę tę autorzy przeprowadzą w dwóch wymiarach. Po pierwsze, proponowana metoda zostanie zweryfikowana na podstawie głównych klasycznych pojęć oceniających jakość metodyki ewaluacji (stosowanych szczególnie przy podejściach „twardych”). Po drugie, refleksji poddana zostanie kwestia, czy i w jakim stopniu stosowanie metody jest obciążone ryzykiem popełnienia błędów ewaluacyjnych czterech rodzajów.

Jakość zastosowanej metody zostanie oceniona na podstawie wybranych obszarów:

1) **Poprawność konstrukcji** (*construct validity*) – czy stosując metodę efektu netto, mierzymy właściwe zjawisko, czy też pomijamy jego ważną część.

Z założenia metoda szacowania efektu netto danej interwencji wymaga wstępnego określenia obszarów, które poddane zostaną badaniu – zmiany mierzymy w konkretnych aspektach za pomocą określonych wskaźników odzwierciedlających te zmiany w wartościach ilościowych. Metoda ta daje możliwość udzielenia odpowiedzi na pytanie, jak dana inwestycja wpłynęła na wybrane wskaźniki (np. bezpieczeństwa, płynności ruchu, poziomu zanieczyszczeń emitowanych do środowiska, poziomu hałasu itp.) – nie pozwala natomiast określić efektów innych niż wybrane i skwantyfikowane na etapie konceptualizacji badania.

2) **Zewnętrzna użyteczność** (ang. *external validity, generalizability*) – czy rezultaty badania można generalizować.

Wyników badań dokonanych metodą szacowania efektów netto nie można generalizować, gdyż – mimo wykorzystania narzędzi analiz ilościowych – stosowana metodologia bliska jest tej, na której opierają się badania typu studium przypadku. Ewaluowane inwestycje/projekty/programy traktuje się w sposób indywidualny (na tym założeniu opiera się dobór jednostek kontrfaktycznych do porównań). Jakościowy charakter badań z zasady nie pozwala na uogólnianie wysuwanych wniosków, ponad uogólnienia typowe dla studiów przypadku.

3) **Wewnętrzna użyteczność** (ang. *internal validity*) – czyli siła, z jaką można ustalić relację przyczynową.

Możliwość ustalania relacji przyczynowych stanowi słabą stronę proponowanej metody ze względu na złożoność otoczenia i wynikającą z niej mnogość czynników oddziałujących na badane zjawisko. Kluczowe dla prawidłowej oce-

ny danej inwestycji/programu okazać się może umiejętne określenie i oszacowanie wpływu czynników niemierzalnych przez zespół ewaluatorów. Ocena ta musiałaby z jednej strony opierać się na dalszych badaniach, z drugiej – na wiedzy i doświadczeniu ewaluatorów w zakresie funkcjonowania infrastruktury drogowej (konieczny udział ekspertów z tej dziedziny w zespole ewaluacyjnym lub zlecenie ekspertyzy ekspertom zewnętrznym). Potrafiliby oni wyjaśnić przyczyny i czynniki, za sprawą których projekt przyniósł takie a nie inne efekty, i tym samym odpowiedzieć na kluczowe dla decydentów pytanie: dlaczego projekt się powiódł lub odniósł porażkę, i jakie z tego płyną wnioski na przyszłość. Zastosowanie metody powinno w tym aspekcie stanowić bazę wyjściową do dalszych badań przy użyciu innych narzędzi badawczych – co jest zresztą dobrą i zalecaną przez KE praktyką prowadzenia badań ewaluacyjnych – odzwierciedlającą zasadę triangulacji zbieranych danych. Należy jednak podkreślić, że pomiar efektu netto może stanowić cenną wskazówkę przy wyborze projektów do dalszych, jakościowych badań ewaluacyjnych.

4) **Użyteczność wniosków** (ang. *conclusion validity*) – czy otrzymane wyniki poprawnie wykrywają obecność czynnika, zależności czy efektu w danej skali i natężeniu.

W tym zakresie metoda oceny efektu netto okazuje się adekwatna, co z kolei zapewnia oparcie wyliczeń na wybranych i odpowiednio skonstruowanych wskaźnikach.

W tym kontekście należy też rozważyć, czy stosowanie metody jest obciążone wysokim ryzykiem popełnienia:

- **błędu I rodzaju** (*type 1 error* – ewaluator stwierdza, że efekty dla badanych grup różnią się, podczas gdy w rzeczywistości tak nie jest);
- **błędu II rodzaju** (*type 2 error* – ewaluator stwierdza, że program nie przyniósł efektów, podczas gdy efekty w rzeczywistości nastąpiły).

W praktyce nie można zanegować możliwości wystąpienia sytuacji określanej przez badaczy jako błąd I lub II rodzaju, niemniej jednak prawdopodobieństwo wystąpienia tych błędów w przypadku szacowania efektu netto ocenia się jako praktycznie niemożliwe ze względu na charakter badań – jakościowy, bez możliwości przeniesienia wyników badań z próby na populację.

Oczywiście ryzyko popełnienia błędu II rodzaju jest dość duże, jednak niższe niż w przypadku pomiaru efektu brutto przy użyciu tych samych wskaźników. Wynika to z dwóch omówionych wcześniej przyczyn – po pierwsze możliwości zaistnienia niemierzalnych lub incydentalnych czynników, które wpłyną na otrzymane wyniki badań, po drugie koncentracji na wybranych aspektach wpływu danej interwencji, co prowadzi do nieuwzględnienia efektów pobocznych (dodatkowych, niezamierzonych). Niemniej jeżeli przyjmemy, że podstawą oceny projektów powinien być stopień osiągnięcia dzięki ich realizacji zakładanych celów, to problem nieuwjęcia efektów dodatkowych nie jest już kluczowy i z powodzeniem może zostać rozwiązany w toku dodatkowych badań. Zaletą metody w tym aspekcie pozostaje natomiast fakt porównania jednostki poddanej interwencji z jednostką kontrfaktyczną, co w przypadku otrzymania ujemnego efek-

tu netto daje podstawy dla przeanalizowania (w toku dalszych badań o charakterze jakościowym – np. wizji lokalnych, wywiadów z członkami społeczności lokalnych, w których realizowana była interwencja) przyczyn sukcesu jednostki porównywanej;

– **błędów typu III i IV** (mierzenie czegoś, co nie da się zmierzyć lub nie istnieje, oraz mierzenie czegoś, co nie ma znaczenia z perspektywy decydentów).

Ryzyko popełnienia tego typu błędów jest znikome, gdyż efekt netto danej inwestycji mierzy się w wybranym obszarze, wykorzystując konkretne wskaźniki, a uzyskane w ten sposób wskaźniki dostarczają nam rzetelnych danych o powodzeniu lub porażce danej interwencji, z tego źródła zaś – przy założeniu uzupełnienia wyników o przyczyny takiego stanu rzeczy – decydenci powinni czerpać wiedzę przy planowaniu podobnych projektów.

Wnioski końcowe

Wnioski do adaptacji metody do badań ewaluacyjnych

Podsumowując podjęte w artykule rozważania, należy stwierdzić, że metoda szacowania efektów netto może z powodzeniem być stosowana w badaniach ewaluacyjnych wybranych projektów z zakresu infrastruktury drogowej. Niemniej należy mieć na uwadze szereg ograniczeń związanych przede wszystkim z dostępnością danych (brakiem rzetelnych baz danych sprzed realizacji interwencji i po niej, w szczególności dotyczących stanu kontrfaktycznego). Istotnym problemem jest niemożność uwzględnienia czynników niemierzalnych lub incydentalnych, które bywają kluczowe w przypadku wydarzeń na drodze. Trudność ta traci na znaczeniu, gdy badaniu efektu netto poddane będą zjawiska takie jak zmiany poziomu hałasu czy zanieczyszczeń emitowanych do środowiska.

W chwili obecnej zasadne wydaje się mierzenie efektu netto w przypadku nowych, dopiero planowanych do wdrożenia projektów polegających na przebudowie danych odcinków dróg.

W przypadku inwestycji drogowych przed ich rozpoczęciem, a po akceptacji wniosku, należy stworzyć bazę obiektów podobnych oraz dobrać obiekty obrazujące stan kontrfaktyczny, a także zmierzyć wybrane wskaźniki (płynność ruchu) dla wszystkich obiektów. Następnie po kilku latach należy ponownie przeprowadzić pomiar dla wszystkich obiektów.

Możliwe jest także przeprowadzenie ewaluacji *ex post*, obejmującej wpływ inwestycji liniowych (np. przebudowy dróg) na bezpieczeństwo ruchu drogowego. Tego typu badania mogą być przeprowadzone już w niedalekiej przyszłości; konieczne jest upłynięcie ok. trzech lat od zakończenia inwestycji w celu zgromadzenia odpowiedniej liczby rzetelnych danych, zwłaszcza dotyczących tak rzadkich zjawisk, jak wypadki śmiertelne (szczególnie na krótszych odcinkach dróg). Obecnie pożądane byłoby także badanie efektu netto w przypadku zrealizowanych już projektów dotyczących szerokiej modernizacji systemów komunikacji miejskiej. W tym zakresie nie przeprowadzano w Polsce badań ewaluacyjnych,

natomiast istniejący system wskaźników jest prawdopodobnie niedoskonały ze względu na szereg czynników zewnętrznych i wewnętrznych, kształtujących wielkość przewozów. Stąd też pomiar efektu netto może okazać się odpowiednią metodyką ewaluacji. Z technicznego punktu widzenia ten typ projektów stanowi stosunkowo liczną grupę (i wciąż jeszcze istnieją systemy niepoddane interwencji), natomiast wskaźniki oddziaływania są silnie obciążone działaniem czynników zewnętrznych. W tym przypadku prawdopodobnie konieczne będzie jednak opisanie stanu kontrfaktycznego przy użyciu danych z kilku innych miast, ze względu na zróżnicowanie lokalnej polityki transportowej. Należy przy tym pamiętać, że statystyczny dobór obiektów podobnych (jako wstępna selekcja) musi być wsparty ich jakościową analizą ekspercką, tak by zarówno zweryfikować podobieństwo, jak i dobrze poznać przyczyny wielkości efektu brutto (np. inne interwencje na szczeblu lokalnym).

Użyteczność metodologii w badaniach ewaluacyjnych

Bez wątpienia metodologia szacowania efektów netto realizowanych projektów wzbogaca wachlarz stosowanych w badaniach ewaluacyjnych narzędzi. Należy podkreślić, że ze względu na wskazane w artykule ograniczenia metoda ta powinna być stosowana nie jako samodzielne narzędzie badawcze, ale raczej jako instrument dodatkowy, wnoszący istotne informacje na temat skuteczności realizowanych projektów. Wynika to z braku możliwości wskazania zjawisk przyczynowo-skutkowych oraz dodatkowych efektów poddawanych ocenie, nieprzewidzianych na etapie konceptualizacji projektów.

Metoda dostarcza rzetelnych informacji, które powinny stanowić bazę do projektowania dalszych narzędzi badawczych o charakterze zarówno jakościowym, jak i ilościowym (np. wywiady ze społecznością lokalną, ankiety wśród użytkowników dróg, opinie ekspertów itp.), które uzupełnią otrzymane wyniki, np. o wskazanie relacji przyczynowo-skutkowych, a być może nawet ich siły. Takie podejście w pełni wpisuje się w dobrą praktykę badań ewaluacyjnych, opierającą się na zasadach triangulacji gromadzonych i analizowanych danych. Wykorzystanie proponowanej metody ewaluacji wpływu w istotny sposób może podwyższyć trafność podejmowanych przez decydentów interwencji.

Literatura

- Bruska A., 2009, „Funkcje ewaluacji a możliwości wykorzystania dorobku nauk o zarządzaniu”, w: *Ewaluacja programów operacyjnych na poziomie regionalnym – teoria i praktyka*, praca zbiorowa, Opole: Urząd Marszałkowski Województwa Opolskiego, Departament Polityki Regionalnej i Przestrzennej, Referat Ewaluacji.
- Górniak J., 2007, „Ewaluacja w cyklu polityk publicznych”, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Kraków: Uniwersytet Ekonomiczny w Krakowie, Małopolska Szkoła Administracji Publicznej.
- Komisja Europejska, 1997, *Ewaluacja programów wydatków Unii Europejskiej. Przewodnik*, styczeń 1997.

- Konarski M., Kornatowski M., 2007, „Zastosowanie metody PSM w ewaluacji ex-post”, w: A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, Warszawa: Polska Agencja Rozwoju Przedsiębiorczości.
- Lennie J., Perez M., Ammassari S., Bisgard J., 2006, *Definitions of Evaluation Types, Approaches and Fields*, Independent Evaluators' Webring.
- Michalek J., Zarnekow N., Stulrajter Z., Klank L., 2008, *Cartographic and Statistical Analysis of RDI in Poland and Slovakia. Regional Report RR3-2*, University of Kiel.
- Michalek J., 2008, materiały IV Konferencji Ewaluacyjnej „Ewaluacja – kluczowy instrument poprawy jakości działań administracji publicznej”, Warszawa, 17 października.
- Niebieska Księga. *Infrastruktura drogowa*, 2008, nowe wydanie, Warszawa: Jaspers.
- Ridder M. i in., 2003, *Halbzeitbewertung des Operationellen Programms Verkehrsinfrastruktur 2000–2006 für das Deutsche Zeil-1-Gebiet*, Münster: MR, Gefra, IVM, Delmenhorst.
- Symon E., 2008, *Wypadki drogowe w Polsce w 2007 roku*, Warszawa: Komenda Główna Policji.
- Wolański M., Gil M., Pander W., 2008, *Wpływ projektów dotyczących zakupu lub modernizacji pojazdów szynowych na osiągnięcie głównego celu poddziałania 1.1.2, tzn. poprawę warunków przejazdów pasażerów transportem kolejowym między aglomeracjami miejskimi i w aglomeracjach*, Warszawa: CASE-Doradcy i EGO.

Strony internetowe:

- Detailed Information on the Amtrak Assessment, <http://www.whitehouse.gov/expectmore/detail/10004000.2005.html#performancemeasures>, dostęp w dniu 10.11.2008.
- Detailed Information on the Railroad Safety Program Assessment, <http://www.whitehouse.gov/omb/expectmore/detail/10001127.2003.html>, dostęp w dniu 10.11.2008.
- Detailed Information on the Federal Motor Carrier Safety Administration Grant Program Assessment, <http://www.whitehouse.gov/omb/expectmore/detail/10000410.2008.html>, dostęp w dniu 10.11.2008.
- Detailed Information on the Highway Infrastructure Assessment, <http://www.whitehouse.gov/omb/expectmore/detail/10000412.2007.html>, dostęp w dniu 10.11.2008.

IMPACT EVALUATION OF EU-CO-FINANCED PROJECTS — A CASE STUDY OF CROSSROADS RECONSTRUCTION

The article presents possibilities of adapting the impact evaluation methodology to the evaluation of public intervention in road infrastructure. In the first part of the article the authors present the principles of the impact evaluation methodology, which serves to evaluate real effects of an intervention, as well as examples of evaluation projects prepared with the use of this method. The authors present their own empiric study, which was the first application of the methodology in the road infrastructure sector. The article concludes with a critical analysis of the method, especially concerning its reliability and potential usefulness in road infrastructure.