

# Ewaluacja w polityce społecznej

## Kryterium skuteczności – badanie skutków

Dr hab. Ryszard Szarfenberg  
(były) Instytut Polityki Społecznej UW

[rszarf.ips.uw.edu.pl/ewalps/dzienne/](https://rszarf.ips.uw.edu.pl/ewalps/dzienne/)

Rok akademicki 2021/2022

# Drugie zadanie analityczne

- 1. Czy w analizowanej ewaluacji zastosowano kryterium skuteczności?**
- 2. Jak zdefiniowano kryterium skuteczności w analizowanej ewaluacji?**  
Czy definicja tego kryterium odpowiada temu, jak skuteczność i jej badanie zostały przedstawiona na wykładach?
- 3. W jaki sposób badano skutki projektu/programu?** Skąd pochodziły dane do identyfikowania skutków? Który ze schematów badania skutków zastosowano? Jak mierzono skalę skutków projektu/programu? Czy przeprowadzono analizę skutków ubocznych?
- 4. Jak oceniano program/projekt pod względem jego skutków zamierzonych i ubocznych?**

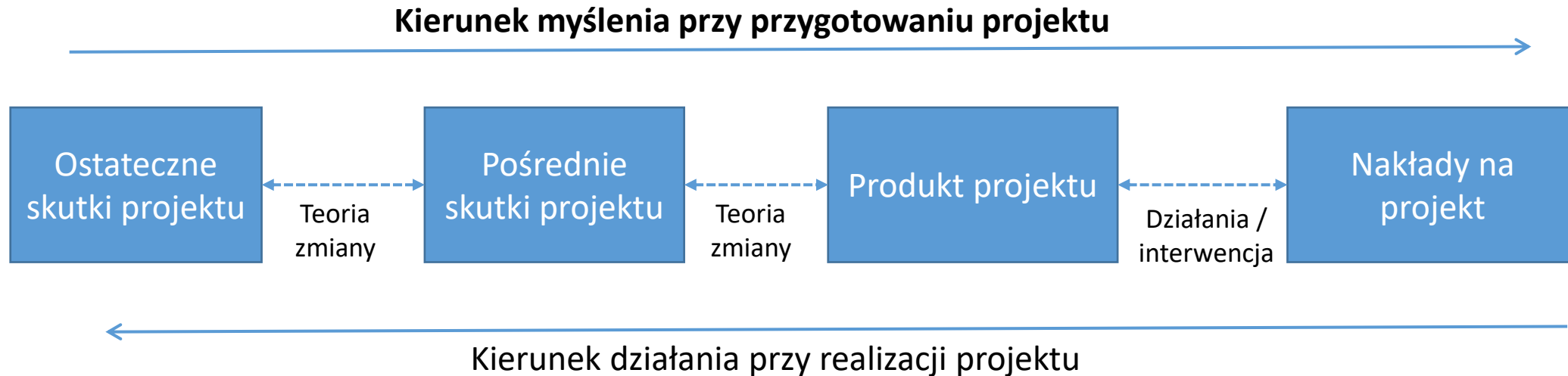
# Plan wykładu

1. **Badanie występowania i rozmiarów skutków** polityki społecznej a ocena ewaluacyjna zbadanych skutków w kontekście **teorii przyczynowo-skutkowej łączącej produkt projektu ze skutkami pośrednimi i ostatecznymi**
2. **Identyfikacja skutków** polityk, programów, projektów społecznych na przykładzie **metody eksperymentalnej, która służy do porównania sytuacji faktycznej z kontrfaktyczną**
3. Schematy badań eksperymentalnych, quasi-eksperymentalnych i nieeksperymentalnych

# Pytania dotyczące skutków i pytania dotyczące oceny skutków

- **Przykład pytań z zakresu badania skutków zamierzonych i ubocznych**
  1. Czy Program 500 Plus spowodował wzrost diety (oficjalny cel), a jeżeli tak, o ile wzrosły wskaźniki diety w porównaniu do sytuacji kontrfaktycznej? – **Pytanie o skutek zamierzony w oficjalnym celu**
  2. Czy Program 500 plus spowodował spadek ubóstwa rodzin z dziećmi (oficjalny cel), a jeżeli tak, o ile zmniejszyły się wskaźniki ubóstwa tych rodzin w porównaniu do sytuacji kontrfaktycznej? – **Pytanie o skutek zamierzony w oficjalnym celu**
  3. Czy Program 500 plus spowodował spadek zatrudnienia matek wychowujących dzieci otrzymujących świadczenie, a jeżeli tak, o ile zmniejszyły się wskaźniki zatrudnienia tej grupy? – **Pytanie o skutek uboczny**
- **Przykład pytań o ocenę skutków** (o ile mamy dowody, że to są rzeczywiście skutki ewaluowanego programu w określonym rozmiarze)
  - Jeżeli Program 500 Plus spowodował określone skutki (np. wzrost diety o x%, spadek ubóstwa o y%, zmniejszenie zatrudnienia o z%), to jaka jest ich ocena, np. czy są zamierzone i pozytywne, czy uboczne negatywne; czy są małe, średnie, duże?
    - **Pytanie o ocenę wartości skutków - pozytywne czy negatywne?**
    - **Pytanie o ocenę skali skutków - czy są duże, małe, średnie?**

# Badanie skutków w ujęciu teorii przyczynowo-skutkowej projektu



M. Koopman i in red., *Performance Measurement in the Dutch Social Rented Sector*, 2008, s. 20. Modyfikacje R.S.

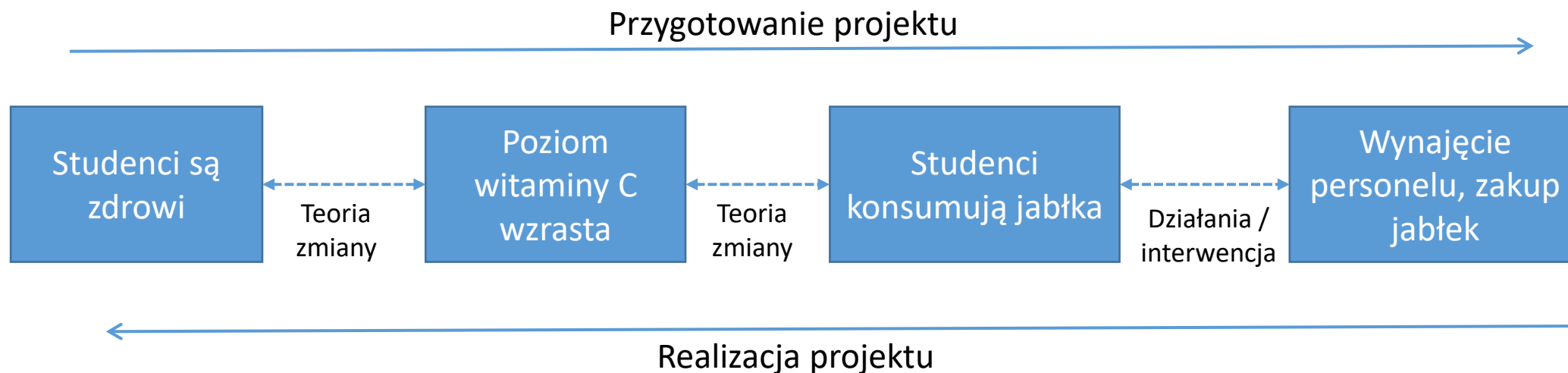
Metoda identyfikacji skutków powinna **dostarczyć dowodów na to, że produkt projektu spowodował skutki pośrednie, a te spowodowały skutki ostateczne w zamierzonych kierunkach i rozmiarach**. Pytania:

1. Czy produkt spowodował zamierzone skutki pośrednie, a jeżeli tak, jaka była ich skala?
2. Czy skutki pośrednie spowodowały skutki ostateczne, a jeżeli tak, jaka była ich skala?

# Zastosowanie teorii przyczynowo-skutkowej do interpretacji wyników ewaluacji

| Czy implementacja interwencji była odpowiednia? | Czy wystarczające były zaangażowanie, rekrutacja, wypełnianie zaleceń przez odbiorców? | Czy skutki pośrednie zostały osiągnięte? | Czy skutki ostateczne zostały osiągnięte? | Interpretacja                                                                     |
|-------------------------------------------------|----------------------------------------------------------------------------------------|------------------------------------------|-------------------------------------------|-----------------------------------------------------------------------------------|
| X                                               | X                                                                                      | X                                        | X                                         | Niepowodzenie implementacji                                                       |
| ✓                                               | X                                                                                      | X                                        | X                                         | Niepowodzenie zaangażowania lub zgodności (pierwszy łącznik przyczynowo-skutkowy) |
| ✓                                               | ✓                                                                                      | X                                        | X                                         | Niepowodzenie teorii (wcześniejszy łącznik przyczynowo-skutkowy)                  |
| ✓                                               | ✓                                                                                      | ✓                                        | X                                         | Niepowodzenie teorii (późniejszy łącznik przyczynowo-skutkowy)                    |
| ✓                                               | ✓                                                                                      | ✓                                        | ✓                                         | <b>Zgodność z teorią</b>                                                          |
| ✓                                               | ✓                                                                                      | X                                        | ✓                                         | Niepowodzenie teorii (inna ścieżka przyczynowo-skutkowa)                          |

## Przykład teorii konkretnego projektu – Jabłka dla studentów



M. Koopman i in red., *Performance Measurement in the Dutch Social Rented Sector*, 2008, s. 20. Modyfikacje i adaptacja R.S.

Metoda identyfikacji skutków powinna **dostarczyć dowodów na to, że produkt projektu (jabłka) spowodował skutki pośrednie (zjedzone jabłka znacząco zwiększyły poziom witaminy C), a te spowodowały skutki ostateczne w określonych rozmiarach (zwiększenie poziomu witaminy C znacząco zmniejszyło poziom zachorowalności studentów)**. Pytania:

1. Czy zjedzenie jabłek spowodowało wzrost poziomu witaminy C, a jeżeli tak, jaka była skala tego wzrostu?
2. Czy wzrost poziomu witaminy C spowodował zmniejszenie zachorowalności studentów, a jeżeli tak, jaka była skala tego zmniejszania?

# Zastosowanie teorii projektu do interpretacji wyników ewaluacji – projekt Jabłka dla studentów

| Czy jabłka zostały dostarczone? | Czy jabłka zostały zjedzone? | Czy poziom witaminy C we krwi wzrósł? | Czy poprawił się poziom zdrowia? | Interpretacja                                                                     |
|---------------------------------|------------------------------|---------------------------------------|----------------------------------|-----------------------------------------------------------------------------------|
| X                               | X                            | X                                     | X                                | Niepowodzenie implementacji                                                       |
| ✓                               | X                            | X                                     | X                                | Niepowodzenie w zaangażowaniu i zgodności zachowań (pierwszy łącznik przyczynowy) |
| ✓                               | ✓                            | X                                     | X                                | Niepowodzenie teorii (wcześniejszy łącznik przyczynowy)                           |
| ✓                               | ✓                            | ✓                                     | X                                | Niepowodzenie teorii (późniejszy łącznik przyczynowy)                             |
| ✓                               | ✓                            | ✓                                     | ✓                                | <b>Zgodność z teorią</b>                                                          |
| ✓                               | ✓                            | ✓/X                                   | ✓/X                              | Częściowe niepowodzenie teorii (działa tylko w pewnych kontekstach)               |
| ✓                               | ✓                            | X                                     | ✓                                | Niepowodzenie teorii (odmienna ścieżka przyczynowa)                               |

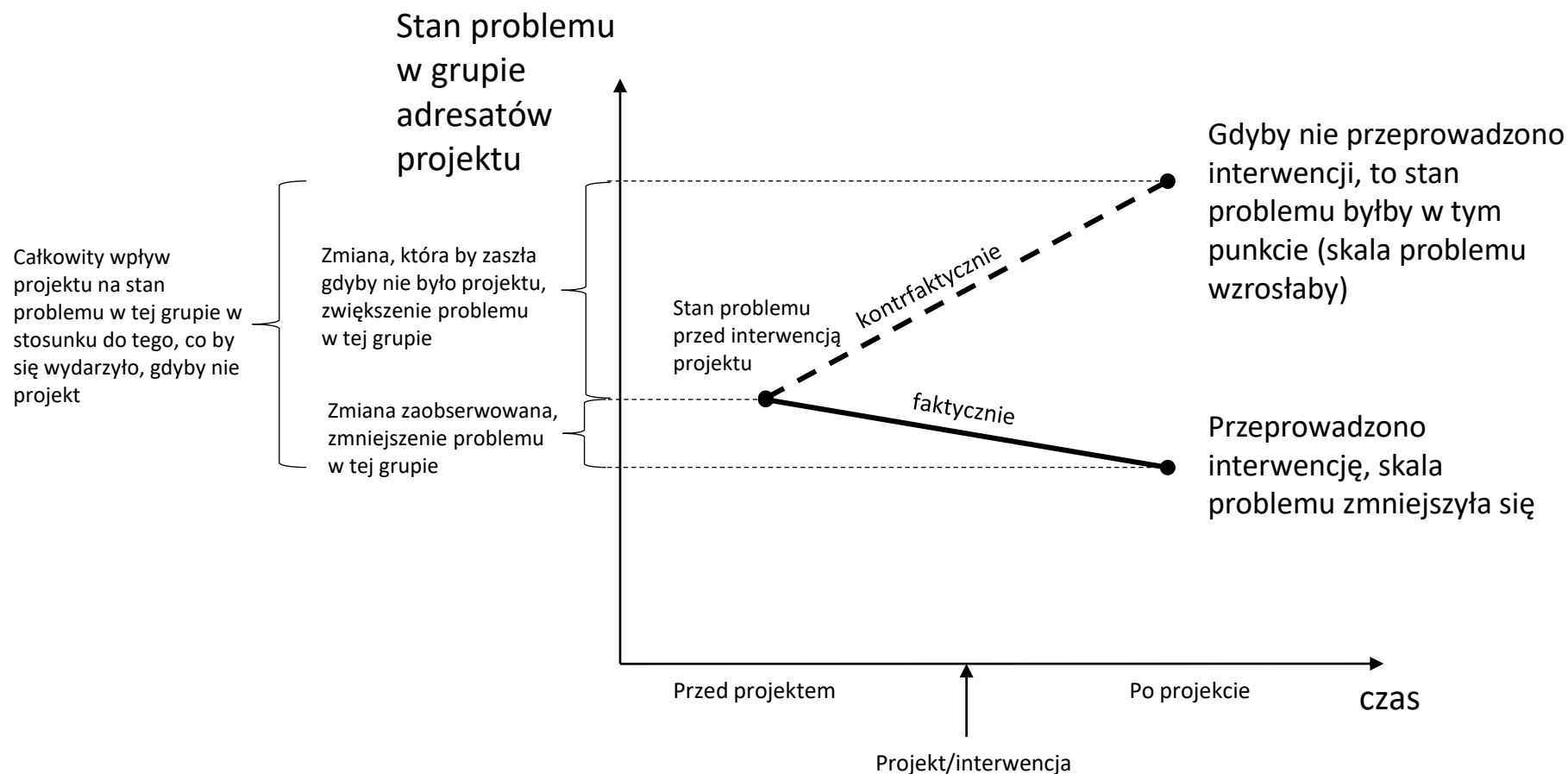
Podejście eksperymentalne do badania skutków  
polityki, programów, projektów społecznych

Nagrodę Nobla z ekonomii w 2019 r. otrzymali Abhijit Banerjee, Esther Duflo oraz Michael Kremer. Za co? Za upowszechnianie **metod eksperymentalnych** w badaniach programów walki z ubóstwem

"Badania przeprowadzone przez tegorocznych laureatów **znacznie poprawiły nasze umiejętności dotyczące walki z ubóstwem na świecie.** W ciągu zaledwie dwóch dziesięcioleci ich **nowe, eksperymentalne działanie zmieniło podejście** do ekonomii rozwoju, która jest obecnie kwitnącą dziedziną badań"

Do czego służą przede wszystkim metody eksperymentalne? Do dostarczania dowodów na to, że interwencje polityki społecznej regulujące zachowania czy dostarczające usług, świadczeń czy rzeczy mają określone skutki

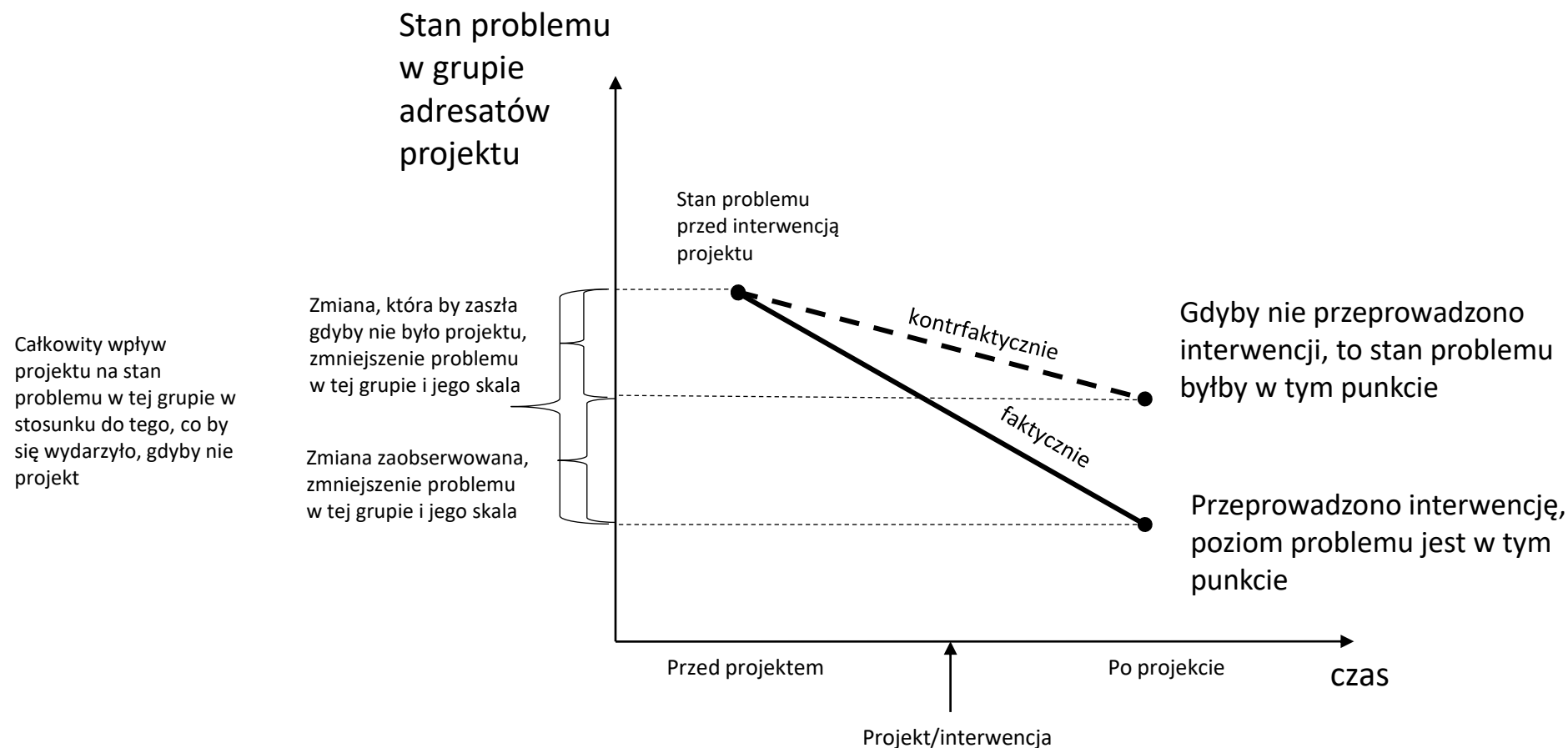
# Przyczynowość kontrfaktyczna przykład 1 – problem by wzrósł, gdyby nie interwencja



## Podstawowe zadania

1. Pomiar faktycznego stanu problemu w grupie odbiorców przed i po ich uczestnictwie w projekcie
2. Pomiar kontrfaktycznego stanu problemu w grupie odbiorców, gdyby nie uczestniczyli w projekcie
3. Porównanie faktycznego i kontrfaktycznego stanu problemu i stwierdzenie, czy występuje różnica i jej skala

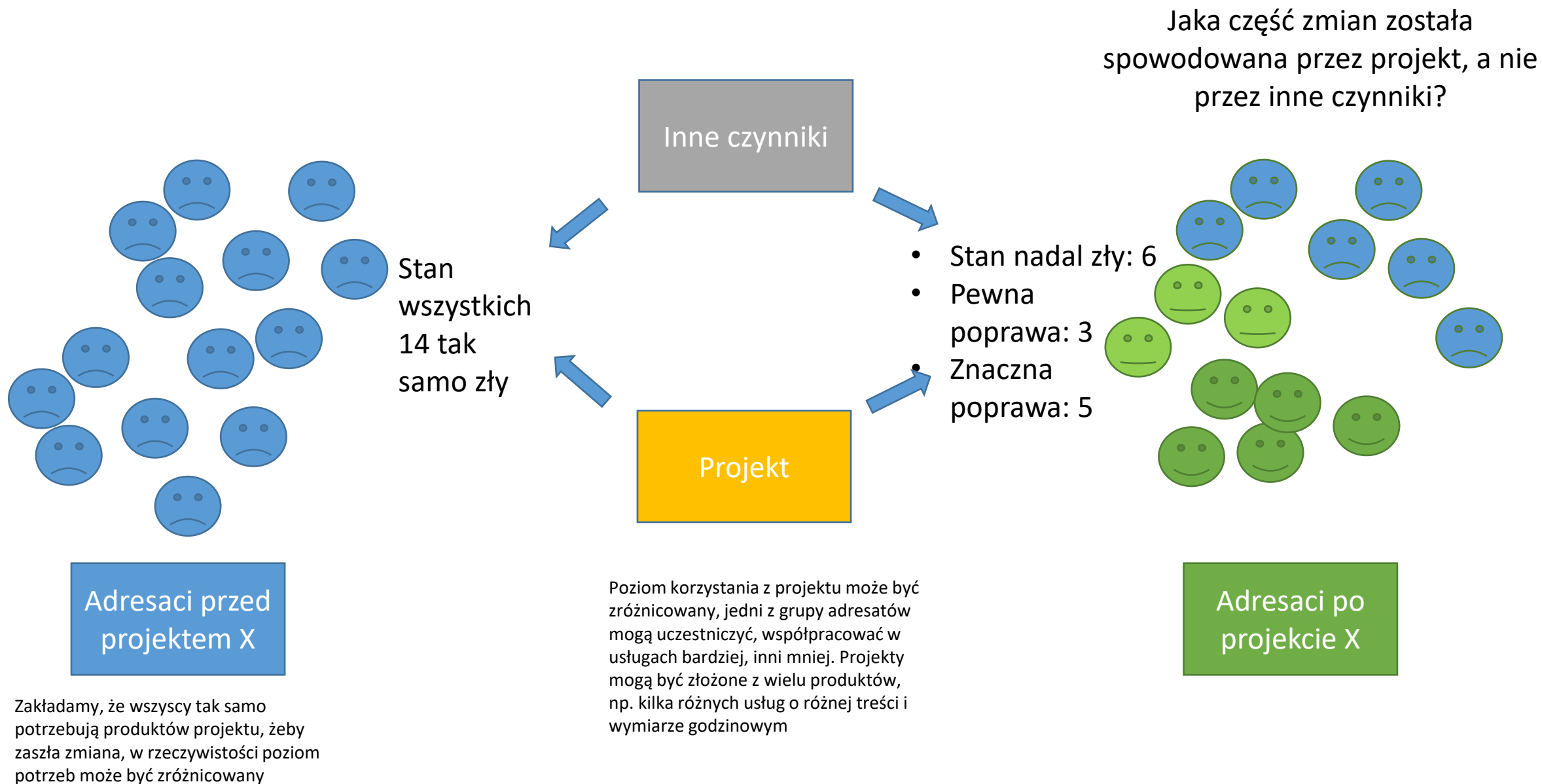
## Przyczynowość kontrfaktyczna przykład 2 – problem by spadł mniej, gdyby nie interwencja



### Podstawowe zadania

1. Pomiar faktycznego stanu problemu w grupie odbiorców przed i po ich uczestnictwie w projekcie
2. Pomiar kontrfaktycznego stanu problemu w grupie odbiorców, gdyby nie uczestniczyli w projekcie
3. Porównanie faktycznego i kontrfaktycznego stanu problemu i stwierdzenie, czy występuje różnica i jej skala

# Adresaci przed i po projekcie uczestniczący w projekcie



Eksperyment: dobieramy grupę kontrolną (porównawczą), która jest możliwie jak najbardziej zbliżona do adresatów projektu, ale nie zostanie nim objęta. Pytanie: co by było gdyby ci sami adresaci nie brali udziału w projekcie?



Zakładamy, że grupa kontrolna jest możliwie jak najbardziej zbliżona do grupy adresatów projektu (w ideale powinny być to te same osoby w takim samym świecie, w którym jednak nie zrealizowano tego projektu)

Porównanie zmiany w grupie adresatów projektu do grupy kontrolnej daje nam możliwość określenia wpływu projektu bez wpływu innych czynników

# Analiza porównawcza zmian w grupie adresatów i w grupie kontrolnej

| Przypadek | Zmiana w grupie adresatów | Interpretacja            | Zmiana w grupie kontrolnej | Interpretacja      | Różnica między przypadkami |
|-----------|---------------------------|--------------------------|----------------------------|--------------------|----------------------------|
| 1         | 0,0                       | brak zmiany              | 0,0                        | brak zmiany        | brak różnicy               |
| 2         | 1,0                       | <b>poprawa</b>           | 0,0                        | <b>brak zmiany</b> | <b>różnica</b>             |
| 3         | 0,0                       | brak zmiany              | 0,0                        | brak zmiany        | brak różnicy               |
| 4         | 1,0                       | poprawa                  | 1,0                        | poprawa            | brak różnicy               |
| 5         | 1,0                       | <b>poprawa</b>           | 0,0                        | <b>brak zmiany</b> | <b>różnica</b>             |
| 6         | 0,0                       | brak zmiany              | 0,0                        | brak zmiany        | brak różnicy               |
| 7         | 0,0                       | brak zmiany              | 0,0                        | brak zmiany        | brak różnicy               |
| 8         | 0,0                       | brak zmiany              | 0,0                        | brak zmiany        | brak różnicy               |
| 9         | 1,0                       | poprawa                  | 1,0                        | poprawa            | brak różnicy               |
| 10        | 0,5                       | <b>częściowa poprawa</b> | 0,0                        | <b>brak zmiany</b> | <b>różnica</b>             |
| 11        | 1,0                       | poprawa                  | 1,0                        | poprawa            | brak różnicy               |
| 12        | 0,0                       | brak zmiany              | 0,0                        | brak zmiany        | brak różnicy               |
| 13        | 0,5                       | częściowa poprawa        | 0,5                        | częściowa poprawa  | brak różnicy               |
| 14        | 0,5                       | częściowa poprawa        | 0,5                        | częściowa poprawa  | brak różnicy               |

W tabeli założono, że w grupie adresatów i w grupie kontrolnej są te same osoby. Pytanie brzmi: **co by było gdyby te same osoby z grupy adresatów nie zostały poddane oddziaływaniu projektu** (stąd przymiotnik „kontrfaktyczny”, faktem jest poddanie grupy oddziaływaniu, a „kontrfaktem” jest przybliżona alternatywna rzeczywistość, w której nie zostały one poddane temu oddziaływaniu)

## Prosty sposób podsumowania różnic

Grupa adresatów (eksperymentalna)

| Liczba braku zmiany | Liczba popraw | Liczba popraw częściowych |
|---------------------|---------------|---------------------------|
| 6                   | 5             | 3                         |

Grupa kontrolna (porównawcza)

| Liczba braku zmiany | Liczba popraw | Liczba popraw częściowych |
|---------------------|---------------|---------------------------|
| 9                   | 3             | 2                         |

Różnica: liczba zmian w grupie kontrolnej odjąć liczba zmian w grupie adresatów

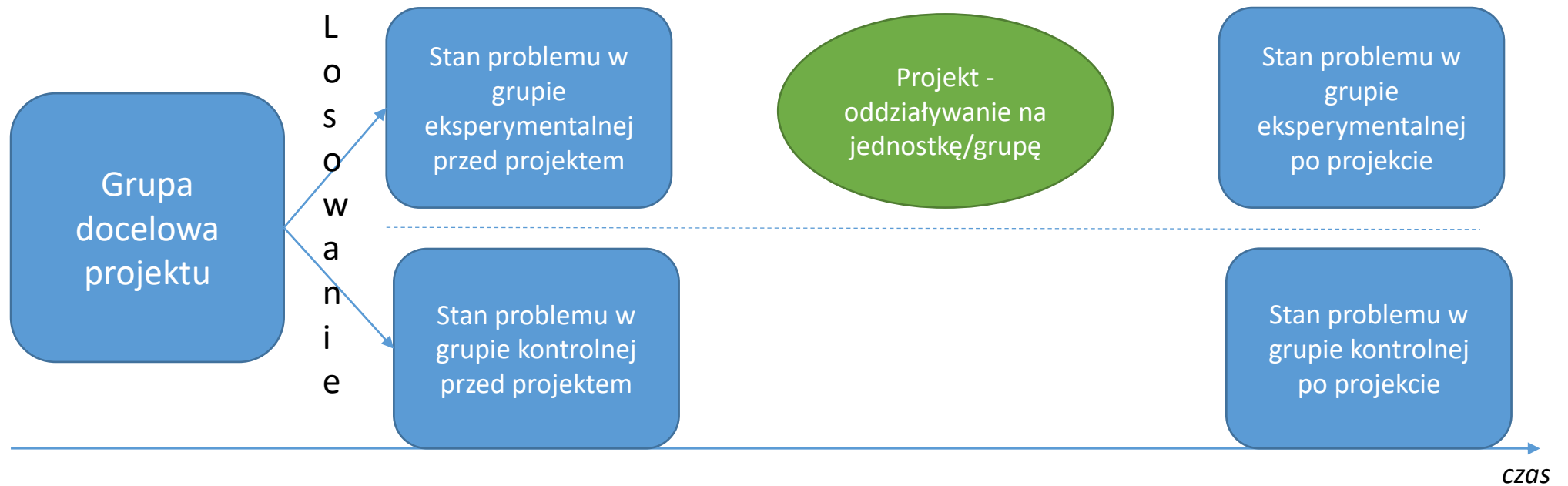
| Liczba braku zmiany | Liczba popraw | Liczba popraw częściowych |
|---------------------|---------------|---------------------------|
| 3                   | -2            | -1                        |

Schematy badań eksperymentalnych i quasi-eksperymentalnych

Schematy badań ilościowych, których celem jest pomiar skutków jakie skorzystanie z produktów projektu wywołało w odbiorcach

- **Eksperyment, czyli pretest – posttest z losowo dobraną grupą eksperymentalną i kontrolną (porównawczą)**
- **Quasi-eksperyment, czyli pretest – posttest, ale odbiegający od ideału eksperymentu**
  - Z grupą porównawczą o charakterze nielosowym
  - Szeregów czasowych z kilkoma pretestami i posttestami
- **Pretest – posttest, czyli pomiar tego, na co chcemy wpłynąć projektem przed jego realizacją i po jego zakończeniu**
- **Tylko posttest, czyli brak pretestu**

# Eksperyment losowy



Klasyczny opis schematu według trzech oznaczeń: R – randomizacja (dobór losowy), O - obserwacja zmiennej zależnej, X - zmienna niezależna (produkt projektu, interwencja)

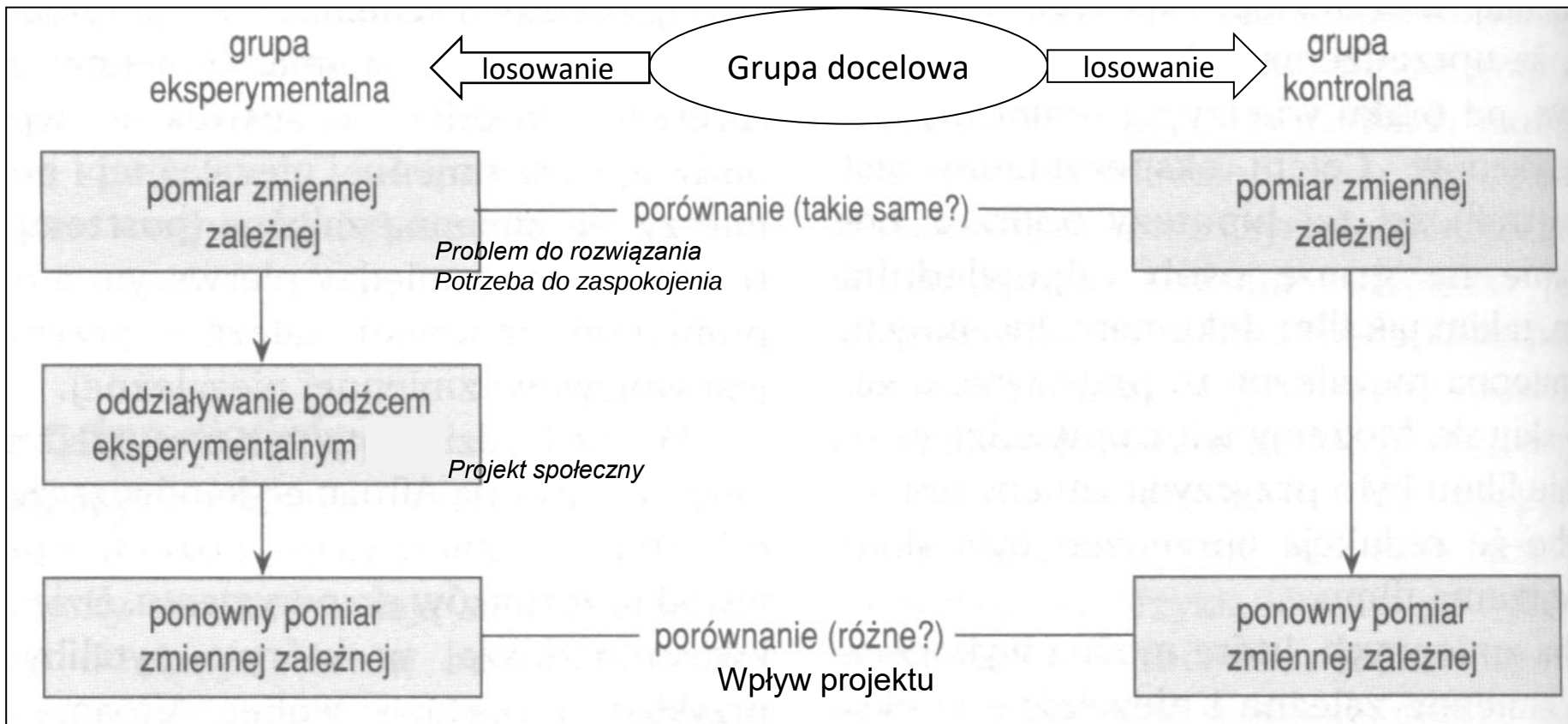
**R O X O**

**R O O**

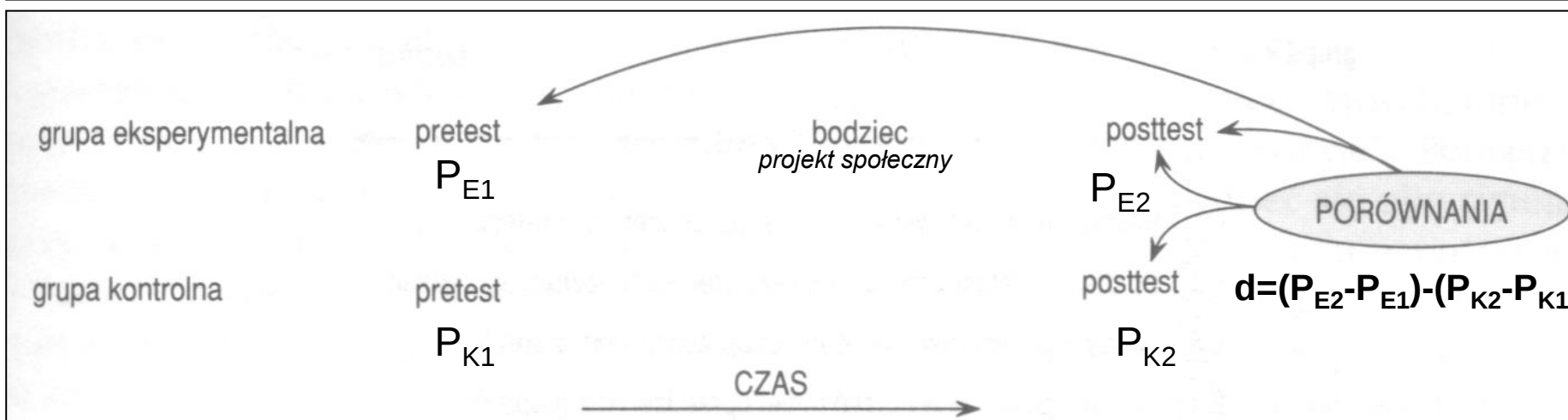
Pomiar stanu problemu w obu grupach, czyli **pretest**

Porównanie wyników pomiarów w celu wnioskowania o wpływie projektu na jego adresatów

Pomiar stanu problemu w obu grupach, czyli **posttest**

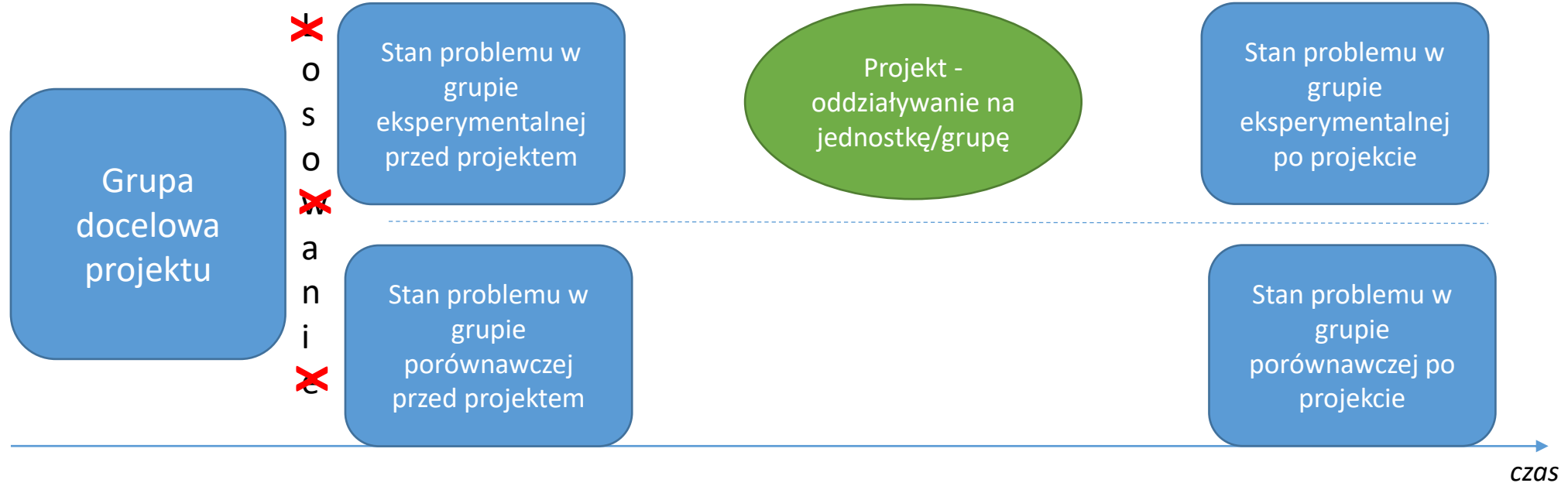


Schemat eksperymentalny jako podstawa do ustalenia skutków projektu



d – wpływ (skutek)  
Jest to istotnie statystyczna różnica między różnicami post- i pretestów dla grupy eksperymentalnej i dla grupy kontrolnej (**różnica w różnicach**)

# Quasi-eksperyment z grupą porównawczą



Pomiar stanu problemu w obu grupach, czyli **pretest**

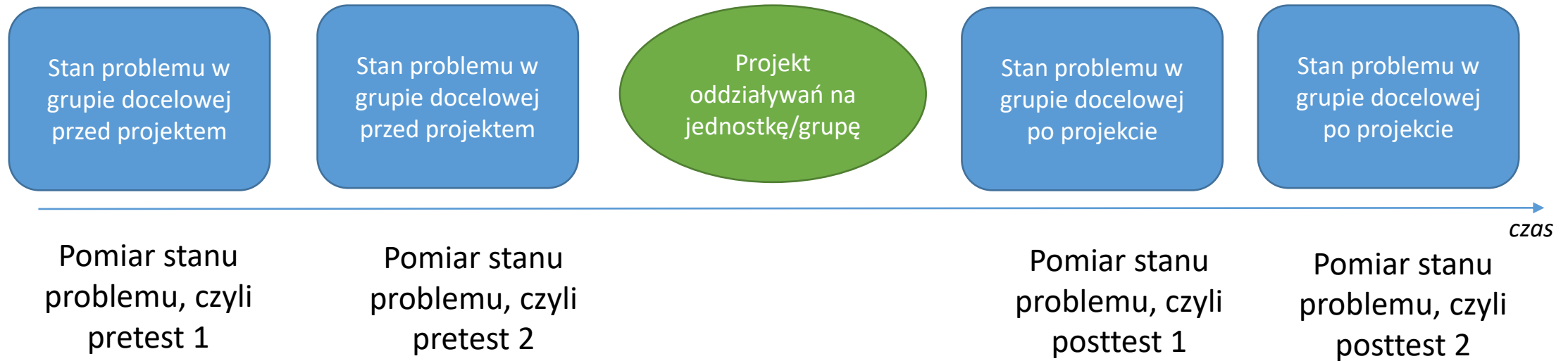
Porównanie wyników pomiarów w celu wnioskowania o wpływie projektu na jego adresatów

Pomiar stanu problemu w obu grupach, czyli **posttest**

W porównaniu ze schematem poprzednim nie przeprowadzono tu losowania. Jest grupa porównawcza, ale nie zapewniono, że rozkład adresatów pomiędzy grupy będzie losowy

OXO  
O O

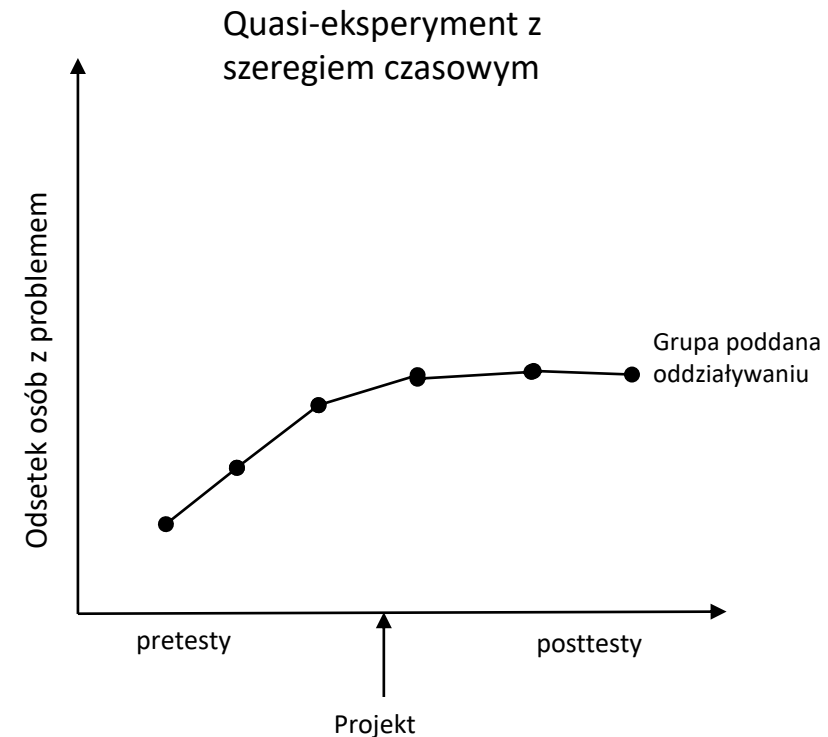
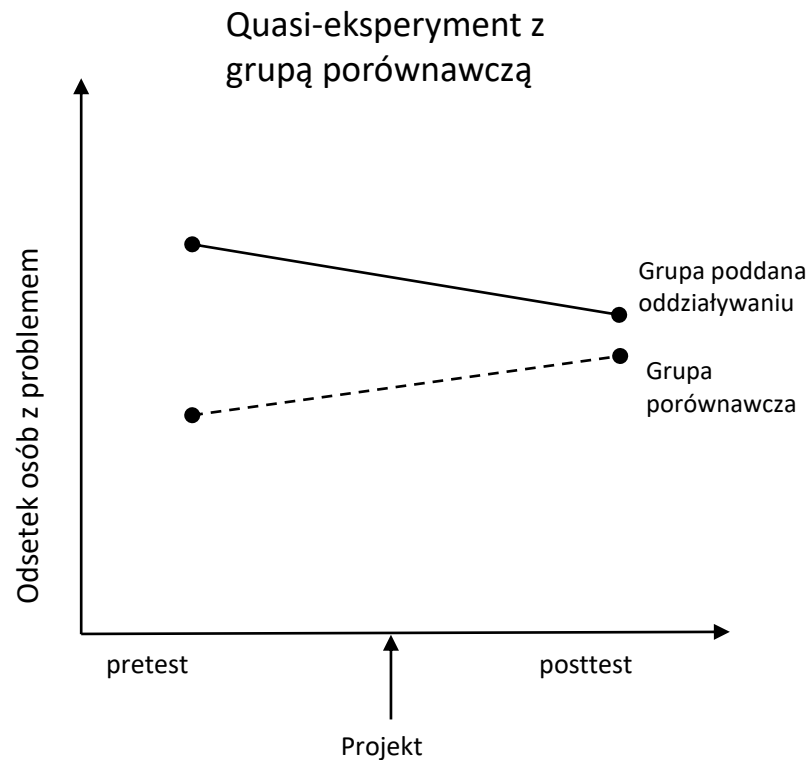
# Quasi-eksperyment z szeregami czasowymi



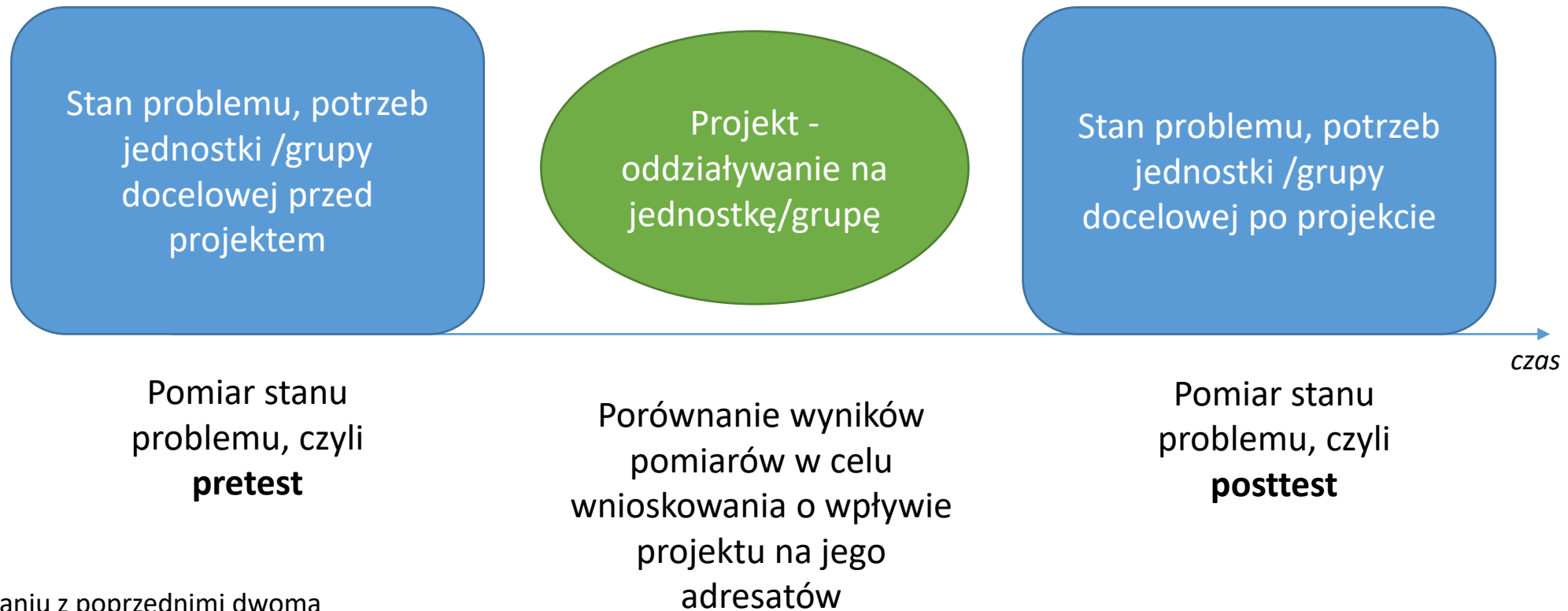
W porównaniu z poprzednimi dwoma slajdami nie przeprowadzono tu losowania, nie ma też grupy porównawczej, ale jest kilka obserwacji przed i po projekcie

00X00

# Wizualizacja możliwych wyników schematów quasi-eksperymentalnych



## Tylko jednorazowy pretest i posttest bez grupy porównawczej

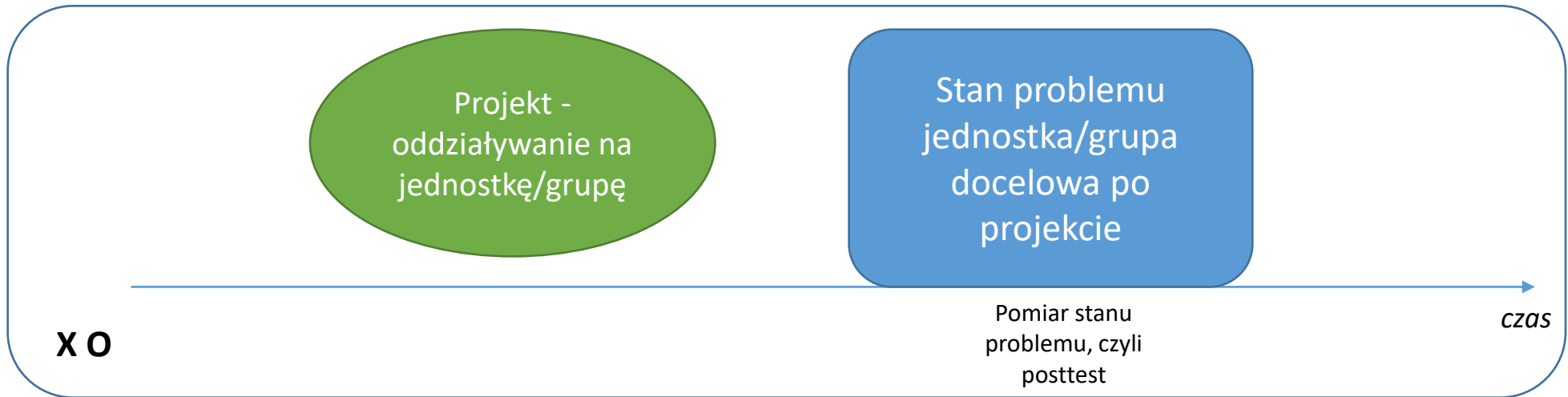


W porównaniu z poprzednimi dwoma slajdami nie przeprowadzono tu losowania, nie ma też grupy porównawczej, są tylko jednorazowe pretest i posttest

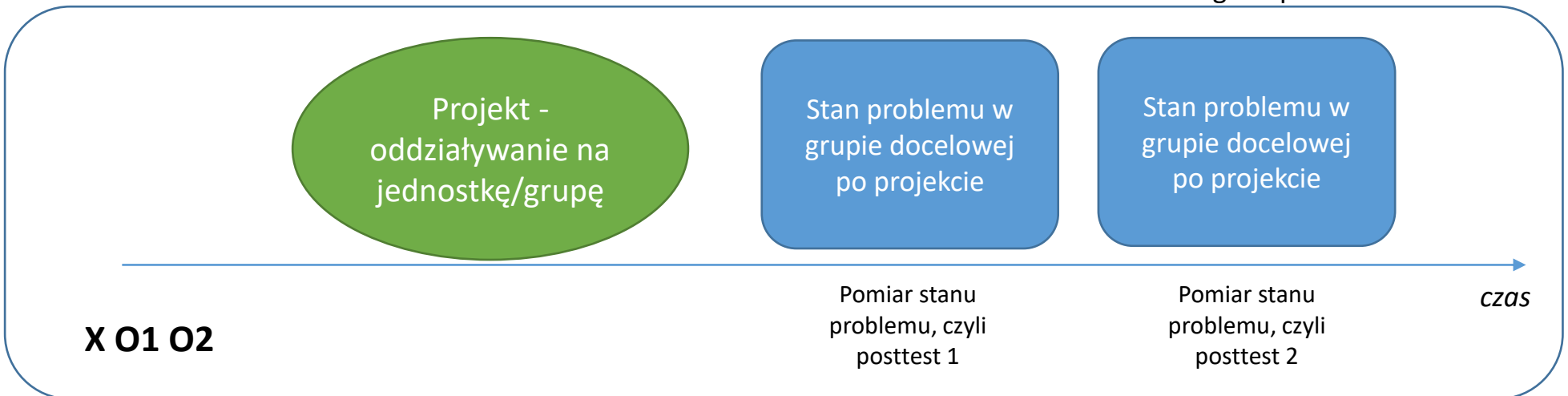
**O X O**

# Tylko posttest w dwóch wariantach

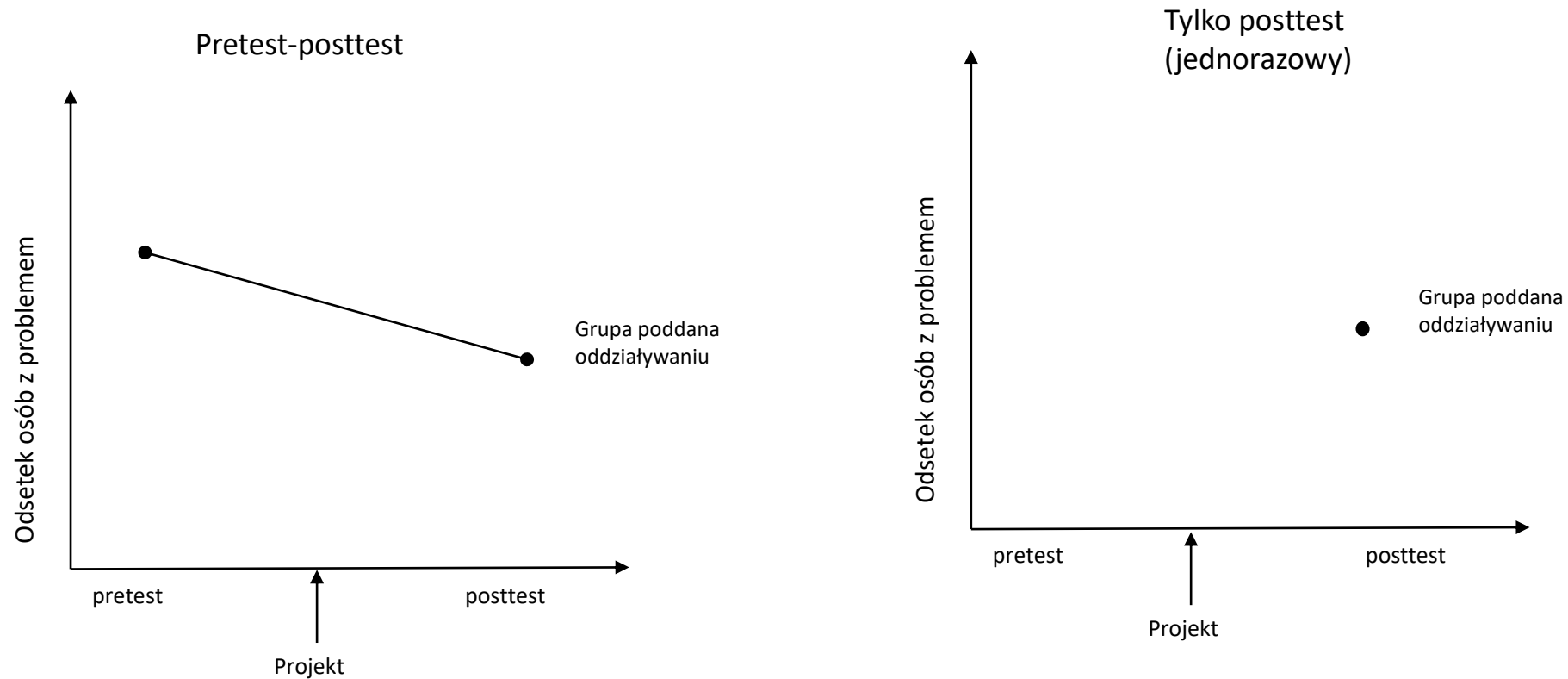
Wariant jednorazowego posttestu



Wariant z szeregiem posttestów



# Wizualizacja możliwych wyników schematów nieeksperymentalnych



[J. D. Hawkins, B. Nederhooft, Podręcznik ewaluacji programów profilaktycznych, 1994 \(oryginał 1987, tłumaczenie i wydanie sfinansowano z programu „Promocja badań ewaluacyjnych w profilaktyce uzależnień”\)](#)

# Uogólnianie wniosków z eksperymentów

- Eksperymenty w ewaluacji projektów społecznych przeprowadzane są po to, aby zidentyfikować i zmierzyć ich wpływ (skutek) na problemy/potrzeby adresatów
- Załóżmy, że mamy **eksperyment przeprowadzony z niewielkimi grupami eksperymentalną i kontrolną**, który zidentyfikował wpływ (skutek) projektu na adresatów i umożliwił pomiar wpływu, który okazał się znaczący
- *Czy można na podstawie tego eksperymentu wnioskować, że podobny projekt zastosowany do podobnej grupy w innej lokalizacji lub w całym kraju będzie miał również zbliżony wpływ?*
- Wątpliwość dotyczy **uogólnienia wyników eksperymentu na całą populację – aby móc uogólniać trzeba dysponować odpowiednio dużymi i reprezentatywnymi próbami** (przykład eksperymentu w Finlandii z bezwarunkowym zasiłkiem dla bezrobotnych w latach 2017-2018 – grupa eksperymentalna wynosiła 2000 bezrobotnych)
- Jeżeli nie ma możliwości przeprowadzenia eksperymentu z reprezentatywnymi grupami, to jest problem z ich wiarygodnością

# Czego się dowiedzieliśmy? Podsumowanie

1. Teoria polityki, programu czy projektu społecznego dokładnie pokazuje związki przyczynowo skutkowe między produktami, skutkami pośrednimi i skutkami ostatecznymi
2. Badanie skutków (wpływu) polityki, programu czy projektu polega na zidentyfikowaniu i zmierzeniu ich skutków (wpływu) zamierzonych i ubocznych na problemy/potrzeby ich adresatów oraz ich otoczenia
3. Ocena skutków polityki, programu, projektu stwierdzonych w badaniu wpływu polega co najmniej na ocenie ich wartości (pozytywne, negatywne) oraz ich rozmiaru (duże, średnie, małe)
4. Główną metodą badania skutków (wpływu) jest metoda eksperymentalna oparta na podejściu kontrfaktycznym: co by było, gdyby nie było interwencji w postaci polityki, programu czy projektu?
5. Metoda eksperymentalna polega na odpowiednim dobraniu uczestników do grupy poddanej oddziaływaniu polityki czy projektu oraz do grupy kontrolnej (muszą one być jak najbardziej podobne do siebie), a następnie na obliczeniu różnicy w różnicach między pretestami i posttestami dla obu grup
6. Najbardziej wiarygodne wyniki dają eksperymenty losowe na dużych reprezentatywnych grupach eksperymentalnej i kontrolnej, ale nie znaczy to, że nie ma innych metod badania skutków (wpływu) polityki społecznej

# Literatura i film z wykładem

- [Anna Kurowska, Schematy badań – metody eksperymentalne i quasi-eksperymentalne, w: B. Szatur-Jaworska, \*Ewaluacja w służbach społecznych\*, rozdział 12, 2010](#)
- [Film z wykładem](#) (2020/2021)