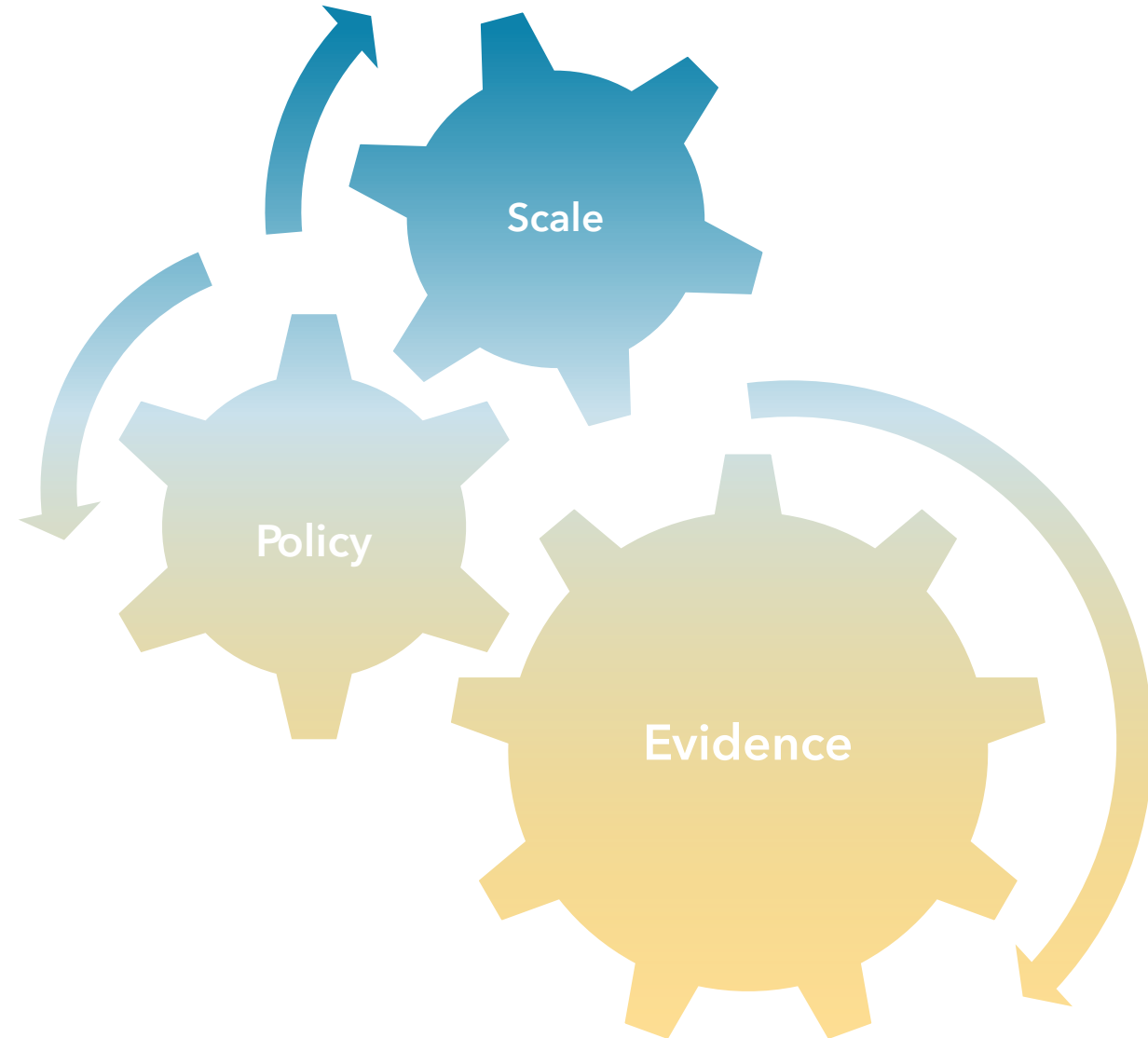


Impact Evaluation Collaborative

Moving Economic
Inclusion to Scale

IE WORKSHOP





Méthodes d'évaluation d'impact quasi- expérimentales

Clara Delavallade

Senior Economist - Banque Mondiale

Africa Gender Innovation Lab

INTRODUCTION

Pourquoi l'évaluation d'impact

Objectif principal:

- Guider les décisions politiques et de programmes sur base de preuves solides.

Argument:

- Plusieurs types d'évidence du plus simple & moins-couteux au plus complexe.
- En cas de doutes sur l'efficacité d'un programme, l'étude d'incidence (ou évaluation d'impact) se justifie.
- Si nous n'avons pas de doutes, nous pouvons nous passer de l'étude d'incidence.

INTRODUCTION

Pourquoi l'évaluation d'impact

- Permet de *mesurer les* impacts directs des politiques publiques, programmes d'infrastructures...
 - Est-ce que le programme a un impact sur l'indicateur choisi ?
 - Comprendre les avantages nets du programme & comparer les différentes options du programme
 - Comprendre la répartition des effets du programme entre différents sous-groupes
 - Comprendre comment les résultats peuvent changer en modifiant certains aspects d'un programme

Autres avantages opérationnels

- Décaissement plus rapide & moins de retards dans la mise en œuvre du projet

INTRODUCTION

Récapitulatif de la séance d'hier

- Evaluation d'impact Vs. évaluation des besoins Vs. évaluation du processus
- Inférence causale et contrefactuel
- Comment randomiser

INTRODUCTION

Evaluation d'impact

- Exemple: Les programmes d'emploi pour les jeunes peuvent-ils réduire le risque de participation à la violence?



INTRODUCTION

Evaluation d'impact

- Exemple: Les garderies peuvent-elles accroître la participation des femmes dans la vie économique?



INTRODUCTION

Evaluation d'impact

- Exemple: La modernisation des routes augmente-t-elle le bien-être de la population rurale?



INTRODUCTION

Inférence causale

- Qu'est-ce que l'impact d'un programme?
Comment le mesure-t-on?
 - *C'est la différence entre les résultats avec programme et sans programme*
- L'objectif de l'évaluation d'impact est de mesurer cette différence de manière à *l'attribuer au programme*, et *seulement au programme*
- Qu'est-ce qu'un contrefactuel?
 - *C'est ce qui arriverait en l'absence du programme ou projet*



INTRODUCTION

Contrefactuel

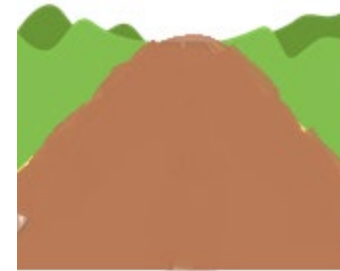
Trouver un groupe de comparaison, le contrefactuel:

Ce qui s'est vraiment passé



Monde réel !!!

Ce qui se serait passé sans programme - contrefactuel

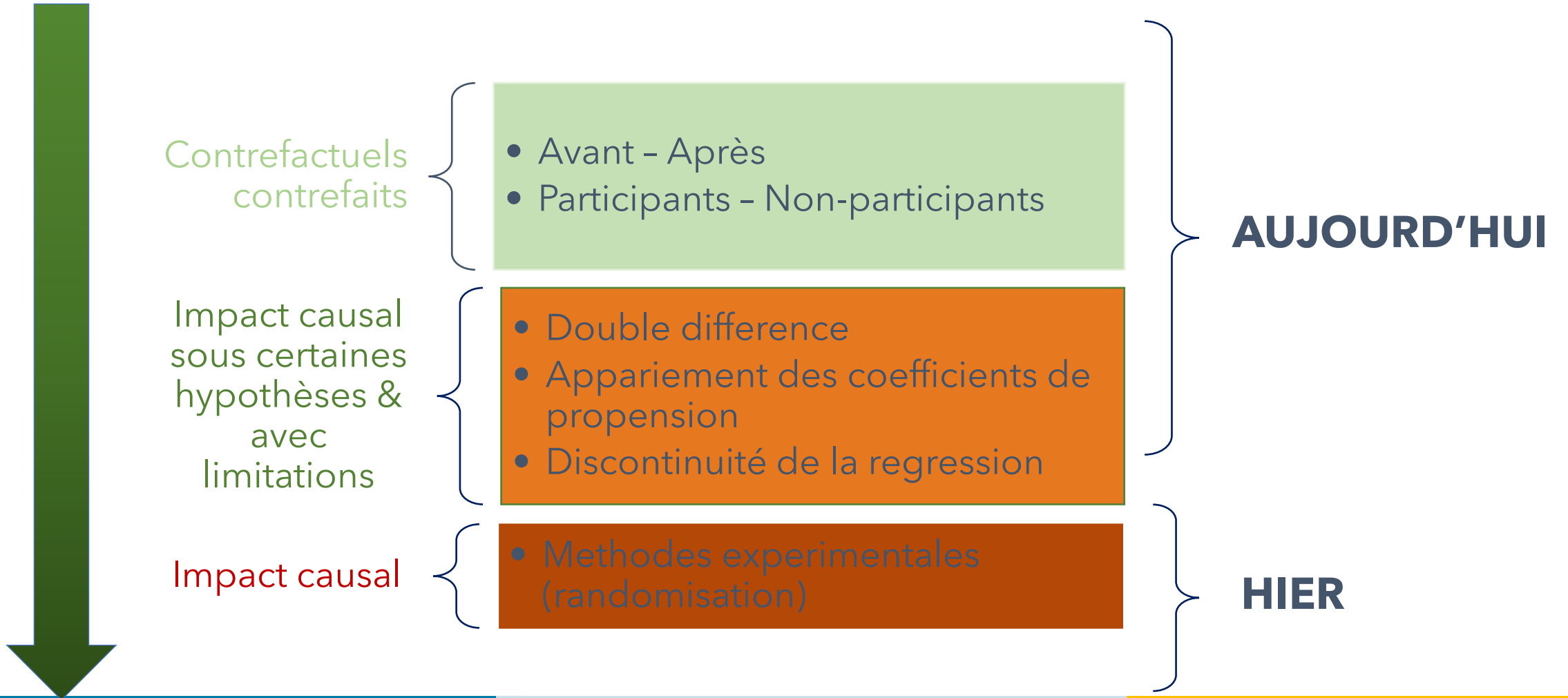


Monde imaginaire !!!

La clé d'une bonne évaluation d'impact est un scénario contrefactuel valide !

Comment trouver un contrefactuel valide?

COMMENT CONSTRUIRE UN CONTREFACTUEL VALIDE POUR MESURER L'IMPACT CAUSAL ?



ETUDE DE CAS

Inclusion économique et bien-être

Problème: La République d'Atlantis veut effectuer un transfert monétaire et de compétences pour améliorer le bien-être socio-économique des ménages les plus pauvres.

Intervention: Le gouvernement décide de lancer un programme pilote pour voir si ce programme sera efficace.

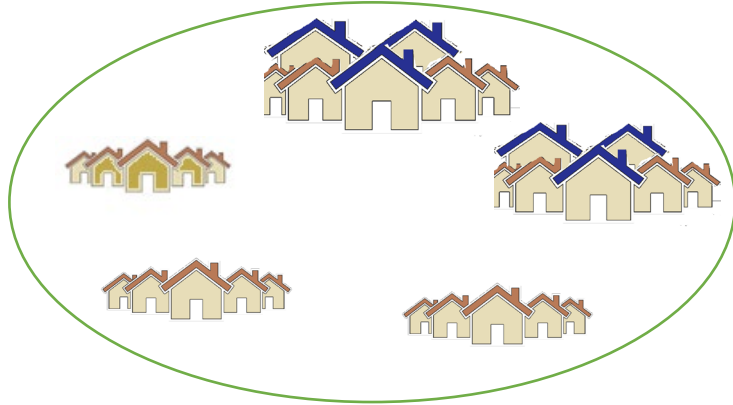
Résultats clés: Bien-être socio-économique, plus particulièrement consommation par habitant des ménages ruraux.

Déroulement :

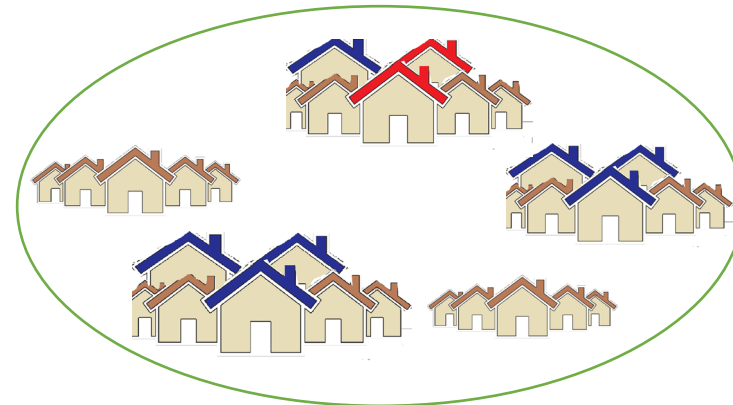
- **Décembre 2019:** Parmi les 9000 ménages ruraux, 2,000 sont identifiés comme éligibles.
- **Mars 2020- mars 2021:** 1021 ménages s'enregistrent à temps et reçoivent le transfert. Les 979 restants sont les non-participants.
- **Avril 2022:** Enquête auprès des ménages des 2000 ménages éligibles

Comparaison avant-après

Participants 2020



Participants 2022



	Participants 2020	Participants 2022	Différence
<i>Consommation des ménages participants</i>	274.4	301.6	27.2 *

Comparaison avant-après

	Participants 2020	Participants 2022	Différence
<i>Consommation dans les villages participants</i>	274.4	301.6	27.2 *

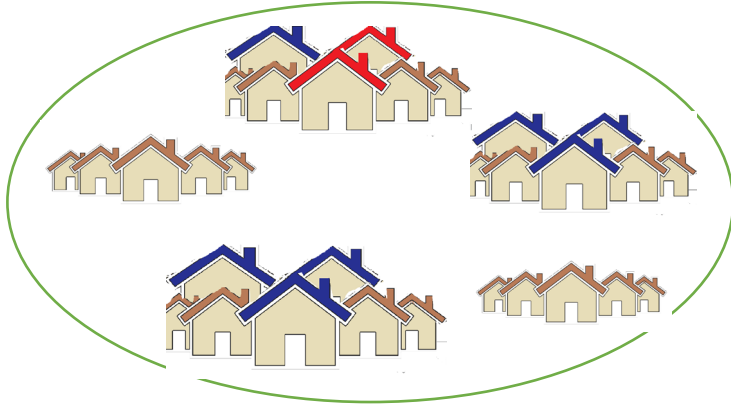
La différence représente-t-elle une estimation non-biaisée de l'impact du programme?

- **Pas forcément: différence dans le temps**
 - Pas que le programme: trop de facteurs changent avec le temps, par ex. dégradation de la conjoncture économique,...
 - Des changements surviennent avec le temps, même sans programme.

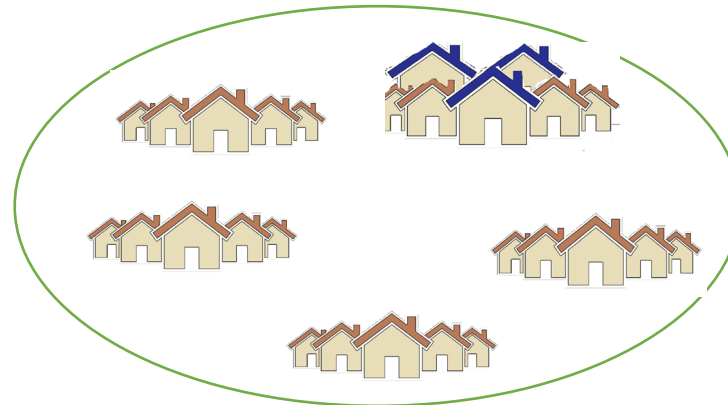
CONTREFACTUELS CONTREFAITS

Comparaison participants / non-participants

Participants



Non-participants



	Participants	Non-participants	Différence
<i>Consommation des ménages</i>	301.6	219.1	82.5 *

CONTREFACTUELS CONTREFAITS

Comparaison participants / non-participants

	Participants	Non-participants	Différence
<i>Consommation des ménages</i>	301.6	219.1	82.5 *

La différence représente-t-elle une estimation non-biaisée de l'impact du programme?

Pas forcément: biais de sélection

- Les ménages participants sont ceux qui ont déposé leur dossier à temps, peut-être plus éduqués, mieux informés (observables)
- Les ménages ont une motivation différente, ont des histoires différentes (non observables)

Méthode de double différence

Compare les différences de résultats entre participants et non-participants dans le programme au fil du temps

Hypothèse d'identification:

- “Tendances parallèles” en l'absence du programme

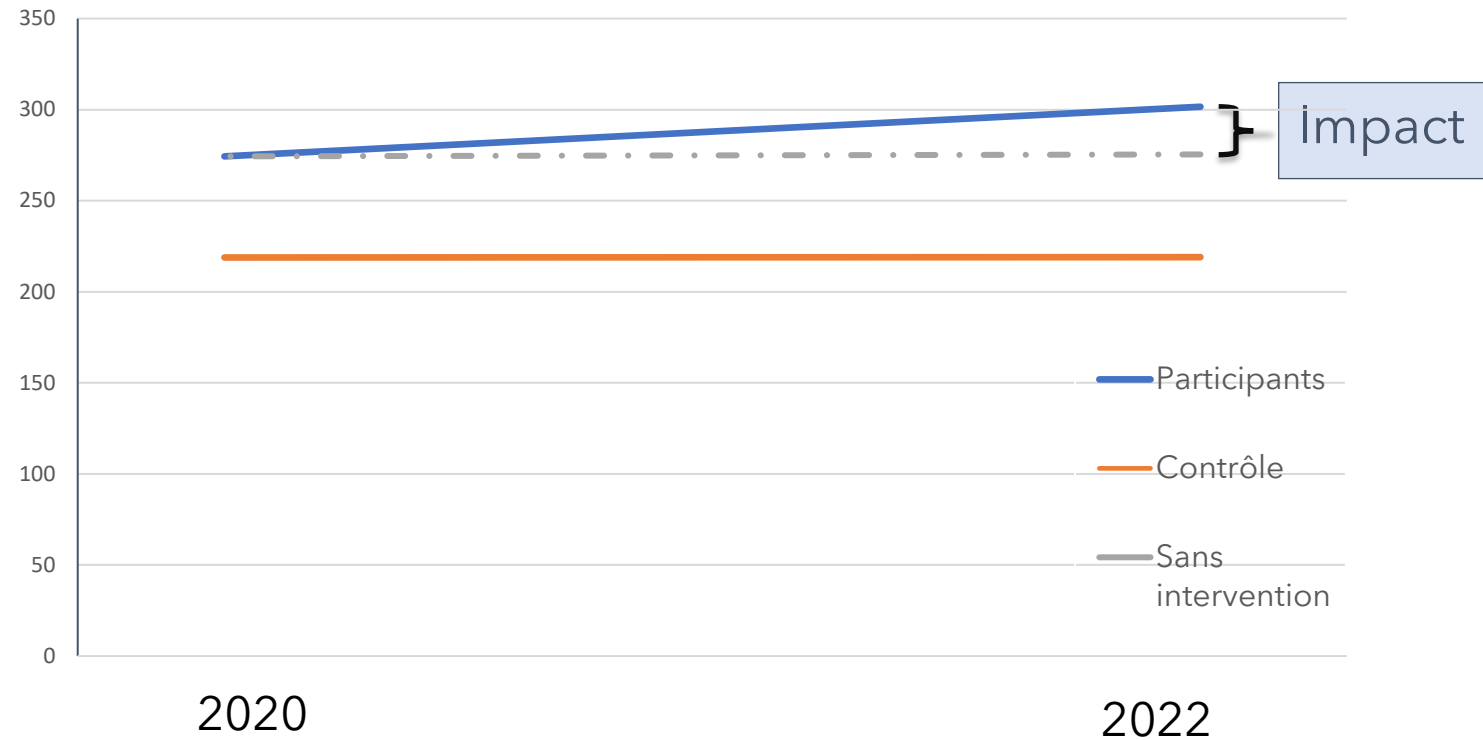
Le contrefactuel

- *C'est le changement au fil du temps pour les non-participants dans le programme*

(Sous certaines hypothèses), la DD peut produire des estimations moins biaisées

Méthode de double différence

Consommation
des ménages



Hypothèse d'identification: "Tendances parallèles" en l'absence du programme

Méthode de double différence

Méthode à suivre:

1. Collecter des données initiales pour chaque groupe avant l'introduction du programme.
2. Collecter les données de suivi pour chaque groupe après que le programme a eu lieu.
3. Calculer la différence avant-après pour chaque groupe
4. Soustraire la différence observée dans le groupe témoin à celle observée dans le groupe traité

Méthode de double différence

	Traitement	Comparaison	Différence
POST- transferts Consommation en 2022	301.6	219.1	82.5
PRÉ-transferts Consommation en 2020	274.4	219	55.4
Différence de niveau de consommation en 2022 et 2020	27.2 (301.6-274.4)	0.1 (219.1-219)	27.1* (301.6-274.4)-(219.1-219) =(Différence-dans la - Différence)

Méthode d'appariement (matching)

Idée: Pour chaque unité traitée, choisir la meilleure unité de comparaison (match) sur la base des similarités dans les caractéristiques observées

Contrefactuel:

Groupe des non-participants avec les mêmes caractéristiques observables que les participants

Hypothèse fondamentale: après appariement...

- Pas de différence systématique dans les caractéristiques non observées influençant la participation et le résultat
- les caractéristiques non observables n'affectent pas l'assignation au traitement ni les résultats étudiés

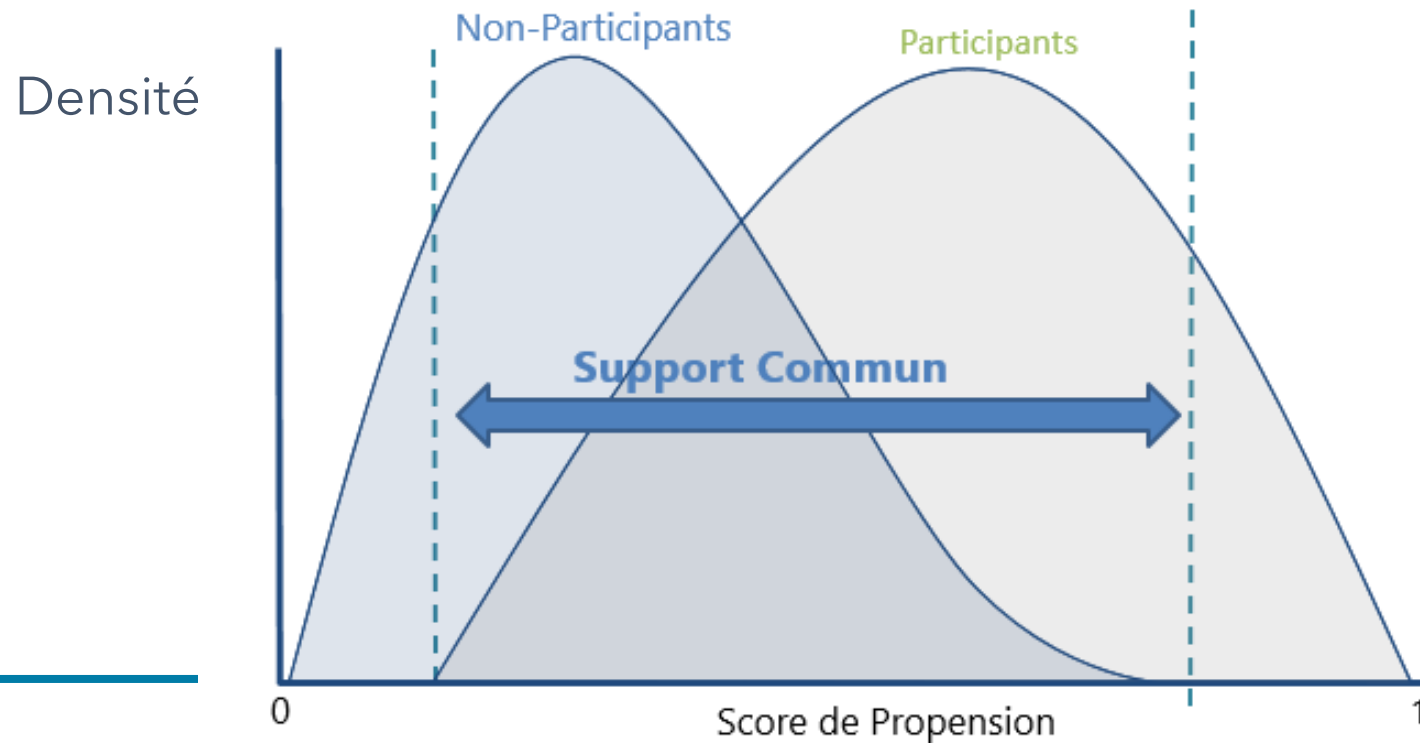
Méthode d'appariement

Apparier sur la base du «score des coefficients de propension»:

- Calculer la probabilité de participation de chacun, basée sur leurs caractéristiques observables.
 - Ex: impact de l'accès à l'eau potable sur la santé des enfants en Inde.
 - Quels sont les facteurs qui prédisent le mieux l'accès à l'eau?
 - Les ménages vivant dans des villages plus grands (en termes de population), des villages avec une école secondaire, une route « pucca » (« goudronnée »), un arrêt de bus, un téléphone, une banque et un marché étaient plus susceptibles d'avoir de l'eau courante.
- Pour chaque participant, trouver un échantillon de non-participants avec un score de propension similaire.
- Comparer les indicateurs des résultats pour chaque observation et le groupe d'observations appariées.

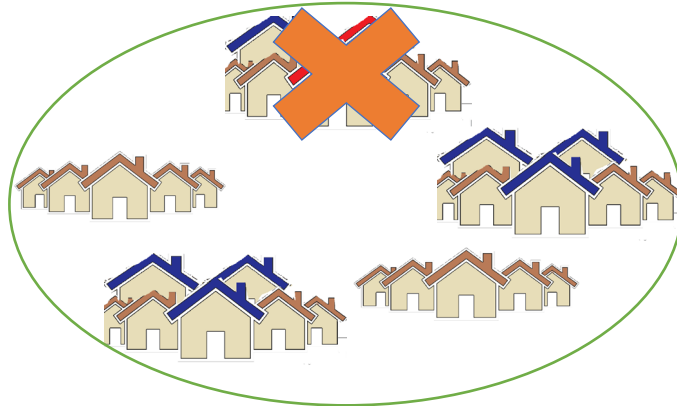
Méthode d'appariement

Densité du score de propension:

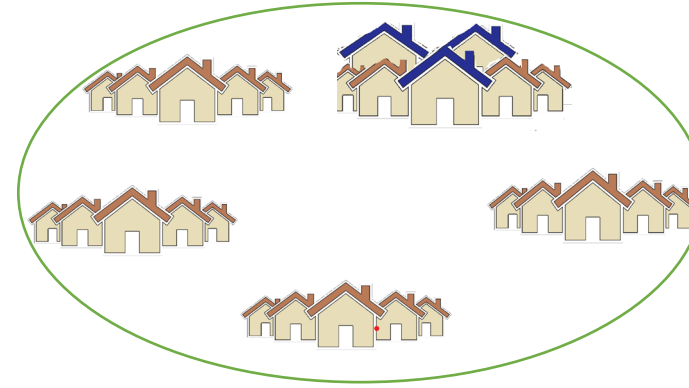


Méthode d'appariement

Participants



Non-participants



	Participants	Comparaison	Différence
Consommation des ménages	290.23	234.41	55.8*
Nombre d'observations utilisées	886	751	

Méthode d'appariement

	Participants	Comparaison	Différence
Consommation des ménages	290.23	234.41	55.8*
Nombre d'observations utilisées	886	751	

- La différence représente-elle une estimation *non-biaisée* de l'impact du programme?
- **Pas forcément: Influence des facteurs non-observables**

MÉTHODES QUASI-EXPERIMENTALES

Méthode de régression discontinue (RD)

Beaucoup de programmes sociaux sélectionnent les bénéficiaires en utilisant un indice ou un score:



Méthode de régression discontinue (RD)

Hypothèse fondamentale:

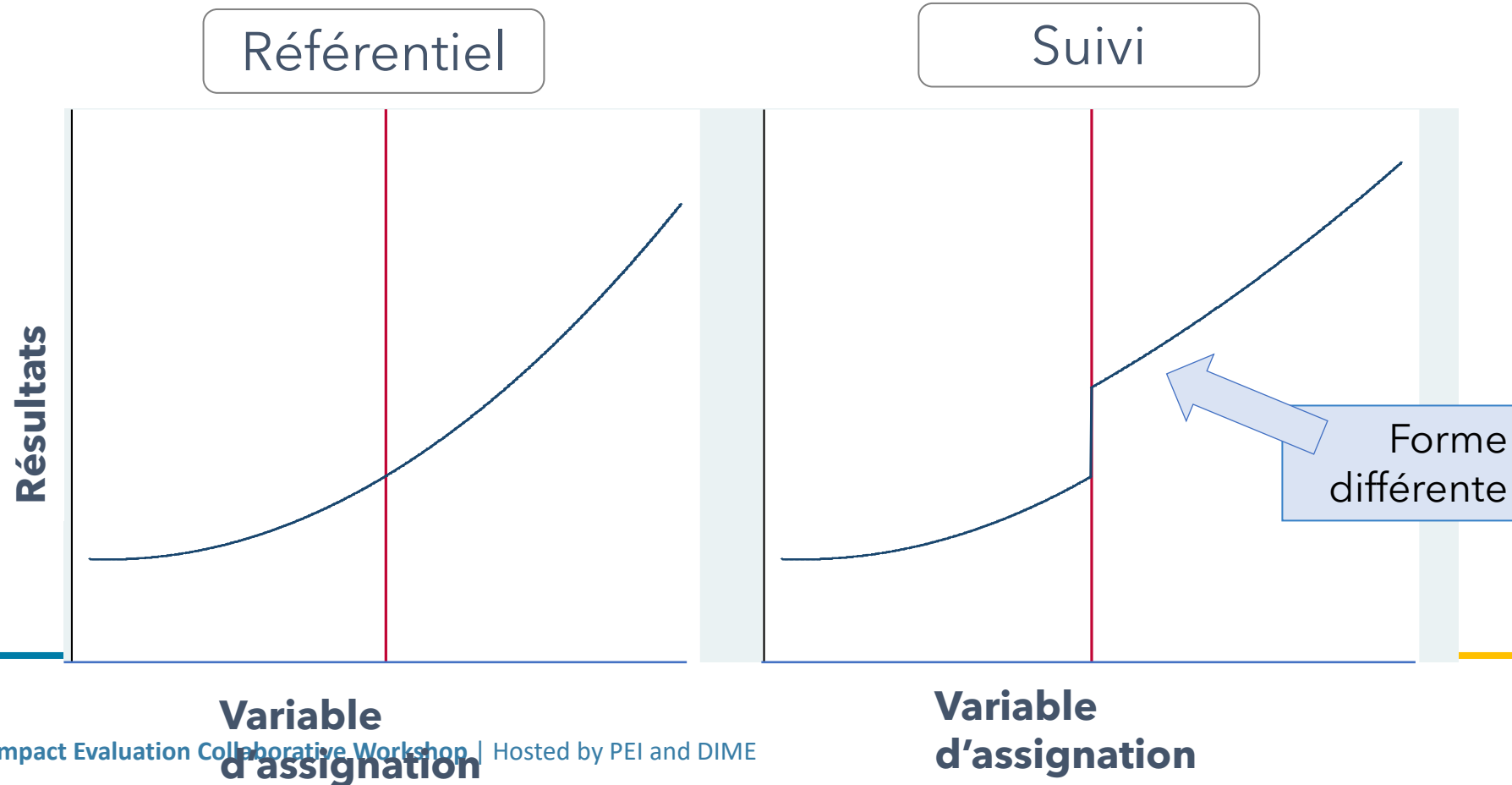
Les unités juste au-dessus du seuil sont comparables à celles juste au-dessous

La RD se base sur la compréhension du processus de sélection:

- Etablir une règle de sélection claire & un score quantifiable simple et continu
- L'assignation du programme est basée sur base d'un seuil
- Cibler les unités autour du seuil pour l'évaluation

Méthode de régression discontinue (RD)

Nous sommes à la recherche de ce type de tendances...



Méthode de régression discontinue (RD)

Exemple: Effet de programme de bourses rurales:

Objectif

- Améliorer la diversification des revenus parmi les femmes des ménages ruraux les plus pauvres

Méthode

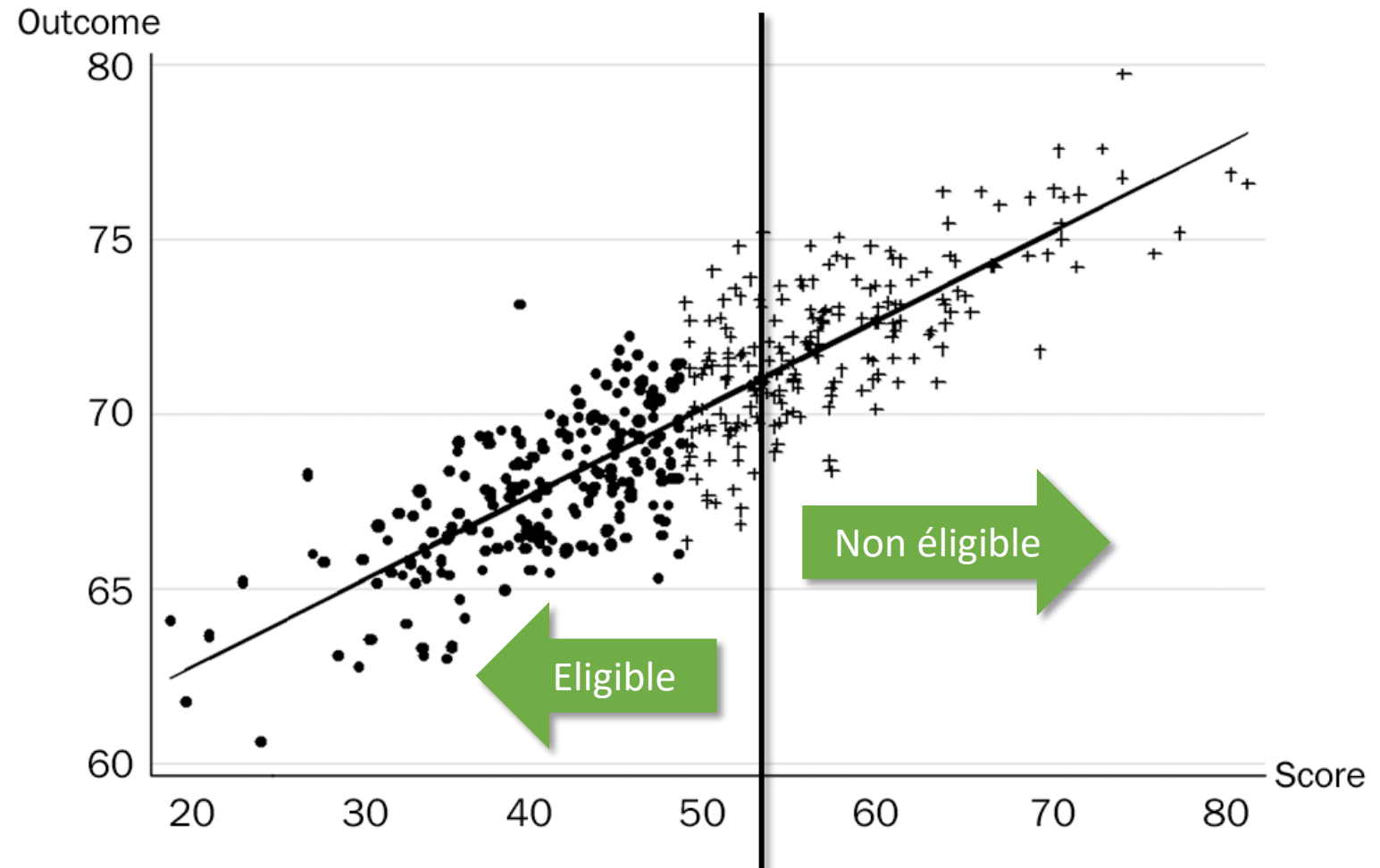
- Ménages avec un score (indice de richesse) ≤ 50 sont pauvres
- Ménages avec un score (indice de richesse) > 50 sont non pauvres

Intervention

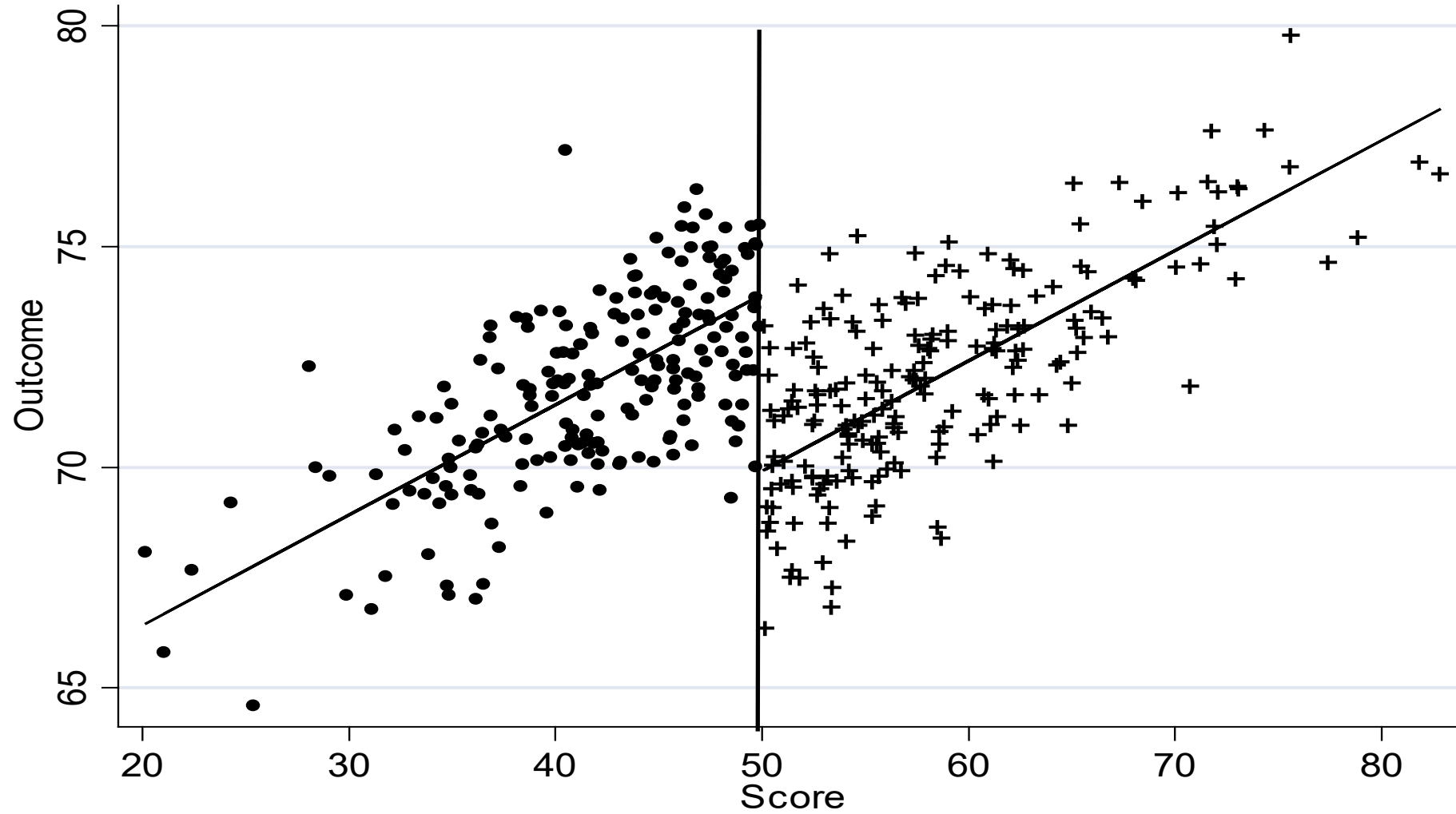
- Les femmes faisant partie des ménages pauvres reçoivent une bourse pour le transport et la commercialisation de leurs produits.

Méthode de régression discontinue (RD)

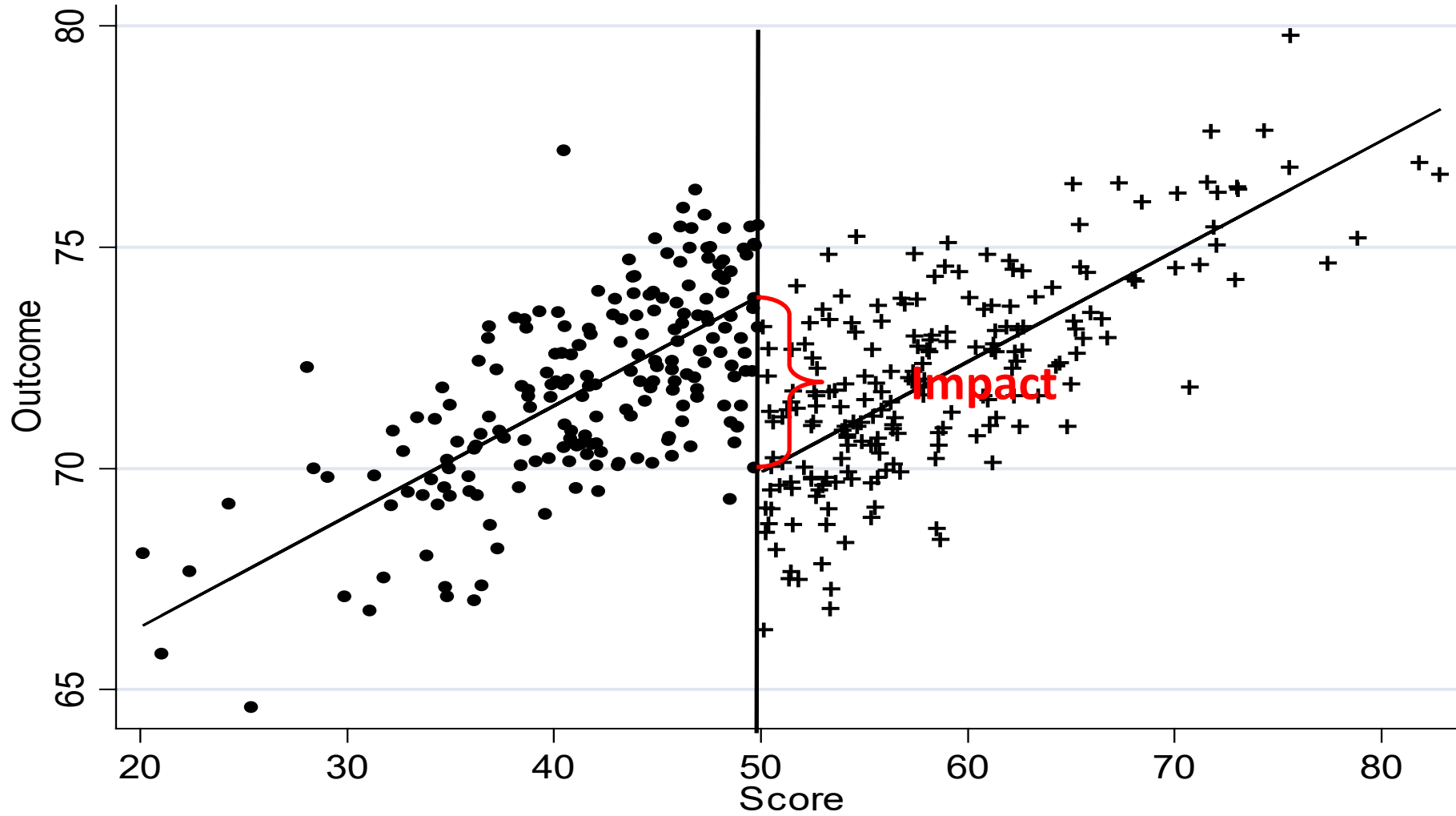
Situation de référence:



Regression Discontinuity Design - Post Intervention



Regression Discontinuity Design - Post Intervention



SYNTHÈSE

Méthodes non-expérimentales

- Toutes les situations ne se prêtent pas à la même méthode.
- Différentes méthodes produisent des résultats différents.
- Les méthodes faibles peuvent conduire à des résultats biaisés.
- Certaines méthodes non-expérimentales nécessitent des hypothèses fortes et de très bonnes données.
- Les résultats de l'évaluation d'impact sont valides seulement si nous utilisons des méthodes rigoureuses.

SYNTHÈSE

Évaluation d'impact

Tout l'art de l'évaluation d'impact consiste à trouver un **bon groupe témoin**.

- La **sélection aléatoire** des bénéficiaires du programme parmi un échantillon d'individus ou de groupes éligibles permet d'obtenir le groupe témoin le plus fiable possible.
- Les évaluations contrôlées randomisées sont donc les plus **crédibles / valides...**
 - Elles sont aussi les plus claires: faciles à expliquer, faciles à comprendre.
 - À condition d'avoir été conçues et mises en oeuvre correctement!

SYNTHÈSE

Évaluation d'impact

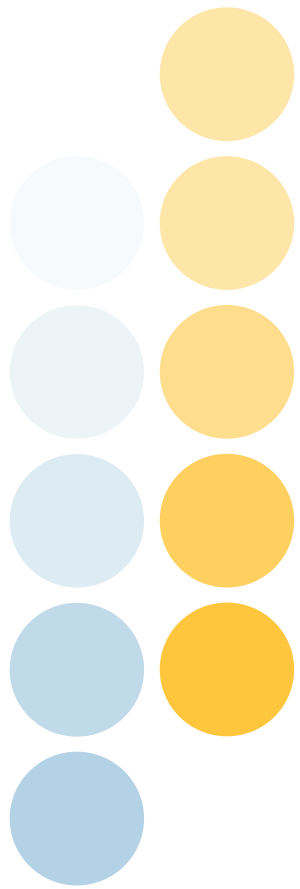
Pour décider des politiques sociales à mettre en oeuvre, il est nécessaire de bien comprendre les relations de cause à effet.

Les relations de cause à effet (e.g. l'impact d'un programme) ne peuvent être estimées correctement que si le "contrefactuel" est estimé correctement.

Sinon il y a des biais de sélection.

Soyez vigilants avec les "résultats" des études non-randomisées.

Demandez-vous toujours comment le contrefactuel a été estimé ...



Clara Delavallade

cdelavallade@worldbank.org

Thank you!

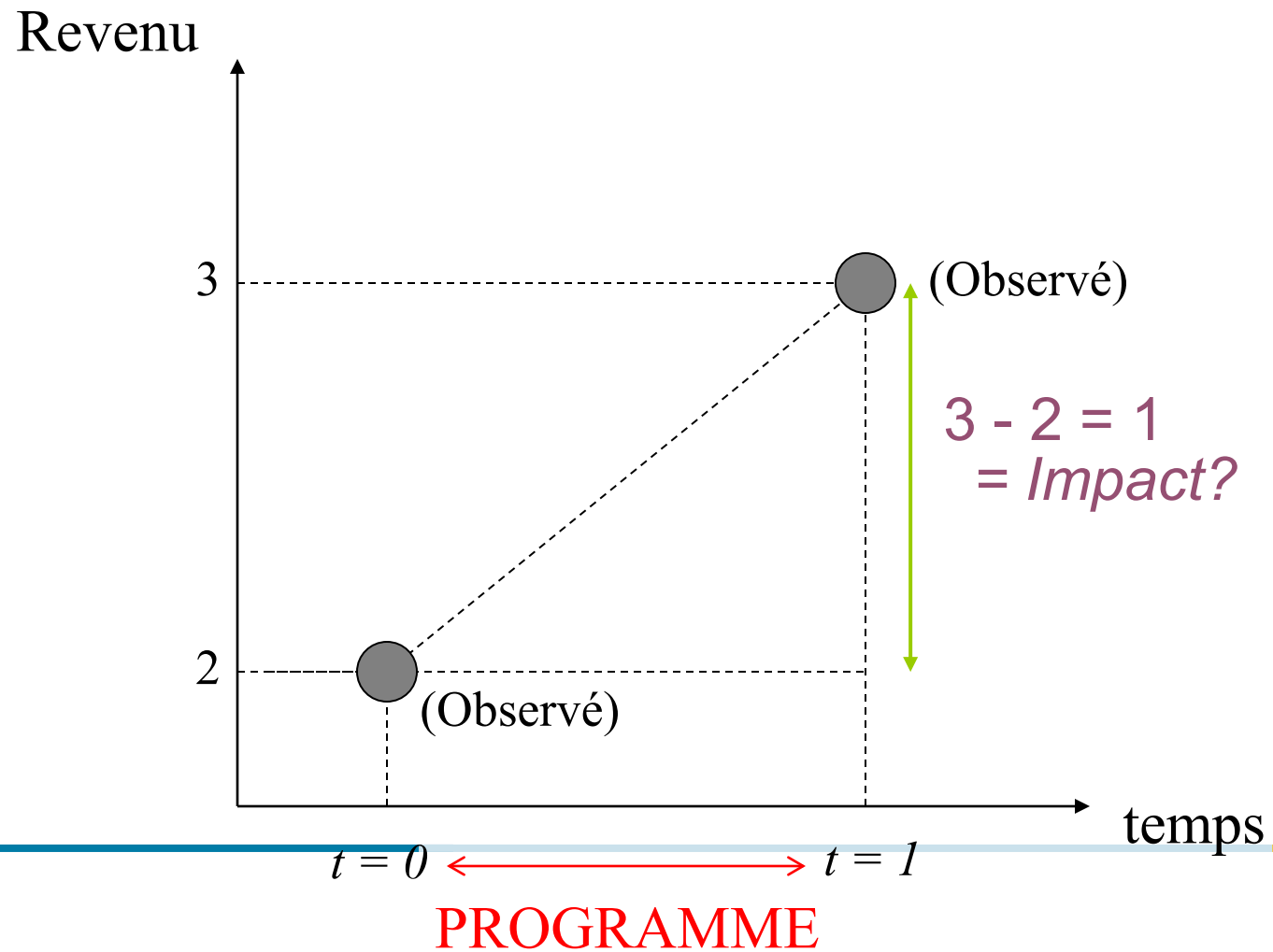


PEI FUNDING PARTNERS

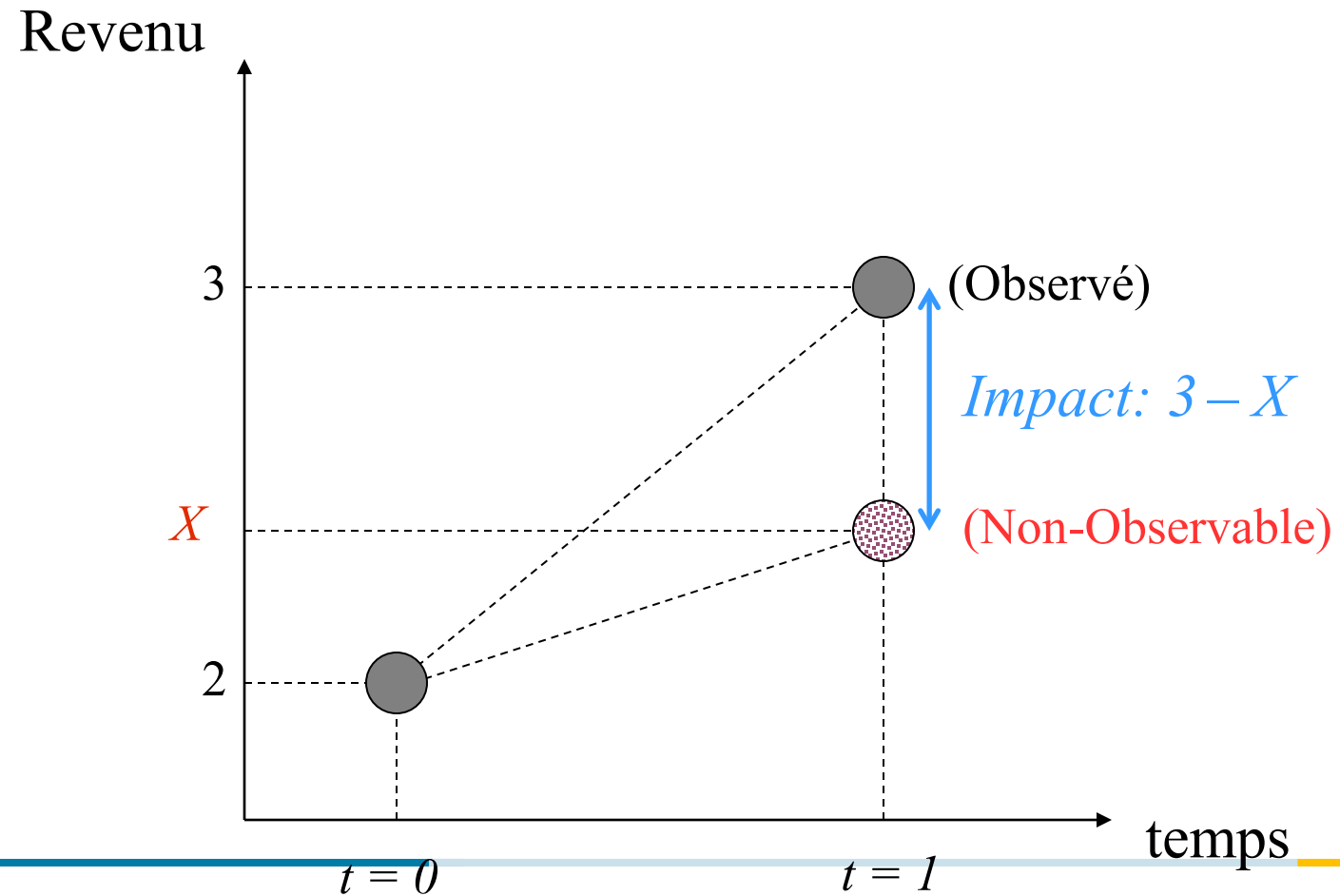


Implemented by

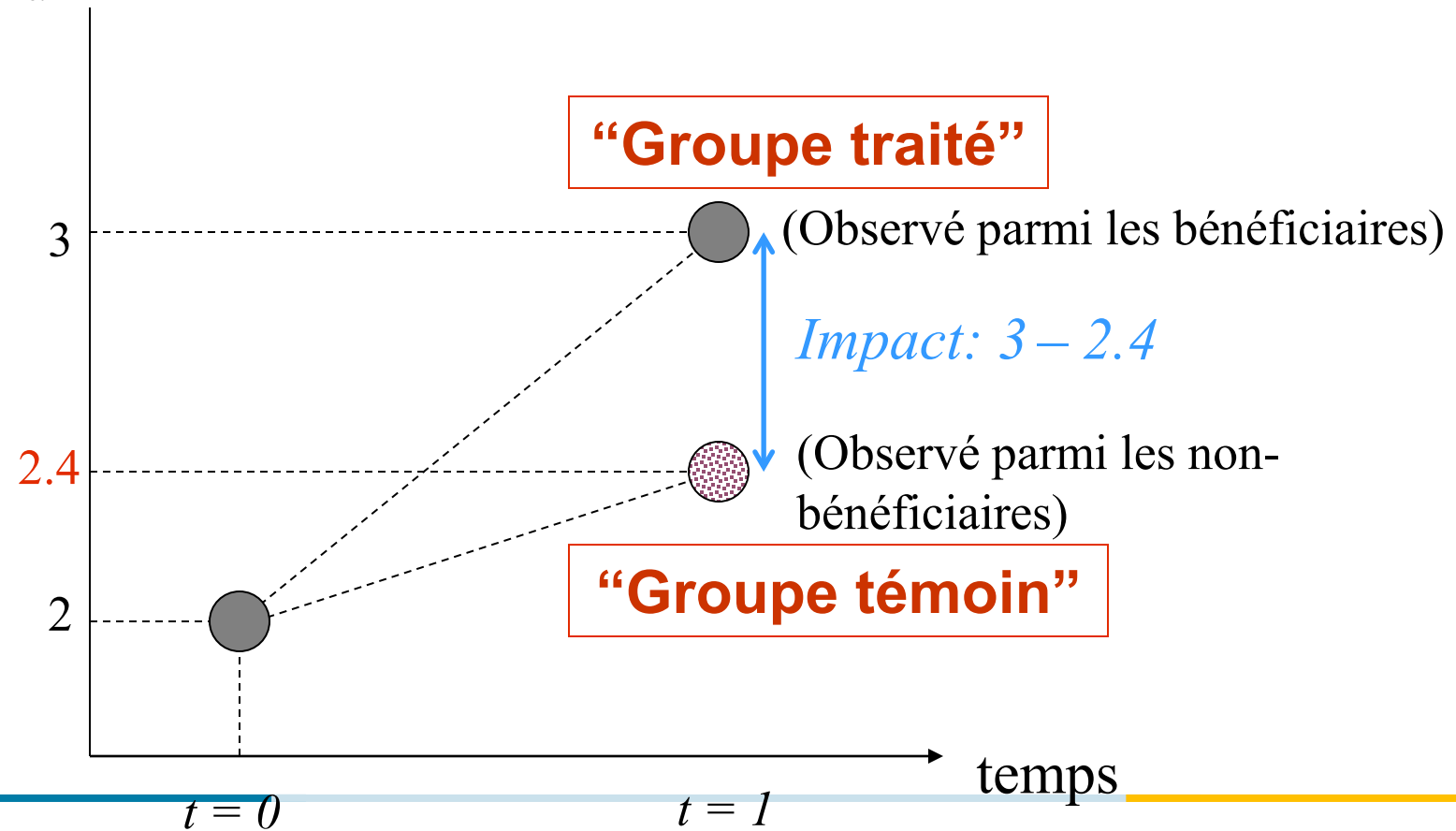


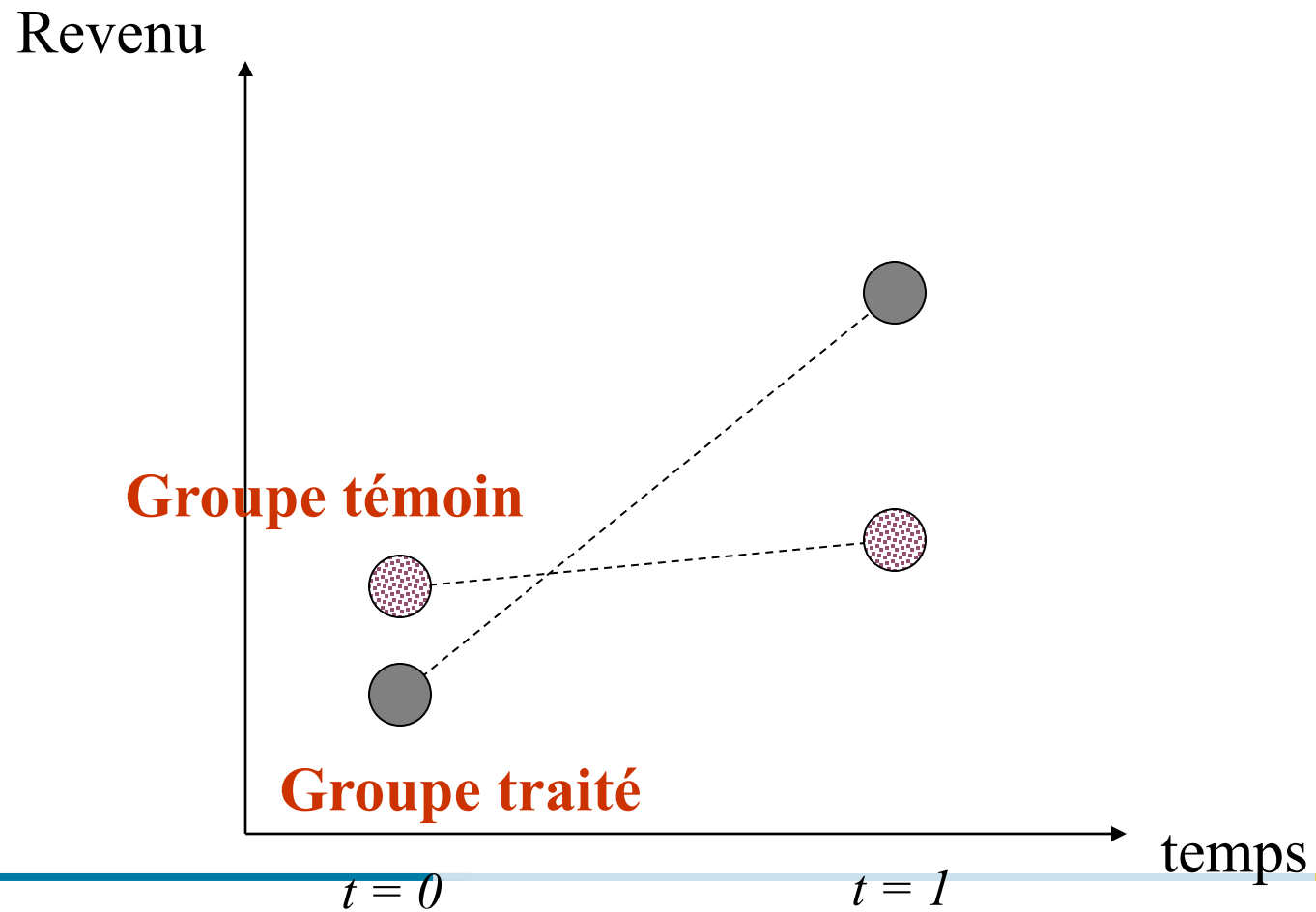


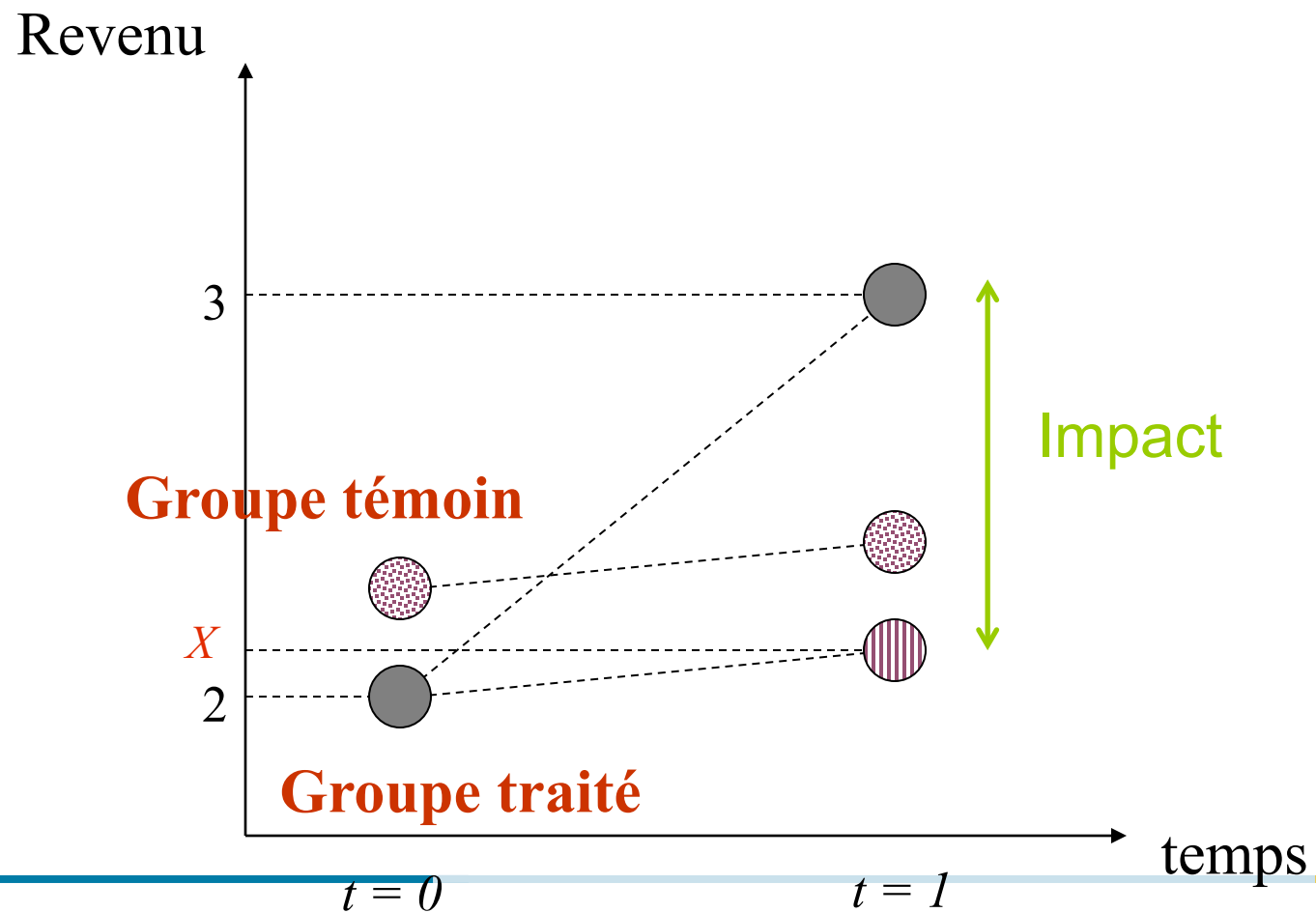
Que se serait-il passé en l'absence du programme?



Revenu







Revenu

