

Szacowanie przyczynowego wpływu interwencji: funkcje, walory, ograniczenia oraz relacje z innymi podejściami w ewaluacji

prof. dr hab. Jarosław Górniak
Uniwersytet Jagielloński

III Międzyregionalna Konferencja Ewaluacyjna

Nowe wyzwania dla ewaluacji programów operacyjnych w perspektywie 2014-2020

Toruń, 21-22 czerwca 2016

Dwa kluczowe podejścia do ewaluacji

- Ewaluacja oparta o teorię programu (w tym ewaluacja realistyczna)
 - Program może być traktowany jako hipoteza, że określonych warunkach wdrożenie określonego zestawu przedsięwzięć w oparciu o założone zasoby przyniesie zaplanowane efekty
 - Celem ewaluacji jest obserwacja i wyjaśnienie zmiany w kategoriach założeń (często nieujawnianych w programie) dotyczących wykorzystania właściwych zasobów w właściwie dobranych działaniach zmierzających do założonych efektów
 - Odtworzenie *teorii programu* – jego *modelu logicznego*, a w ewaluacji realistycznej: mechanizmu generującego efekty w danych warunkach – jest podstawą trafności ewaluacji i użyteczności także dla innych programów
- Ewaluacja efektów
 - Jak najlepszy pomiar efektów: trafny, rzetelny, obiektywny
 - Pomiar to zestaw obserwacji, które **zmniejszają** niepewność i które można wyrazić jako liczbę
 - Ustalenie efektu netto, czyli wielkości zmiany wywołanej przyczynowo przez program: różnica między uzyskanym wynikiem, a oszacowanym dla sytuacji, gdyby nie podjęto działania – przykład: edukacyjna wartość dodana
- Obecnie główny nurt teorii ewaluacji opowiada się za syntezą tych podejść – ale jak taka synteza powinna wyglądać?

Ewaluacja oparta o teorię programu

- Istotą podejścia opartego o teorię programu było odtworzenie założonego przez decydentów modelu zależności zaplanowanych efektów od dobranych instrumentów działania.
- Rekonstrukcja teorii programu i teorii wdrożenia pozwala na zrozumienie przebiegu interwencji poprzez odczytanie motywów decyzji.
- Ewaluacja podąża za procesem wdrożenia interwencji dostarczając lepszego zrozumienia uwarunkowań powodzenia i niepowodzenia w realizacji programu oraz określenia użyteczności uzyskanych wyników.
- TBE daje dobry wgląd w złożoność warunków realizacji i sukcesu bądź porażki programu oraz w powiązania pomiędzy czynnikami składającymi się na te warunki.
- Problemem TBE jest słabe oddzielenie efektów programu od efektów oddziaływania innych czynników w otoczeniu i stąd zakłócenie obrazu wielkości efektu, który można przypisać interwencji oraz najczęściej jakościowy charakter wyników.

Trzy problemy, przed którymi staje ewaluacja

James Heckmann

- Ewaluacji wpływu (przyczynowego) przeszłych interwencji na obserwowane efekty, w tym w ich wpływ na dobrobyt osób, na które interwencja jest ukierunkowana.
- Przewidywanie wpływu interwencji już wdrożonych w jednym środowisku w innych środowiskach, z uwzględnieniem dobrobytu.
- Przewidywanie efektów interwencji nigdy nie wprowadzanych w nowych warunkach środowiskowych, z uwzględnieniem skutków dla dobrobytu

(Heckmann 2005, s. 8)

Poza eksperyment

- „Złotym standardem” badania przyczynowego jest eksperyment randomizowany
- W ewaluacji ograniczone są możliwości opierania się na eksperymentach
- Szacowanie efektu netto w oparciu o metody dające zbliżony efekt do:
 - Planu randomizowanego – *propensity score matching*
 - Randomizowanego planu blokowego – *corsend exact matching*
- Wyniki eksperymentu i ekwiwalentnych planów kontrfaktycznych mają ograniczone walory poznawcze:
 - Otrzymujemy oszacowanie efektu
 - Nie mamy wglądu w mechanizm społeczny w ramach którego program działa

Problemy z randomizowanym planem eksperymentalnym

Ograniczenia planu randomizowanego

- Może wystąpić tzw. „obciążenie randomizacyjne” oszacowań (Heckman i Smith, 1995): mierzony jest raczej efekt zamiaru uczestnictwa niż efekt uczestnictwa w programie (uczestnicy mogą odpaść, a nie-uczestnicy mogą znaleźć substytuty)
- Także badania eksperymentalne mogą prowadzić do błędnych wniosków
- Badania skoncentrowane jedynie na ustaleniu wielkości efektu netto mają ograniczone walory dla procesu uczenia się decydentów publicznych
- Konieczne jest zrozumienie wpływu przyczynowego programu w szerszym kontekście jego działania.

prosty eksperyment

X – bodziec (randomizacja)

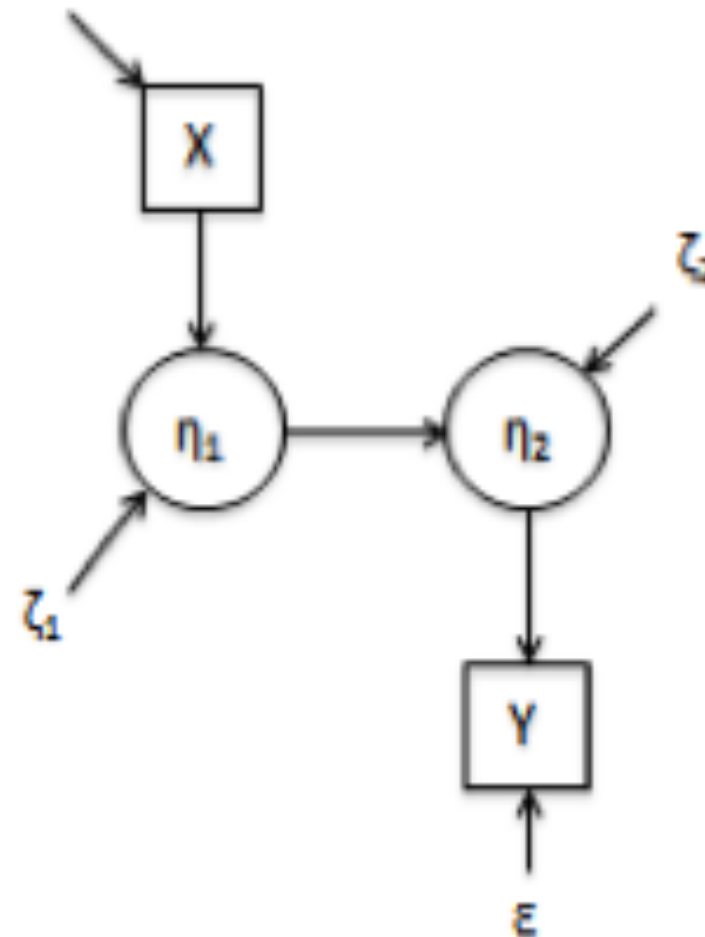
Y – wynik pomiaru efektu

η_1, η_2 – zmienne latentne (np. potrzeba domknięcia i skłonność do uprzedzeń)

Problemy:

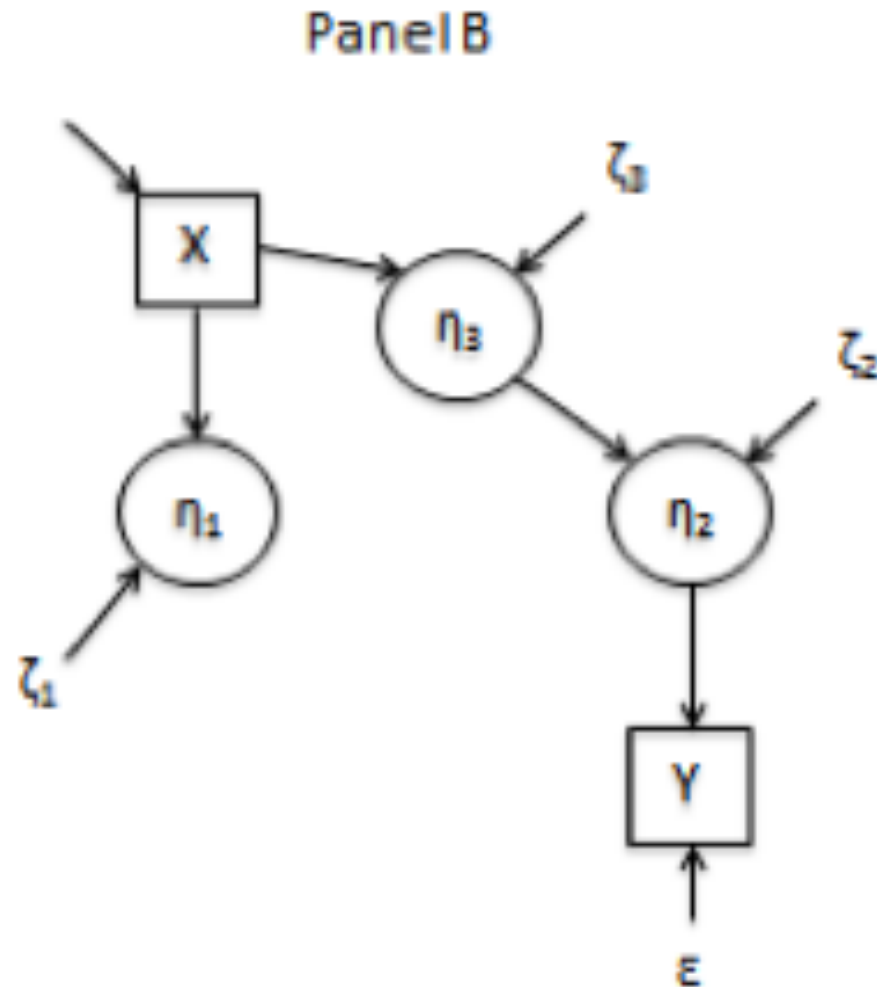
- Jak silnie bodziec działa na η_1 ?
- Jak dobry jest pomiar efektu?
- Jeśli słabo – wynik może być nieistotny nawet gdy wpływ jest silny

Panel A



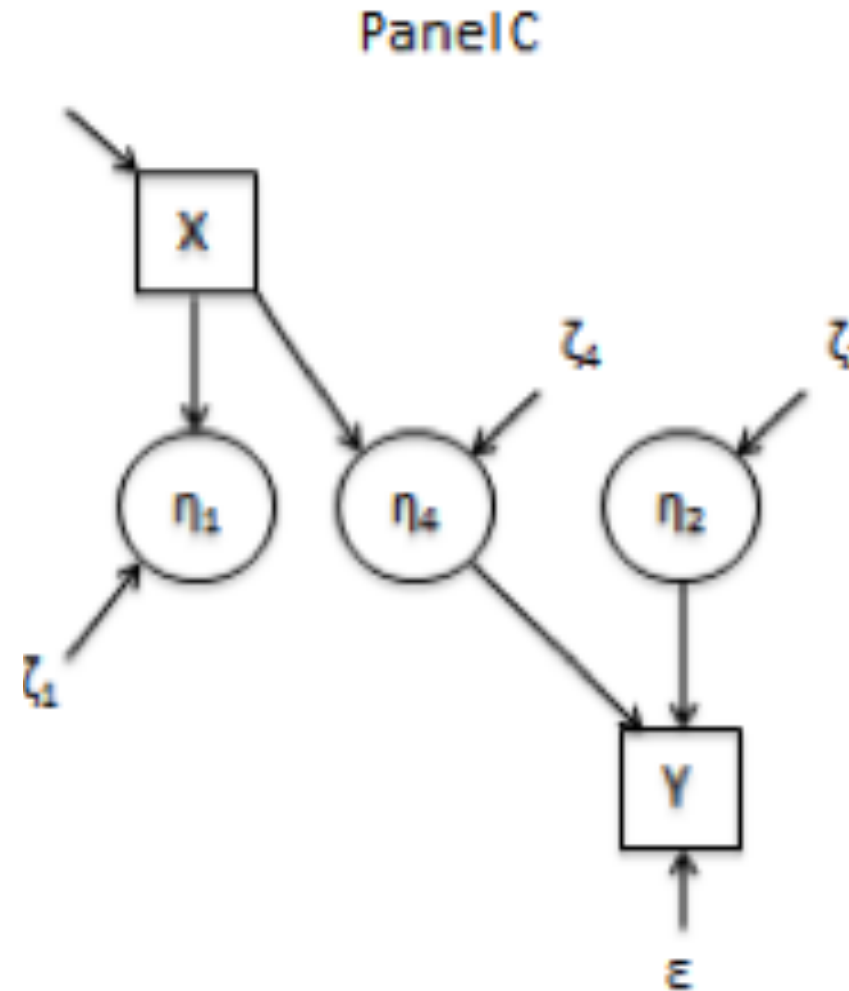
Inny mechanizm przyczynowy?

- X działa na inną zmienną, która wpływa na latentną zmienną zależną
- X i Y pozostają w istotnej statystycznie zależności



Inny mechanizm 2

- Zależność między X a Y jest istotna
- η_1 nie wpływa na η_2
- Bodziec wpływa na inną zmienną - η_4 , która wpływa na η_2



W kierunku syntezy

Wielkość efektu a rozliczalność wydatków publicznych

- Misją ewaluacji jest wsparcie rozliczalności/odpowiedzialności (*accountability*) instytucji publicznych wobec obywateli.
- Presja polityczna na dowody racjonalności w zakresie wydatków publicznych i oczekiwanie koncentracji programów i ewaluacji na efektach (Barca i in. 2011)
- Rola dowodów z ewaluacji w procesie uzasadnienia decyzji w politykach publicznych: czy dowody te mają znaczenie?
- Wyniki ewaluacji jako wkład do decyzyjnego „kosza na śmieci” (Cohen, March, Olsen): poza teorię programu

White'a theory based impact evaluation (2009)

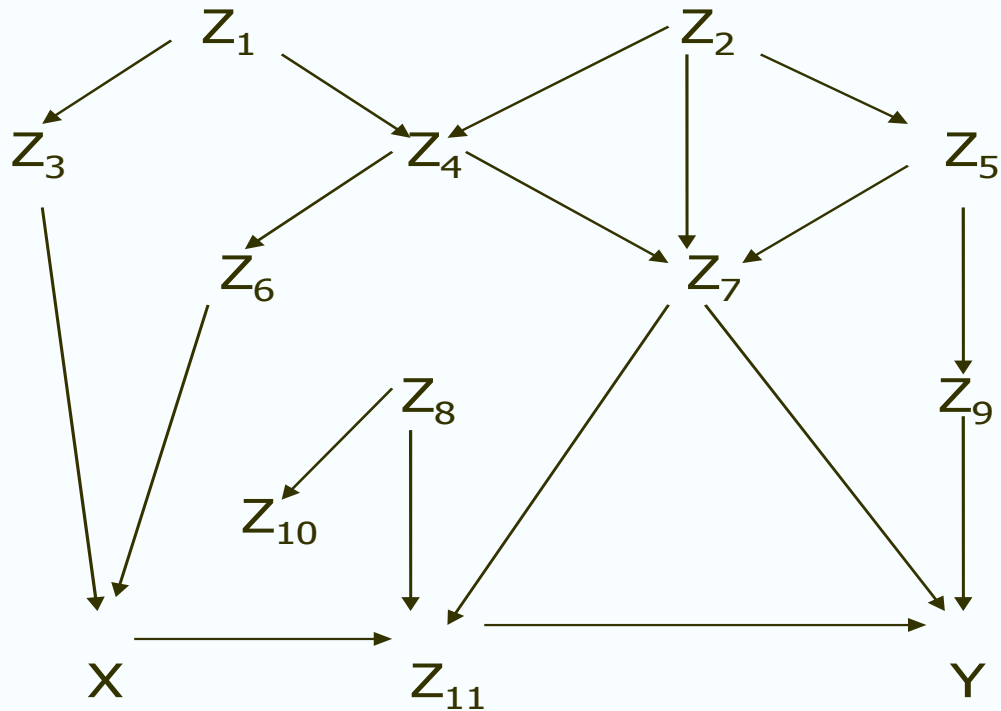
- 1. Nakreśl (*map out*) łańcuch przyczynowy (teorię programu)
- 2. Zrozum kontekst
- 3. Antycypuj heterogeniczność
- 4. Rygorystyczna ewaluacja wpływu z użyciem wiarygodnego podejścia kontrfaktycznego
- 5. Rygorystyczna analiza stanu faktycznego
- 6. Używaj metod mieszanych

Jak powiązać podejście oparte o teorię programu z badaniem wpływu przyczynowego interwencji

- Wynikiem fazy TBE powinien być model mechanizmu przyczynowego w ramach którego interwencja prowadzi do oczekiwanej zmiany.
- Należy wykorzystać język Acyklicznych Grafów Ukierunkowanych (ADG) do opisu modelu tego mechanizmu z koncentracją na interwencji jako bodźcu (*exposure*) i założonej zmianie jako efekcie.
- Należy ustalić minimalny zestaw cech, których kontrola jest niezbędna dla wyeliminowania zakłócenia obrazu zależności między interwencją a efektem przez tzw. „tylne drzwi”
- Na tej podstawie należy zaplanować proces pomiaru ilościowego zoperacjonalizowanych zmiennych chodzących w skład minimalnego zestawu
- Wreszcie, należy dokonać oszacowania efektu przyczynowego w warunkach kontroli potencjalnych zmiennych zakłócających.

Skierowany graf acykliczny (Pearl 2000)

Co muszę zmierzyć, by wiedzieć, jak X wpływa na Y?



Jak zbudować DAG

1. Ustalić co jest interwencją, a co wynikiem
2. Przeprowadzić główną ścieżkę/i przyczynowego oddziaływania interwencji na efekt uwzględniając wszystkie potencjalne mediatory.
3. Należy dodać do modelu wszystkie zmienne, które mogą przyczynowo wpływać na wynik interwencji.
4. Wskazać oddziaływania przyczynowe pomiędzy wszystkimi zmiennymi wpływającymi na wynik interwencji a interwencją.
5. Ustalić, czy występują jeszcze jakiegokolwiek zależności przyczynowe pomiędzy którymkolwiek zmiennymi uwzględnionymi w modelu i uwzględnić reprezentujące je strzałki w modelu.
6. Ustalić dla każdej pary zmiennych, czy istnieją jakiegokolwiek zmienne, które wpływają przyczynowo na parę.
 - Ta lista kroków została sporządzona na podstawie znakomitego wykładu Scotta Vennersa umieszczonego w youtube.com: <https://www.youtube.com/watch?v=pJhU4fimHBQ>

Proponowane kroki w ewaluacji

1. Najpierw działamy w ramach podejścia ewaluacji opartej na teorii programu, najlepiej w bardziej zaawansowanym wydaniu, zbliżonym do ewaluacji realistycznej.
2. W oparciu o ustalenia etapu 1. konstruujemy ukierunkowany graf acykliczny (DAG) lub panel SWIG reprezentujący ścieżkę działania interwencji na efekt i ustalamy minimalny zestaw zmiennych podlegających kontroli, by można było bez zakłócenia przez tzw. „tylne drzwi” oszacować zależność przyczynową .
3. Projektujemy plan gromadzenia danych zgodnie z wynikami analizy DAG z uwzględnieniem wymogów pomiaru, m.in. ilustrowanych wyżej.
4. Realizujemy starannie badanie.
5. Szacujemy parametry modelu przyczynowego.
6. Podsumowujemy wyniki ewaluacji opisując mechanizm działający u podstaw programu oraz przedstawiając wielkość efektu wywołanego przez interwencję.