

*Monika Maksim\**

## WYBRANE PROBLEMY OCENY AKTYWNYCH PROGRAMÓW RYNKU PRACY

### 1. WSTĘP

Aktywna polityka rynku pracy stanowi obok polityki makroekonomicznej i polityki elastyczności i deregulacji rynku pracy jeden z kierunków polityki państwa skierowanych na ograniczanie nierównowagi na rynku pracy. Pomimo licznych uwag krytycznych pod jej adresem pozostaje ona potencjalnie ważnym narzędziem walki z bezrobociem. Aktywna polityka rynku pracy dysponuje szerokim wachlarzem instrumentów, które bardzo często aplikowane były w zasadzie w dowolny sposób, bez dokonania wcześniej pogłębionej analizy problemów występujących na rynku pracy i czynników je powodujących. W rezultacie poszczególne programy nie zawsze kierowane były do osób najbardziej ich potrzebujących, a ich faktyczne efekty były zazwyczaj niższe od oczekiwanych. W chwili obecnej podejście do polityki rynku pracy ulega istotnym przemianom. Przede wszystkim kładzie się większy nacisk na działania prewencyjne. W celu lepszego adresowania programów coraz częściej wykorzystuje się techniki profilowania bezrobotnych. Obserwuje się ponadto wzrost innowacyjności w zakresie stosowanych instrumentów. Polityka ta stopniowo odchodzi od przedsięwzięć o charakterze reaktywnym zmierzając jednocześnie w kierunku działań wspierających proaktywne postawy bezrobotnych na rynku pracy. Dąży się przy tym do większej indywidualizacji wykorzystywanych instrumentów poprzez upowszechnienie metodyki opartej na indywidualnych planach działania, a także znaczącą wagę nadaje się szkoleniom i przekwalifikowaniom zawodowym dla poszukujących pracy. Oprócz wyżej wspomnianych zmian mających na celu poprawę skuteczności aktywnej polityki, istotnym elementem tego nowego podejścia powinny być obligatoryjne badania ewaluacyjne służące wiarygodnej ocenie realizowanych programów.

Coraz częściej mówi się o konieczności ekspansji aktywnej polityki rynku pracy, a przede wszystkim o przypisaniu jej roli nadrzędnej w stosunku do rozwiązań pasywnych. Wydatki na aktywną politykę mierzone udziałem w PKB

---

\* Mgr, Katedra Gospodarowania Zasobami Pracy, Uniwersytet Mikołaja Kopernika w Toruniu.

oscylują w krajach Piętnastki w przedziale 0,5 –1,5%, natomiast przeciętny ich udział w wydatkach na politykę rynku pracy ogółem kształtuje się na poziomie 40% (Martin, Grubb 2001). Nie należy się jednak spodziewać, że w najbliższym czasie nakłady na aktywne formy zostaną w znaczącym stopniu zwiększone. Po pierwsze w warunkach wysokiego bezrobocia a więc i większego zapotrzebowania na zasiłki możliwości transferu środków z form pasywnych na rzecz aktywnych są znacznie ograniczone. Po drugie, wątpliwe efekty aktywnych programów wynikające po części z niedoskonałości stosowanych metod ewaluacji nie dostarczają argumentów przemawiających za zwiększeniem nakładów na tę politykę. Badania oceniające programy dają niejednokrotnie sprzeczne rezultaty, co utrudnia decydom wyciąganie prawidłowych wniosków i podejmowanie odpowiednich działań. Po trzecie, na co zwracają uwagę de Koning i Mosley (2002), istnieją jeszcze rezerwy wzrostu efektywności aktywnych programów rynku pracy, które można wykorzystać poprzez działania polegające między innymi na lepszej alokacji wydatków na te programy, usprawnieniu systemów zarządzania publicznymi służbami zatrudnieniowymi, wprowadzeniu zasady finansowania zorientowanego na efekty, benchmarkingu działalności publicznych służb zatrudnieniowych, czy nawet kontraktowaniu usług zatrudnieniowych i prywatyzacji. Trzeba zauważyć jednak, że punktem wyjścia do podejmowania tego typu przedsięwzięć jest systematyczna ocena aktywnej polityki rynku pracy. W wielu krajach, szczególnie tych które ostatnio dołączyły do Wspólnoty ewaluacja polityki rynku pracy nadal pozostaje w obszarze o znaczeniu marginalnym. Również o ile w krajach Europy Zachodniej metody ewaluacji są coraz lepiej rozpowszechnione – o tyle niepokojące jest, że w krajach byłego Bloku Wschodniego, w tym także w Polsce wykorzystywane są stosunkowo rzadko, szczególnie na płaszczyźnie praktycznej.

## 2. CELE I TYPY BADAŃ EWALUACYJNYCH

Szczyt rozwoju badań w zakresie oceny efektów aktywnych programów rynku pracy przypada na koniec lat 80. i lata 90. ubiegłego stulecia. Ogólnie ujmując, ewaluacja oznacza badanie charakterystyk, zalet oraz wad ewaluowanego programu. Dostarcza ona informacji na temat efektywności badanego przedsięwzięcia, w celu zoptymalizowania jego wyników, jakości i sprawności działania. Zawężając tę definicję do aktywnej polityki rynku pracy stwierdza się, że ewaluacja polega na wykorzystaniu metod i technik badań społecznych do pomiaru efektów programów działania państwa nakierowanych na zwalczanie bezrobocia. Może swym zakresem obejmować ocenę zasadności wprowadzenia danego programu, projekt i wdrożenie, wpływ programu i jego

efekty oraz efektywność wyrażaną jako relację korzyści do kosztów programu (Rossi, Lipsey, Freeman 2003). Biorąc pod uwagę aktywną politykę rynku pracy chodzi przede wszystkim o uzyskanie odpowiedzi na pytania: czy programy te są efektywne, jaki faktyczny wpływ wywierają, czy też jakie skutki powodują z punktu widzenia zatrudnienia? Innymi słowy studia ewaluacyjne umożliwiają badanie celowości i efektywności aktywnych form przeciwdziałania bezrobociu. Badania ewaluacyjne dostarczają informacji, które mogą być wykorzystane do różnych celów, w tym między innymi do (World Bank 2002):

- projektowania i konstruowania nowych programów,
- udoskonalania programów już realizowanych,
- lepszego adresowania programów do grup celowych,
- identyfikowania i eliminowania programów nieefektywnych.

W porównaniu do ocen ukierunkowanych na program (*programme – oriented analysis*) preferowane obecnie oceny zorientowane na cel (*target-oriented policy evaluation*) mają znacznie bardziej złożony i kompleksowy charakter (Schmid, O'Reilly, Schömann 1996). Prowadzone są w szerszej perspektywie, obejmują studia na poziomach mikro i makroekonomicznym z wykorzystaniem szeregów czasowych i danych przekrojowych uwzględniających okresy przed i po wprowadzeniu programu. Polegają na krytycznej analizie celów, procesu implementacji i efektów końcowych tych programów.

W literaturze przedmiotu rozróżnia się następujące typy ewaluacji (por. European Comission 2003; Purdon, Lessof, Woodfield, Bryson 2001; Kwiatkowski 1998):

- ocenę celów i wdrożenia programów (*process evaluations*),
- monitoring skuteczności (*performance monitoring*),
- ocenę efektów netto programów (*impact evaluations*),
- analizę kosztów i korzyści programów (*cost-benefit analysis*).

Pierwszy z wymienionych typów ewaluacji polega między innymi na analizie celów realizowanych programów pod kątem ich zgodności z priorytetami polityki społeczno-gospodarczej. W ramach tej oceny bada się przede wszystkim wszelkie aspekty organizacyjne związane z funkcjonowaniem programu, a także działania, jakie podejmowane są w trakcie jego implementacji. W badaniach ewaluacyjnych tego typu wykorzystuje się metodę sondażową, przeważnie ilościową, a także metody jakościowe oparte na wywiadach pogłębionych, grupach dyskusyjnych (*focus groups*) i studiach przypadków.

Pojęcie monitorowania polityki rynku pracy wymaga większej precyzji terminologicznej. Chociaż w literaturze przedmiotu monitoring uchodzi za jeden z typów badań ewaluacyjnych, coraz częściej wskazuje się na cechy odróżniające go od klasycznej ewaluacji. Różnice w postrzeganiu systemu monitorowania można wyjaśnić w następujący sposób. Po pierwsze, monitoring dostarcza bogatej informacji na temat stopnia zaawansowania realizacji

poszczególnych programów oraz podejmowanych działań i ich wyników, ale nie wyjaśnia dlaczego dany program jest skuteczny bądź nie. Odpowiedzi na takie pytania daje natomiast systematyczna ewaluacja. Na system monitorowania programów rynku pracy składają się specjalnie opracowane wskaźniki efektywności wyrażające stopień realizacji zakładanych celów poszczególnych programów. Pokazują one jednak tylko efekty brutto działania programów, gdyż uzyskane wyniki nie obejmują efektów pozornych. Po drugie, w Stanach Zjednoczonych i Kanadzie badania ewaluacyjne aktywnych programów zatrudnieniowych mają już dosyć długą tradycję, podczas gdy w Europie jeszcze do niedawna polegały właśnie przede wszystkim na monitorowaniu realizacji programów przy pomocy finansowych i fizycznych wskaźników (Auer, Kruppe 1996). Istotne zmiany w podejściu do problemu ewaluacji aktywnej polityki rynku pracy w Europie zapoczątkowały dopiero kraje takie jak Wielka Brytania i Szwecja (European Commission 2003).

Podkreśla się jednak, że monitoring może być efektywną formą kontroli wdrażanych programów rynku pracy, gdyż w odróżnieniu od ewaluacji, która jest procesem czasochłonnym, w miarę szybko zapewnia informację zwrotną o efektach poszczególnych instrumentów. Dlatego też panuje zgoda co do tego, że system monitorowania programów rynku pracy powinien stanowić wyjściową bazę potrzebną do prowadzenia dalszych ewaluacji.

Natomiast badania efektów netto aktywnych programów rynku pracy rozwinęły się w dwóch kierunkach. Obok analiz na poziomie mikroekonomicznym podejmowane są również przedsięwzięcia badawcze, które koncentrują się na ocenie efektów zagregowanych (*aggregated impact analysis*). Te pierwsze ograniczają się do zbadania wpływu uczestnictwa jednostek w aktywnych programach zatrudnieniowych na ich pozycję na rynku pracy. Pozycja ta może być określana przez ich status na tym rynku po zakończeniu programu przy pomocy mierników ilościowych (na przykład czy jednostka znalazła zatrudnienie, czy też nie, długości okresu pozostawania bez pracy bądź też wysokości osiągniętych dochodów) jak i jakościowych odnoszących się na przykład do rodzaju umowy o pracę, związku wykonywanej pracy z przedmiotem szkolenia, stopnia satysfakcji z pracy (OECD 1995; O'Leary, Nesporova, Samorodov 2001). Te drugie zmierniki odnoszą się do określenia wpływu rozwoju aktywnych programów na wybrane zmienne makroekonomiczne, którymi są najczęściej bezrobocie i stopa bezrobocia, zatrudnienie, odpływy z bezrobocia do regularnych miejsc pracy oraz płace. Za zmienne określające rozwój programów zatrudnieniowych przyjmuje się liczbę uczestników tych programów lub wydatki potrzebne na ich sfinansowanie (szerzej na ten temat zob. Bellmann, Jackman 1996). Do analiz na poziomie makroekonomicznym wykorzystuje się często model zmodyfikowanej funkcji dopasowań (Kwiatkowski, Tokarski 1997; Puhani 1999b).

Natomiast analizy kosztów i korzyści (*cost-benefit analysis*) należą do tych bardziej tradycyjnych sposobów oceny efektywności instrumentów aktywnej polityki rynku pracy (Delander, Niklasson 1996). Zmierzają do zidentyfikowania wszystkich kosztów i korzyści, jakie wynikają z programu w celu oszacowania jego całościowego efektu. Do korzyści zalicza się efekty netto odnoszące się do pojedynczych uczestników programu, jak również te, które ujawniają się w skali całej gospodarki. Natomiast koszty obejmują wszystkie wydatki związane z funkcjonowaniem programu oraz niezamierzone uboczne skutki działania tego programu, których zasięg oddziaływania może wykraczać poza grono osób biorących w nim udział (Purdon i inni 2001).

Z zaprezentowanej wyżej metodologii wynika, że istnieją różne podejścia do problemu ewaluacji aktywnych programów zatrudnieniowych i różne sposoby ich oceny. Poniższa tabela stanowi próbę usystematyzowania typów badań ewaluacyjnych aktywnych programów rynku pracy (tab. 1).

Tabela 1

## Typy badań ewaluacyjnych aktywnych programów rynku pracy

Typ ewaluacji		Podejście do ewaluacji	
		Oceny zorientowane na program	Oceny zorientowane na cel
Ocena celów i wdrożenia programów ( <i>process evaluations</i> )		-	+
Monitoring skuteczności		+	+
Ocena efektów netto ( <i>impact evaluations</i> )	Na poziomie mikroekonomicznym ( <i>microevaluations</i> )	-	+
	Na poziomie makroekonomicznym, ocena efektów zagregowanych ( <i>macroevaluations</i> )	-	x
Analiza kosztów i korzyści ( <i>cost-benefit analysis</i> )		+	+

- ten typ ewaluacji nie jest wykorzystywany
- + ten typ ewaluacji jest wykorzystywany
- x ten typ ewaluacji może ale nie musi być wykorzystywany

Źródło: opracowanie własne

Decydując się na określony typ ewaluacji *ex post* trzeba pamiętać, że nie wszystkie efekty ujawniają się w tym samym czasie, w stosunkowo krótkim okresie od zakończenia programu. Standardowo badania ewaluacyjne prowadzone są 3-, 6-, 12 i 24 miesiące po programie, chociaż uchwycenie niektórych efektów, w szczególności tych niezamierzonych może wymagać dłuższej perspektywy czasowej. Przyjęcie do badań odpowiedniego horyzontu czasowego jest jednym z czynników warunkujących rzetelność i jakość prowadzonych ewaluacji. Oceny dokonane zbyt wcześnie mogą dawać sprzeczne rezultaty i w związku z tym często postrzegane są jako mało wiarygodne.

### 3. METODOLOGICZNE ASPEKTY OCENY EFEKTÓW NETTO AKTYWNYCH PROGRAMÓW ZATRUDNIENIOWYCH NA POZIOMIE MIKROEKONOMICZNYM

Nie negując znaczenia pozostałych rodzajów ewaluacji, można pokusić się o stwierdzenie, że najważniejszym obszarem oceny aktywnych programów jest ocena ich efektów końcowych. Zagadnienie wpływu uczestnictwa bezrobotnych w tych programach na ich przyszłą pozycję na rynku pracy budzi w środowiskach naukowych wiele kontrowersji. Szacunki efektów netto są skomplikowane i trudne, niektórzy sądzą nawet, że ich dokładne wyodrębnienie jest po prostu niemożliwe. Spory dotyczą więc doboru odpowiednich metod, które miałyby dostarczać rzetelnych wyników.

Badania efektów netto aktywnych programów rynku pracy mają stosunkowo krótką historię i bazują w dużej mierze na dorobku naukowym znanego noblisty w dziedzinie ekonomii Jamesa Heckmana. W Stanach Zjednoczonych, gdzie mają one swój początek, analizowane są przeważnie efekty dochodowe aktywnych programów, natomiast w Europie Zachodniej, jak i w krajach, które niedawno przeszły transformację kładzie się przede wszystkim nacisk na efekty zatrudnieniowe. Różnica w wyborze odpowiednich miar sukcesu programu wynika prawdopodobnie z odmiennej sytuacji na rynku pracy w Stanach Zjednoczonych i poszczególnych krajach europejskich. W tych ostatnich, mianowicie, zagrożenie długookresowym bezrobociem jest poważne, a aktywne programy są między innymi nakierowane na ograniczanie tego uciążliwego zjawiska.

Mikroewaluacje koncentrują się na pomiarze efektów określanych jako wartość dodatkowa (*additionality*), którą generują aktywne programy. Innymi słowy zmierzają więc one do uzyskania odpowiedzi na fundamentalne pytanie postawione przez Heckmana a mianowicie (Heckman, Smith 1996): *Jakie są efekty programu w stosunku do uczestniczących w nim osób?* Lub inaczej *Jak w wyniku działania programu zmienia się pozycja tych osób na*

ryнку pracy w relacji do pozycji, jaką mieliby, gdyby nie uczestniczyli w programie?

Z tak postawionego pytania wynika, że do zbadania faktycznego wpływu programu niezbędne są informacje nie tylko o efektach, jakie mają miejsce w przypadku, gdy program działa, ale także o tych efektach, jakie pojawiłyby się, gdyby program nie był realizowany. Obserwowanie zachowań osób będących jednocześnie w dwóch różnych stanach jest niemożliwe. Powoduje to, że sytuacji uczestnictwa osób w programie przeciwstawia się sytuację odwrotną, która jest stanem hipotetycznym i informuje o tym co wydarzyłoby się gdyby ów program nie był realizowany. W anglojęzycznej literaturze przedmiotu (Orr 1999; Purdon 2002; Puhani 1999a) określa się to terminem *counterfactual*. Chodzi więc o skonstruowanie właściwej płaszczyzny odniesienia dla ewaluowanego programu. Chociaż z punktu widzenia teorii zdefiniowanie sytuacji kontrfaktycznej jest rzeczą prostą, to w praktyce może wystąpić pewna niejednoznaczność. Przystępując do ewaluacji warto rozważyć, czy stanem alternatywnym dla badanego programu jest sytuacja, w której realizowane są inne programy o podobnym celu i charakterze, czy też sytuacja, w której nie stosuje się żadnych programów.

Oszacowanie faktycznego wpływu programu wymaga porównania wielkości ekonomicznych uzyskiwanych w sytuacji uczestnictwa bezrobotnych w programie, czyli tych rzeczywistych z wielkościami analogicznymi, jakie zostałyby osiągnięte w przypadku odwrotnym, czyli gdyby bezrobotni nie brali udziału w programie. Efekt realny programu jest więc różnicą efektów uzyskanych w dwóch skrajnie odmiennych sytuacjach. Przyjęta w pracy definicja efektu realnego programu (*causal effect*), jak również poniższe oznaczenia, które obecnie są powszechnie stosowane w literaturze przedmiotu pochodzą od D.B. Rubina (1974). Efekt realny można zapisać w następujący sposób:

$$\Delta_i = Y_{1i} - Y_{0i}, \quad (1)$$

gdzie:

$\Delta_i$  oznacza efekt realny programu w stosunku do  $i$ -tej osoby w nim uczestniczącej (*impact lub additionality*);  $Y_{1i}$  oznacza wynik uzyskany przez  $i$ -tą osobę (na przykład czy znalazła zatrudnienie czy też nie), w sytuacji gdy uczestniczy w programie;  $Y_{0i}$  oznacza wynik uzyskany przez  $i$ -tą osobę (czy znalazła zatrudnienie czy też nie), w sytuacji gdy nie uczestniczy w programie.

Teoretycznie każdej osobie trzeba jednocześnie przypisać zarówno odpowiednie wartości  $Y_1$  jak i  $Y_0$ . O ile określenie wielkości  $Y_1$  nie jest sprawą metodologicznie zbyt trudną, gdyż  $Y_1$  jest zmienną obserwowalną, o tyle kwestią problematyczną pozostaje estymacja  $Y_0$ , co oznacza bowiem oszacowanie efektów, jakie mogłyby pojawić się w przypadku nie uczestniczenia bezrobotnych w programie, a więc w sytuacji będącej pewnym stanem hipotetycznym określanym jako kontrfaktyczny. Jak już wspomniano żadna

osoba nie może znajdować się w dwóch stanach jednocześnie, w istocie więc jedynym obserwowalnym dla każdej jednostki efektem jest:

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}, \quad (2)$$

gdzie:  $D_i$  jest zmienną zerojedynkową oznaczającą uczestnictwo w programie.

W konsekwencji faktyczny wpływ programu nie jest nigdy bezpośrednio obserwowany. W tym sensie szacowanie efektów realnych aktywnych programów sprowadza się do problemu brakujących danych. W starszej literaturze na temat ewaluacji przyjmowano, że efekt programu jest taki sam w odniesieniu do każdej osoby w nim uczestniczącej. Przyjęcie założenia o homogeniczności efektów oznaczało, że:

$$\Delta_i = \Delta \text{ dla wszystkich } i.$$

Obserwowany w ostatnich latach konceptualny postęp w dziedzinie ewaluacji jest wynikiem odejścia od pierwotnych założeń i uznania, że programy generują różne efekty w stosunku do poszczególnych osób w nich uczestniczących, dlatego między innymi, że osoby te różnią się między sobą pod względem płci, wieku, wykształcenia, miejsca zamieszkania, czy też motywacji do poszukiwania pracy. Traktowanie efektów jako heterogeniczne oznacza w gruncie rzeczy uznanie składnika osobowego jako czynnika mającego istotny wpływ na efektywność programów (Smith 2000). Po odpowiednim przekształceniu równania (2) otrzymuje się:

$$Y_i = Y_{0i} + \Delta_i \cdot D_i \quad (3)$$

W zaprezentowanej metodologii do badania efektów netto trzeba przyjąć pewne ograniczenia. Zakłada się bowiem, że wpływ programu w odniesieniu do konkretnego uczestnika jest stały, co natomiast oznacza, że efekt uzyskiwany przez jedną osobę nie zależy od tego czy inne osoby również uczestniczą w programie. Przykładowo, zakłada się więc, że znalezienie zatrudnienia przez jednego bezrobotnego uczestniczącego w programie jest niezależne od tego czy inni i ilu bezrobotnych wzięło także udział w tym programie. W języku angielskim założenie to nosi nazwę *a stable-unit-treatment-value assumption* (Rubin 1980) i gwarantuje, że przeciętny efekt programu może zostać oszacowany niezależnie od liczebności i struktury populacji, którą poddano interwencji.

Heterogeniczność uzyskiwanych efektów programów zatrudnieniowych sprawia, że szacowane parametry mogą być różnie definiowane. Rzeczywisty wpływ programów analizuje się najczęściej przy pomocy dwóch parametrów. Pierwszy z nich określa się jako przeciętny efekt programu w odniesieniu do populacji bezrobotnych (*average treatment effect*), co można zapisać w następujący sposób:

$$E[\Delta_i] = E[Y_{1i} - Y_{0i}] = E[Y_{1i} - Y_{0i} | D_i = 1]P(D_i = 1) + E[Y_{1i} - Y_{0i} | D_i = 0]P(D_i = 0) \quad (4)$$



Ten parametr informuje o tym, jaki przeciętnie będzie oczekiwany efekt interwencji dla całej populacji bezrobotnych, czyli w stosunku do wszystkich osób mogących, choć niekoniecznie biorących udział w programie. Badaczy jednak częściej interesuje uzyskiwany efekt programu w odniesieniu do osób w nim uczestniczących, a nie do tych, którzy być może z takiej formy pomocy nigdy nie skorzystają. Dlatego najczęściej szacuje się przeciętny efekt programu w stosunku do jego uczestników (*average effect of the treatment on the treated*):

$$E[\Delta_i | D_i = 1] = E[Y_{1i} - Y_{0i} | D_i = 1] = E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 1] \quad (5)$$

Estymując ten parametr można poza tym dowiedzieć się, jaki prawdopodobnie byłby efekt programu, gdyby podobna grupa osób bezrobotnych została poddana interwencji. W przypadku gdyby efekty programów traktowane były jako homogeniczne, wyżej opisane parametry byłyby równoważne.

Jak już wyżej wspomniano, o ile  $E[Y_{1i} | D_i = 1]$  można stosunkowo łatwo oszacować wykorzystując dane o uzyskiwanych przez bezrobotnych uczestniczących w programie wynikach, o tyle trudnym zadaniem pozostaje estymacja  $E[Y_{0i} | D_i = 1]$ , gdyż  $Y_{0i}$  jest nieobserwowalne dla bezrobotnych, dla których  $D_i = 1$ . Sposobem na rozwiązanie tego problemu jest konstrukcja odpowiedniej grupy kontrolnej będącej lustrzanym odbiciem grupy uczestników programu. Oznacza to bowiem, że szacunek  $E[Y_{0i} | D_i = 0]$  będzie wiarygodny i przekonujący tylko wtedy, kiedy grupa kontrolna będzie identyczna jak grupa uczestników programu pod względem wszystkich czynników, które mogą wpływać na osiągane wyniki. Do czynników tych zalicza się cechy społeczno-demograficzne, cechy osobowe, wysokość osiąganych ostatnio dochodów, dotychczasowe uczestnictwo w programach rynku pracy i szkoleniach zawodowych, jak i uwarunkowania lokalne. Jak widać mogą być to cechy obserwowalne jak i nieobserwowalne, czyli takie, których nie można bezpośrednio zmierzyć. Jedyną cechą różniącą obie grupy jest to, że tylko jedna z nich bierze udział w programie. Ważną kwestią pozostaje rozróżnienie czy uczestnictwo w programie ma charakter obligatoryjny i wynika z przydziału, czy też udział ten jest dobrowolny. Może na przykład okazać się, że bezrobotni zgłaszający się do programu są po prostu osobami silnie dążącymi do zmiany swego statusu na rynku pracy i ich lepsza pozycja na tym rynku po zakończeniu programu w porównaniu z grupą kontrolną będzie w pewnej mierze wynikać z silniejszej motywacji.

Utworzenie właściwej grupy kontrolnej jest niezwykle trudne w praktyce, zazwyczaj powstaje efekt selekcji (Heckman, Robb 1985; Manski 1989) będący skutkiem istniejących obserwowalnych i/lub nieobserwowalnych różnic pomiędzy grupą osób uczestniczących i nie uczestniczących w programie, co zapisuje się w następujący sposób:

$$E[Y_{0i} | D_i = 1] \neq E[Y_{0i} | D_i = 0]. \quad (6)$$

Różnice te będą zakłócać prawidłowy szacunek generowanej przez program wartości dodatkowej, czy też jak inni określają jego efektu netto. Najwięcej kontrowersji związanych z ewaluacją dotyczy właśnie wyboru odpowiedniej grupy kontrolnej. Stosowane do ewaluacji metody powinny korygować istniejący efekt selekcji, gdyż w przeciwnym razie będzie on zaciemniać rzeczywisty efekt działania programu.

Na koniec warto wspomnieć jeszcze, że w dotychczasowych rozważaniach metodologicznych na temat ewaluacji programy zatrudnieniowe mają charakter w miarę jednolity, czy nawet homogeniczny. Programy rynku pracy stają się coraz bardziej różnorodne. Powstają kolejne odmiany stosowanych już instrumentów, bądź też tworzone są całkiem nowe ich rodzaje. Dążenie do większej indywidualizacji programów powoduje, że coraz częściej są one zróżnicowane pod względem czasu trwania, zawartych w nich treści merytorycznych i dostępnego oprzyrządowania. Bezrobotni, a szczególnie ci zaliczani do grup wysokiego ryzyka uczestniczą zazwyczaj nie w jednym, a w kilku różnych następujących po sobie programach. Tych aspektów metodologicznych nie uwzględniano w dotychczasowych badaniach, ograniczając się do ewaluacji pojedynczych programów ignorując przy tym ich różnorodność, bądź też zawężając analizy do zbioru w miarę jednolitych programów pomijając natomiast kwestie ich interakcji z innymi programami (Speckesser 2004). Coraz częściej wskazuje się bowiem na konieczność wypracowania takich rozwiązań metodologicznych, które uwzględnić będą aspekty związane z heterogenicznością programów zatrudnieniowych. Pierwsze próby takiej oceny podjęte zostały w pracach M. Lechnera (2001).

#### 4. WYBRANE METODY QUASI-EKSPERYMENTALNE<sup>135</sup>

W metodach quasi-eksperymentalnych dobór do obu badanych grup jest nielosowy, co wiąże się z występowaniem efektu selekcji. Jednak jak sama nazwa wskazuje istotą tych metod jest upozorowanie eksperymentu. Liczne ich odmiany wynikają bowiem z różnych sposobów tworzenia grupy kontrolnej i zastosowania różnych technik estymacji.

W większości badaniach ewaluacyjnych przeprowadzanych przy użyciu metod nieeksperymentalnych główny wysiłek kierowany był na opracowanie takiego estymatora, który rozwiązywały problem efektu selekcji. Na przełomie lat ekonometrycy zaproponowali wiele różnych estymatorów, jednak szacunki uzyskiwane przy ich pomocy były często niejednoznaczne i trudne do zinterpretowania. W swojej pracy z 1986 roku, uznanej później za jedną

---

<sup>135</sup> Pojęcia "metody quasi-eksperymentalne i "metody nieeksperymentalne są w niniejszym opracowaniu stosowane zamiennie.

z ważniejszych w dziedzinie ewaluacji Lalonde porównał szacunki wpływu szkolenia zawodowego uzyskane drogą eksperymentu z różnymi szacunkami efektu tego samego programu, które obliczono metodami quasi – eksperymentalnymi. Rezultat badań nie był zadowalający. Liczne szacunki, jakie uzyskano w wyniku zastosowania metod nieeksperymentalnych były bardzo różnicowane, tylko niektóre z nich pokrywały się z wynikami eksperymentu. Autor słusznie zauważył, że w sytuacji gdyby brakowało punktu odniesienia w postaci rezultatów eksperymentu, trudno byłoby ocenić, które szacunki wygenerowane przy użyciu metod nieeksperymentalnych są wiarygodne.

W ostatnich latach badania w zakresie metodologii oceny programów zatrudnieniowych rozwinęły się w innym co nieco kierunku. Ekonometrycy odeszli bowiem od poszukiwania „cudownego” estymatora (*magic bullet estimator*), przywiązując natomiast większe znaczenie do odpowiedniej jakości wykorzystywanych do badań danych i procesu ich gromadzenia.

Podobnie jak w eksperymencie, metody quasi – eksperymentalne opierają się na porównaniu położenia na rynku pracy uczestników programu z położeniem grupy kontrolnej. Nie występuje tutaj jednak dobór losowy i w związku z tym problem, jaki powstaje tyczy się prawidłowego skonstruowania grupy porównawczej, która ma być podstawą do oszacowania efektów, które wystąpiłyby, gdyby dany program nie był realizowany. Stosowane są bowiem różne rozwiązania, najpopularniejsze koncepcje zaprezentowane zostaną poniżej.

### **Koncepcja „przed versus po”**

W sytuacji, gdy środki przeznaczone na ewaluację są ograniczone, typowym podejściem, jakie się wówczas stosuje jest utworzenie grupy kontrolnej składającej się z osób uczestniczących w programie. W istocie więc badana jest jedna i ta sama grupa, porównuje się bowiem położenie uczestników, najczęściej wysokość osiąganych dochodów lub perspektywy zatrudnieniowe przed rozpoczęciem programu z odpowiednimi wielkościami osiąganymi po jego zakończeniu. Ujemną stroną tego rozwiązania jest założenie, że średni efekt jaki uzyskalaby populacja osób poddanych interwencji w sytuacji nie uczestniczenia w programie jest stały w czasie. Przyjmuje się bowiem, że w krótkim przedziale czasowym profil badanej grupy nie zmieni się i podobnie nie wystąpią istotne zmiany w gospodarce. Oznaczałoby to, że grupy byłyby niejako automatycznie dopasowane zarówno pod względem obserwowalnych, jak i nieobserwowalnych cech. Rzadko jednak dzieje się tak, że założenie to może zostać spełnione, wobec czego:

$$E[Y_{0it-1} | D_i = 1] \neq E[Y_{0it} | D_i = 1] \quad (7)$$

Ponadto należy w tym miejscu wspomnieć, że poważnym zagrożeniem dla słuszności wyników uzyskiwanych tą metodą jest występowanie zjawiska

określanego w literaturze jako Ashenfelter's Dip. Zostało ono po raz pierwszy zaobserwowane przez O. Ashenfeltera (1978) przy ewaluacji programów szkoleniowych i dotyczyło nieproporcjonalnego spadku dochodów w okresie poprzedzającym uczestnictwo w programie. Jak się później okazało, zjawisko to może wystąpić również w sytuacji gdy badane są efekty zatrudnieniowe, na co dowody w swoich pracach wskazują J. Heckman, R. Lalonde i J. Smith (1999). Perspektywa uczestnictwa w programie obniża aktywność bezrobotnych w zakresie poszukiwania pracy, wobec czego prawdopodobieństwo znalezienia zatrudnienia zmniejsza się i podobnie maleją zarobki. Wskutek tego efekty ewaluowanych programów mogą zostać przeszacowane.

### **Metody bazujące na dopasowaniu (*matching*) i regresji**

Ostatnio w licznych badaniach ewaluacyjnych coraz częściej do skonstruowania grupy kontrolnej wykorzystuje się metody bazujące na dopasowaniu. Korygowanie efektu selekcji w tym przypadku polega na wychwyceniu różnic w cechach obserwowalnych osób należących do obu badanych grup. Dopasowanie jak i regresja opierają się na założeniu warunkowej niezależności (*conditional independence assumption*), co można zapisać w następujący sposób:

$$E[Y_{0i} | D_i = 1, X] = E[Y_{0i} | D_i = 0, X] \quad (8)$$

W najprostszej wersji grupę kontrolną można utworzyć spośród osób, które w tym samym okresie co uczestniczący w aktywnych programach były bezrobotne, ale nie korzystały z wsparcia w formie udziału w tych programach. Profile obu badanych grup powinny być porównywalne. Przy zastosowaniu odpowiednich metod statystycznych grupy te mogą zostać dopasowane pod względem takich cech jak: wiek, płeć, poziom wykształcenia, poprzednio wykonywany zawód, typ zamieszkiwanego regionu, czy typ zamieszkiwanej miejscowości.

W bardziej wyrafinowanym podejściu grupę kontrolną można utworzyć stosując tzw. algorytm dopasowania (szerzej na ten temat zob. Fraker, Maynard 1987). Oprócz typowych cech społeczno-demograficznych uwzględnia się ponadto aspekty dotyczące na przykład wysokości osiąganych dochodów, czasu trwania bezrobocia, długości okresu pobierania zasiłku, udziału w szkoleniach i innych programach zatrudnieniowych. Ogólnie rzecz ujmując polega to na tym, że każdej osobie z grupy uczestniczących w aktywnych programach przyporządkowuje się osobę bardzo do niej podobną pod względem określonych obserwowalnych cech, czyli najlepiej do niej pasującą, która jednak nie bierze udziału w tych programach. Skonstruowane w ten sposób grupy określa się mianem „bliźniaków statystycznych”.

W sytuacji gdy grupy uznane są za homogeniczne, to w obu podejściach, jak również w koncepcji „przed versus po” odpowiednią techniką estymacji

efektów netto może być regresja parametryczna na współzmiennych. Jednak przy rosnącej popularności w badaniach ewaluacyjnych analizy nieparametrycznej modele regresji są ostatnio w literaturze krytykowane jako zbyt mało elastyczne i jednocześnie bazujące na założeniach trudnych do spełnienia (Speckesser 2004).

Poważnym zagrożeniem dla metod bazujących na dopasowaniu jest zjawisko określane w literaturze jako „przekleństwo wielowymiarowości”. Poszczególne osoby mogą różnić się między sobą w wielu wymiarach, co często powoduje, że przy ograniczonych danych prawidłowe dopasowanie jest po prostu niemożliwe. Rozwiązaniem tego problemu może być wykorzystanie metody opartej na ocenie skłonności (*propensity score*) rozwiniętej przez P. Rosenbauma i D. Rubina (1983). Zamiast przyporządkowania względem wektora obserwowalnych cech  $X$ , zaproponowali oni dopasowywanie poszczególnych osób w oparciu o ich prawdopodobieństwo udziału w programie  $P(X)$ . W ten sposób wiele wymiarów  $X$  zostaje zredukowanych do jednego wskaźnika, upraszczając tym samym procedurę dopasowania.

#### *Estymator „różnica w różnicach” (difference-in-differences)*

Czasami nie wystarczy korygować efekt selekcji uwzględniając tylko różnice w cechach obserwowalnych, gdyż jak się okazuje pozostałe istniejące rozbieżności w obu badanych grupach, które trudno jest zauważyć „gołym okiem” mogą obciążać szacunki efektów programu. Jedną z ciekawszych metod zapewniającą rozwiązanie tego problemu jest estymator definiowany jako „różnica w różnicach” (zob. np. Moffit 1991). Stanowi on rozszerzenie opisanej już wcześniej koncepcji „przed versus po”. Idea tej metody jest prosta. Dla obu badanych grup porównuje się różnicę w wynikach osiągniętych po programie z różnicą w wynikach uzyskanych przed programem.

Można to zapisać w następujący sposób:

$$\hat{\Delta} = [(\bar{Y}_{1t} - \bar{Y}_{0t-1}) | D = 1] - [(\bar{Y}_{0t} - \bar{Y}_{0t-1}) | D = 0] \quad (9)$$

W okresie  $t-1$  nikt nie jest poddany interwencji, natomiast w okresie  $t$  grupa eksperymentalna uczestniczy w programie. Efekt selekcji nie występuje wtedy, kiedy:

$$E[Y_{0it} | D_i = 1] - E[Y_{0it} | D_i = 0] = E[Y_{0it-1} | D_i = 1] - E[Y_{0it-1} | D_i = 0] \quad (10)$$

Metoda ta opiera się na założeniu, że efekt selekcji jest niezmienny w czasie. Wobec tego obie grupy nie mogą różnić się między sobą w okresie poprzedzającym udział w programie, co można zapisać w następujący sposób:

$$E[Y_{0it-1} | D_i = 1] = E[Y_{0it-1} | D_i = 0] \quad (11)$$

Jednak efekty uzyskiwane w okresie  $t-1$  przez osoby wchodzące w skład grupy eksperymentalnej mogą być obciążone przez czynniki związane z ich

przyszłym uczestnictwem w programie. Wobec tego postuluje się sprawdzanie adekwatności obu grup odpowiednio wcześniej przed rozpoczęciem programu.

Reasumując, stwierdza się, że metodologia dysponuje szerokim wachlarzem instrumentów do oceny programów zatrudnieniowych. Koniecznym zatem staje się doskonalenie ich na płaszczyźnie praktycznej. Wybór odpowiedniej metody zależy od istniejących uwarunkowań i przede wszystkim od jakości dostępnych danych, do których należy przywiązywać większe znaczenie.

## 5. EWALUACJA AKTYWNYCH PROGRAMÓW W POLSCE

Polska jest krajem, w którym brakuje skutecznego systemu ewaluacji polityki rynku pracy. Wprawdzie na początku lat 90. grupa konsultantów z Instytutu Badań Zatrudnienia (*W.E. Upjohn Institute for Employment Research*) w ramach projektu TOR 2 opracowała odpowiednie wskaźniki umożliwiające pomiar efektywności poszczególnych programów. Mierniki zostały skonstruowane przy uwzględnieniu przyjętej hierarchii celów aktywnych programów rynku pracy (Analiza efektywności programów rynku pracy – poradnik, 1995). Jednak nie wszystkie urzędy pracy wcieliły zaprojektowany system monitorowania w życie. Ocena aktywnych programów zatrudnieniowych w Polsce opiera się raczej na skromnym, wątpliwej jakości monitoringu. Z bogatego zestawu wskaźników publiczne służby zatrudnieniowe do badania efektywności wykorzystują praktycznie dwa, tj. stopę ponownego zatrudnienia i koszt ponownego zatrudnienia. Również w raportach o rynku pracy i zabezpieczeniu społecznym opracowanych przez Ministerstwo Gospodarki i Pracy badanie efektywności aktywnych form przeciwdziałania bezrobociu ogranicza się do analizy tych dwóch wskaźników (zob. Polska 2004. Raport o rynku pracy oraz zabezpieczeniu społecznym, <http://www.praca.gov.pl>) Jest to postępowanie płytkie, gdyż po pierwsze wskaźniki te są jedynie miarami brutto, po drugie nie uwzględniają jakościowych aspektów tego zatrudnienia i po trzecie porównywanie wielkości poszczególnych wskaźników w przekroju regionalnym bez użycia odpowiedniej metodologii jest obarczone błędem. Trzeba pamiętać, że na efektywność stosowanych instrumentów wpływa ponadto szereg czynników egzogenicznych niezależnych od programu i instytucji go wdrażającej, na przykład stan koniunktury, uwarunkowania lokalne. Aktywna polityka rynku pracy realizowana jest w obszarach, które zapewne zróżnicowane są pod względem struktury demograficznej, poziomu rozwoju gospodarczego, czy sytuacji na rynku pracy. Obok regionów atrakcyjnych, czyli takich, w których bezrobotni stosunkowo szybko znajdują prace są też i takie, w których szanse zatrudnieniowe są znikome.

Statystyki dotyczące finansowania polityki rynku pracy w Polsce nie są imponujące. W warunkach wysokiego bezrobocia w Polsce w ostatnich latach zwiększenie środków na aktywizację bezrobotnych jest znacznie utrudnione. Mimo, że uprawnieni do zasiłku stanowią 12,3% ogółu zarejestrowanych (stan na koniec III kwartału 2005 r.), wydatki na pasywne działania absorbują znaczną część Funduszu Pracy (Bezrobocie rejestrowane I-III kwartał 2005). Środki przeznaczane na aktywne formy są raczej skromne i w tej sytuacji ocena racjonalności gospodarowania tymi środkami jest tym bardziej zasadna. Niektóre problemy związane z finansowaniem aktywnych programów można częściowo rozwiązać poprzez efektywne wykorzystanie pieniędzy pochodzących jeszcze z funduszy przedakcesyjnych w ramach programu Phare – Spójność Społeczna i Gospodarcza, jak i środków Europejskiego Funduszu Społecznego, które uruchomione zostały po przystąpieniu Polski do Wspólnoty. Unia Europejska wypracowała już pewne standardy ewaluacyjne. Wykorzystywanie więc środków unijnych również wymusi konieczność ich ewaluacji.

Do tej pory badania ewaluacyjne aktywnych programów zatrudnieniowych w Polsce prowadzone były raczej sporadycznie przez zespoły naukowe w kraju, jak i za granicą. Opierały się one przede wszystkim na danych pochodzących z dodatkowego badania modułowego do Badania Aktywności Ekonomicznej Ludności, jakie GUS przeprowadził dwukrotnie, tj. w 1994 i 1996 roku. Oprócz analiz o charakterze quasi-eksperymentalnym podejmowane były także przedsięwzięcia badawcze typu jakościowego, które koncentrowały się wyłącznie na uczestnikach programów. Przeprowadzone badania wskazują między innymi na znaczne problemy związane z poprawnym adresowaniem programów, potwierdzają jednak pozytywne efekty zatrudnieniowe szkoleń zawodowych bezrobotnych. Sugerują również, że prace interwencyjne czy roboty publiczne powinny odgrywać w aktywizacji bezrobotnych coraz mniejsze znaczenie, z uwagi przede wszystkim na ich wątpliwe efekty (Puhani 1999a; Kluge, Lehman, Schmid 2000; Kubiak 2005).

Powracając do problemu stworzenia skutecznego systemu ewaluacji, należy rozstrzygnąć kwestię wyboru podmiotów odpowiedzialnych za ocenę aktywnych programów rynku pracy. W zależności od tego kto przeprowadza ocenę można mówić o ewaluacji wewnętrznej i zewnętrznej (Pierre 1999). O ile pełne monitorowanie działań aktywnych powinno niezaprzeczalnie realizowane być w systemie publicznych służb zatrudnienia o tyle zadania związane z oceną efektów końcowych czy efektów netto należałoby powierzyć całkiem niezależnym jednostkom. Wymagają one bowiem fachowej wiedzy i umiejętności w zakresie skomplikowanych metod i technik szacowania efektów, które opierają się przede wszystkim na zaawansowanej matematyce i ekonometrii. Ponadto w sytuacji prowadzenia ewaluacji programów przez instytucje jednocześnie odpowiedzialne za projektowanie i implementację tych

programów, zwiększa się ryzyko manipulowania wynikami. Istnieje bowiem presja polityczna nastawiona na osiąganie wysokich efektów, co stawia pod znakiem zapytania obiektywizm dokonywanych ocen.

Rządy wielu krajów OECD doszły jednak do wniosku, że badaniem efektywności programów zatrudnieniowych winny zajmować się podmioty zewnętrzne, w tym przede wszystkim instytucje naukowo-badawcze<sup>136</sup>, organizacje pozarządowe lub firmy konsultingowe. Gwarantuje to z pewnością większy obiektywizm jak i bardziej przychylny odbiór społeczny uzyskiwanych wyników.

Na koniec warto wskazać bariery, jakie stoją na drodze do wprowadzenia pełnego systemu monitorowania i ewaluacji aktywnej polityki rynku pracy w Polsce. Za najważniejsze uważa się:

- brak środków na prowadzenie systematycznej ewaluacji,
- niewłaściwe procedury i słabo rozwinięte systemy informacyjne w ramach urzędów pracy,
- niska jakość lub brak odpowiednich danych, koniecznych do przeprowadzenia pogłębionych analiz,
- niechęć pracowników publicznych służb zatrudnienia w stosunku do przedsięwzięć mających na celu ocenę efektywności wdrażanych przez nich instrumentów,
- niewystarczające kwalifikacje, słabe przygotowanie instytucji odpowiedzialnych za politykę rynku pracy do wypełniania funkcji kontrolnej i oceniającej.

Najpoważniejszym ograniczeniem jest brak odpowiednich jakościowo danych. Bogatym źródłem informacji mogłyby być zasoby danych administracyjnych powiatowych urzędów pracy. Należałoby jednak zmodyfikować pewne procedury i udoskonalić istniejące systemy informacyjne. Przeprowadzone dotychczas w Polsce badania ewaluacyjne pokazują, że w kraju istnieje zaplecze naukowo-badawcze, które byłoby w stanie realizować zadania związane z oceną efektów aktywnych programów rynku pracy. Tymczasem ewaluacją działań na rzecz aktywizacji bezrobotnych współfinansowanych ze środków unijnych zajmują się przeważnie firmy konsultingowe.

## 6. WNIOSKI

W latach 90. ubiegłego stulecia nastąpił znaczący postęp metodologiczny w zakresie badania wpływu aktywnych programów rynku pracy. Na gruncie europejskim na szeroką skalę rozpowszechniły się przede wszystkim metody

---

<sup>136</sup> Takie rozwiązanie zostało zastosowane między innymi w Republice Federalnej Niemiec i Wielkiej Brytanii.



quasi-eksperymentalne. Jak pokazuje praktyka, dostępne metody oceny, oprócz licznych walorów, mają również pewne ograniczenia, związane najczęściej z przyjęciem określonych założeń, które są trudne do spełnienia. Problemy te można wyeliminować kładąc większy nacisk na poprawę jakości wykorzystywanych w badaniach danych i proces ich gromadzenia.

Ewaluacja aktywnej polityki rynku pracy w Polsce nie jest realizowana w sposób systematyczny, a istniejący system monitorowania jest wykorzystywany w bardzo okrojonej formie. Utrzymujące się nadal wysokie bezrobocie, jak i napływ funduszy pomocowych Unii Europejskiej sprzyjają ekspansji aktywnych programów zatrudnieniowych, szczególnie szkoleń zawodowych. Rzetelna ewaluacja wybranych instrumentów prowadzi do bardziej racjonalnego gospodarowania środkami przeznaczonymi na formy aktywne i większej efektywności aktywnej polityki rynku pracy w zakresie ograniczenia i łagodzenia skutków bezrobocia. Dlatego w Polsce należy podjąć działania, które przyczynią się do udoskonalenia metod oceniających te programy i szerszego wykorzystania ich w praktyce.

#### SUMMARY

Active labour market policies are still one of the major tools to fight unemployment and improve functioning of the labour market. However, growing scepticism regarding the effects of active policies and very different evaluation findings are the main reason of growing interest in developing the evaluation methods of these policies. The paper outlines the key theoretical issues concerning ALMP evaluation and also intends to explain the need of evaluation of these measures, especially in countries which have just joined European Union, for example Poland.

The first part describes evaluation purposes and briefly characterizes differences between various types of evaluation. The second part covers the methodological aspects of ALMP microevaluations, defines the fundamental evaluation problem and reviews some quasi-experimental methods, commonly used in the EU – 15, like before-after designs, matching and parametric regression, and conditional difference in differences estimation, underlying their strengths and weaknesses. The last part of the article presents some problems of ALMP evaluation in Poland. The paper concludes that the main barriers of implementing evaluation system in Poland are: the lack of funds to conduct systematic evaluations, bad procedures and poorly developed information systems used by public employment services, low quality or the lack of proper statistical data needed for evaluations, and insufficient qualifications of PES' employees to control and evaluate the effects of ALMP.

#### BIBLIOGRAFIA

- Ashenfelter O. (1978), *Estimating the Effect of Training Program on Earning*, "The Review of Economics and Statistics", No. 60.
- Auer P., Kruppe T. (1996), *Monitoring of Labour Market Policy in EU Member States*, (w:) G. Schmid, J. O'Reilly, K. Schömann (red.), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar, Cheltenham.

- Bellmann L., Jackman R. (1996), *Aggregate Impact Analysis*, (w:) G. Schmid, J. O'Reilly, K. Schömann (red.), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar, Cheltenham.
- Delander L., Niklasson H. (1996), *Cost-benefit Analysis*, (w:) G. Schmid, J. O'Reilly, K. Schömann (red.), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar, Cheltenham.
- European Commission (2003), *Evaluating Socio Economic Development, Sourcebook 1: Active Labour Market Policies*, Przewodnik Komisji Europejskiej z zakresu ewaluacji, <http://www.evalseid.info>.
- Fraker T., Maynard R. (1987), *The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs*, "Journal of Human Resources", Vol.22, No.2.
- Fredriksson P., Johansson P. (2003), *Program Evaluation and Random Program Starts*, Department of Economics, Uppsala University "Working Paper", No. 1.
- Friedlander D., Robins P.K. (1995), *Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods*, "The American Economic Review", Vol. 85, No. 4.
- GUS, Bezrobocie rejestrowane I-III kwartał 2005 r., Warszawa 2005.
- Heckman J.J., Robb R. (1985), *Alternative Methods for Evaluating the Impact of Interventions*, (w:) J.J. Heckman, B. Singer (red.), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press, Cambridge.
- Heckman J.J., Smith J.A. (1996), *Experimental and Nonexperimental Evaluation*, (w:) G. Schmid, J. O'Reilly, K. Schömann (red.), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar, Cheltenham.
- Heckman J.J., Lalonde R.J., Smith J.A. (1999), *The Economics and Econometrics of Active Labor Market Programs*, (w:) O. Ashenfelter, D. Card (red.), *Handbook of Labor Economics*, North Holland, Amsterdam.
- Kluve J., Lehmann H., Schmidt Ch. M. (2000), *Disentangling Treatment Effects of Polish Active Labor Market Policies: Evidence from Matched Samples*, Preliminary Draft.
- Koning J. de, Mosley H. (2002), *How can active policies be made more effective?*, (w:) Schmid G., Gazier B.(red.), *The Dynamics of Full Employment. Social Integration Through Transitional Labour Markets*, Edward Elgar, Cheltenham.
- Kubiak P. (2005), *Efekty uczestnictwa bezrobotnych w aktywnych programach rynku pracy w Polsce*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Kwiatkowski E. (1998), *Aktywna polityka państwa na rynku pracy w Polsce*, (w:) E. Kryńska, E. Kwiatkowski, H. Zarychta, *Polityka państwa na rynku pracy w Polsce w latach dziewięćdziesiątych*, IPISS, Warszawa.
- Kwiatkowski E., Tokarski T. (1997), *Efekty polityki państwa wobec rynku pracy w Polsce. Analiza na podstawie funkcji dopasowań*, "Ekonomista", nr 3.
- Lalonde R.J. (1986), *Evaluating the Econometric Evaluations of Training Programs with Experimental Data*, "American Economic Review", No.76.

- Lechner M. (2001), *Identification and estimation of causal effects of multiple treatments under the conditional independence assumption*, (w:) M. Lechner, F. Pfeiffer (red.), *Econometric Evaluation of Labour Market Policies*, "ZEW Economic Studies", No. 13, Physica-Verlag, Heidelberg.
- Manski C.F. (1989), *Anatomy of the Selection Problem*, "Journal of Human Resources", No. 24.
- Martin J.P., Grubb D. (2001), *What works and for whom: A review of OECD countries' experiences with active labour market policies*, "Swedish Economic Policy Review", No. 8.
- Moffit R. (1991), *Program Evaluation with Nonexperimental Data*, "Evaluation Review", No. 15.
- Ministerstwo Gospodarki i Pracy, *Polska 2004. Raport o rynku pracy oraz zabezpieczeniu społecznym*, <http://www.praca.gov.pl> (04. 2005).
- OECD (1995), *Employment Outlook*, Paris.
- O'Leary Ch., Nesporova A., Samorodov A. (2001), *Manual on Evaluation of Labour Market Policies in Transition Economies*, ILO, Geneva.
- Orr L.L., (1999), *Social Experiments. Evaluating Public Programs with Experimental Method*, Sage Publications, London.
- Pierre G. (1999), *A framework for active laour market policy evaluation*, ILO Geneva, „Employment and Training Papers”, No. 49.
- Puhani P. (1999a), *Evaluating Acitve Labour Market Policies. Empirical Evidence for Poland During Transition*, "ZEW Economic Studies", No. 5, Physica-Verlag, Heidelberg.
- Puhani P. (1999b), *Estimating the Effects of Public Training on Polish Unemployment by Way of the Augmented Matching Function Approach*, ZEW Mannheim, „Discussion Paper”, No. 38.
- Purdon S. (2002), *Estimating the Impact of Labour Market Programmes*, Department for Work and Pensions, London.
- Purdon S., Lessof C., Woodfield K., Bryson C. (2001), *Research Methods For Policy Evaluation*, National Centre for Social Research, "Research Working Paper ", No. 2.
- Rosenbaum P.R., Rubin D. B. (1983), *The central role of the propensity score in observational studies for causal effects*, "Biometrika", No. 70.
- Rossi P.H., Lipsey M.W., Freeman H.E. (2003), *Evaluation: A Systematic Approach*, Sage Publications, Thousand Oaks.
- Rubin D.B (1974), *Estimating Causal Effects of Treatments In Randomized and Nonrandomized Studies*, "Journal of Educational Psychology", Vol.66, No.5.
- Rubin D.B. (1980), Comment on Basu, D.: *Randomization Analysis of Experimental Data: The Fisher Randomization Test*, "Journal of the American Statistical Association", No. 75.
- Schmid G., O'Reilly J., Schömann K. (1996), *Theory and Methodology of Labour Market Policy and Evaluation: An Introduction* (w:) G. Schmid, J. O'Reilly,

K. Schömann (red.), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar, Cheltenham 1996.

Smith J.A. (2000), *A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies*, "Zeitschrift für Volkswirtschaft und Statistik", Vol. 136(3).

Speckesser S. (2004), *Essays on Evaluation of Active Labour Market Policy*, Doctoral dissertation, Mannheim University, Mannheim.

World Bank (2002), *Impact Evaluation. Techniques for Evaluating Active Labor Market Programs*, "Employment Policy Primer", No.2.

WUP w Poznaniu (1995), *Analiza efektywności programów rynku pracy – poradnik*, Poznań.