

# **Atelier de formation virtuel sur les Évaluations d'Impact Contrefactuelles (EIC)**

**C4ED – EUTF**  
Septembre 2021

# **Bienvenue à l'atelier de formation sur les Évaluations d'Impact Contrefactuelles (EIC)**

Le contenu de cet atelier a été développé avec le soutien financier de l'Union Européenne. Ce contenu est la seule responsabilité de C4ED et ne reflète pas nécessairement les opinions de l'Union Européenne.

## Jour 2 - Agenda

<b>09:00 – 09:30</b>	Discussion de l'exercice sur les Méthodes Expérimentales
<b>5 min</b>	Pause
<b>9:35 – 10:35</b>	Session 4A: Introduction aux Méthodes Quasi-Expérimentales: Double Différence (DD), Appariement (Matching), Variable Instrumentale (VI), Régression sur Discontinuité (RSD)
<b>15 min</b>	Pause
<b>10:50 – 12:00</b>	Session 4B: Introduction aux Méthodes Quasi-Expérimentales: DD, Matching, VI, RSD + Exercice d'application
<b>45 min</b>	Pause déjeuner
<b>12:45 – 13:05</b>	Discussion de l'exercice sur les Méthodes Quasi-Expérimentales
<b>13:05 – 14:05</b>	Session 5A: Gestion des attentes – échéances, données (sources de données et taille d'échantillon) et budget
<b>15 min</b>	Pause
<b>14:20 – 15:00</b>	Session 5B: Gestion des attentes – échéances, données (sources de données et taille d'échantillon) et budget
<b>15:00 – 15:30</b>	Q&R

# Communication durant l'atelier

---

- Posez vos questions dans la zone de discussion
- Réagissez (« like ») aux questions posées par d'autres, pour nous indiquer qu'elles sont particulièrement pertinentes ou urgentes
- Nous répondrons aux questions au fur et à mesure
- Profitez des longues pauses pour poser plus de questions
- Suggérez des améliorations s'il vous est difficile de suivre ou si vous n'êtes pas d'accord (nous sommes ouverts aux critiques constructives et aux suggestions pour nous améliorer)
- Pour plus de questions et de feedback (en particulier pour la session Q&R):

Envoyez un email à Marc Gillaizeau  
([m.gillaizeau@c4ed.org](mailto:m.gillaizeau@c4ed.org))

# EI Expérimentale - Récapitulatif

---

## Evaluation d'Impact (EI) expérimentale

- Utilise un contrefactuel qui assure que les unités dans les groupes T et C sont statistiquement similaires en moyenne en termes de caractéristiques (non) observables, grâce à l'allocation aléatoire de l'intervention
- Cadre expérimental pas toujours réalisable:
  - Randomisation n'est peut-être pas acceptable politiquement ou socialement
  - Randomisation n'est peut-être simplement pas possible
  - EI élaborée seulement après le début de la mise en place de l'intervention

# Méthodes d'EI Non Expérimentales

---

## EI Non-Expérimentale

- Le contrefactuel est construit sur la base d'hypothèse qui permettent d'établir des groupes T et C comparables
- Hypothèses ne peuvent pas être testées scientifiquement (en général)

# L'impact a souvent été mesuré...

---

## COMPARAISON AVANT-APRÈS

# Comparaison Avant-Après

---

- On peut comparer les résultats des participants au programme AVANT at APRÈS l'intervention
- Requis: données avant et après l'intervention

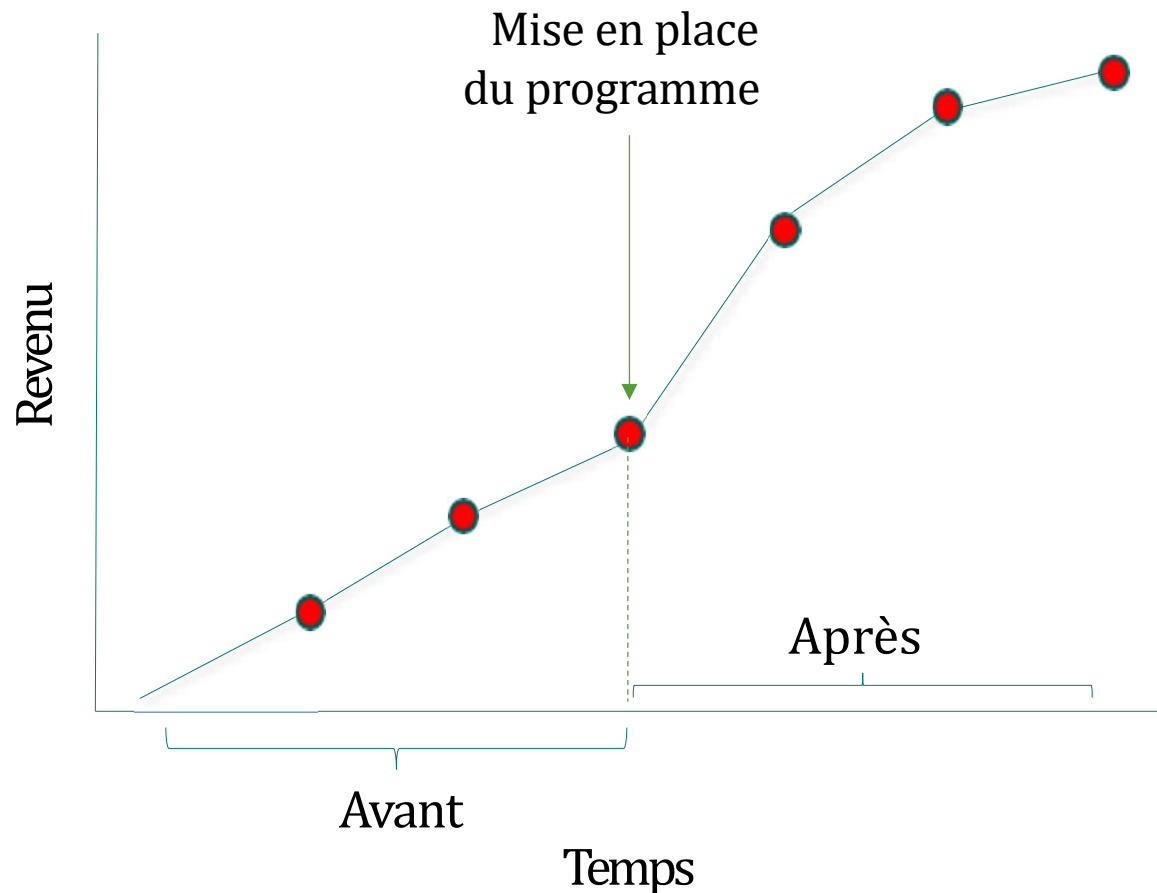


# Comparaison Avant-Après

## EXEMPLE

Programme: formation professionnelle pour les jeunes

Groupe de traitement: groupe des jeunes inscrits dans le programme

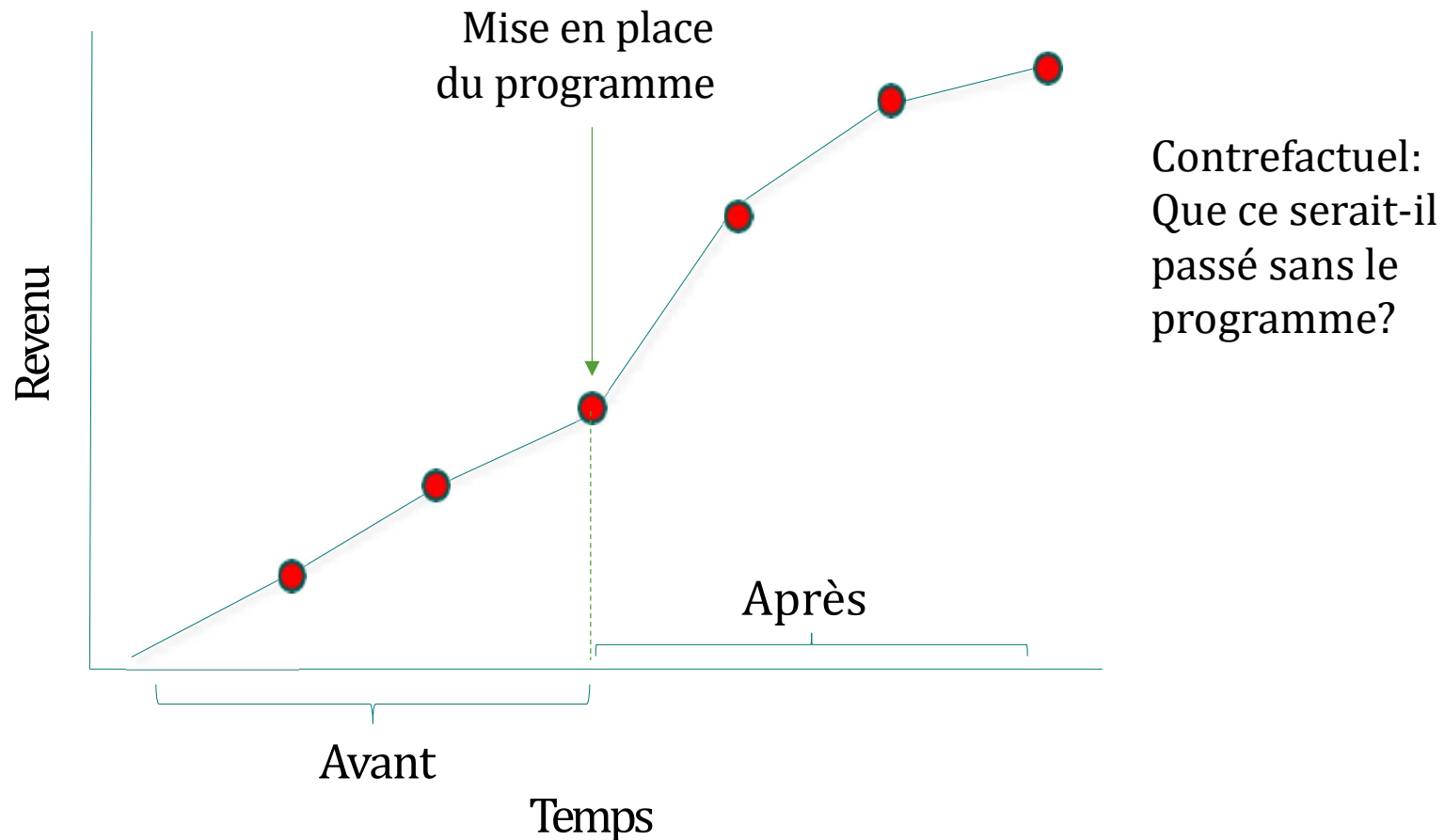


# Comparaison Avant-Après

## EXEMPLE

Programme: formation professionnelle pour les jeunes

Groupe de traitement: groupe des jeunes inscrits dans le programme

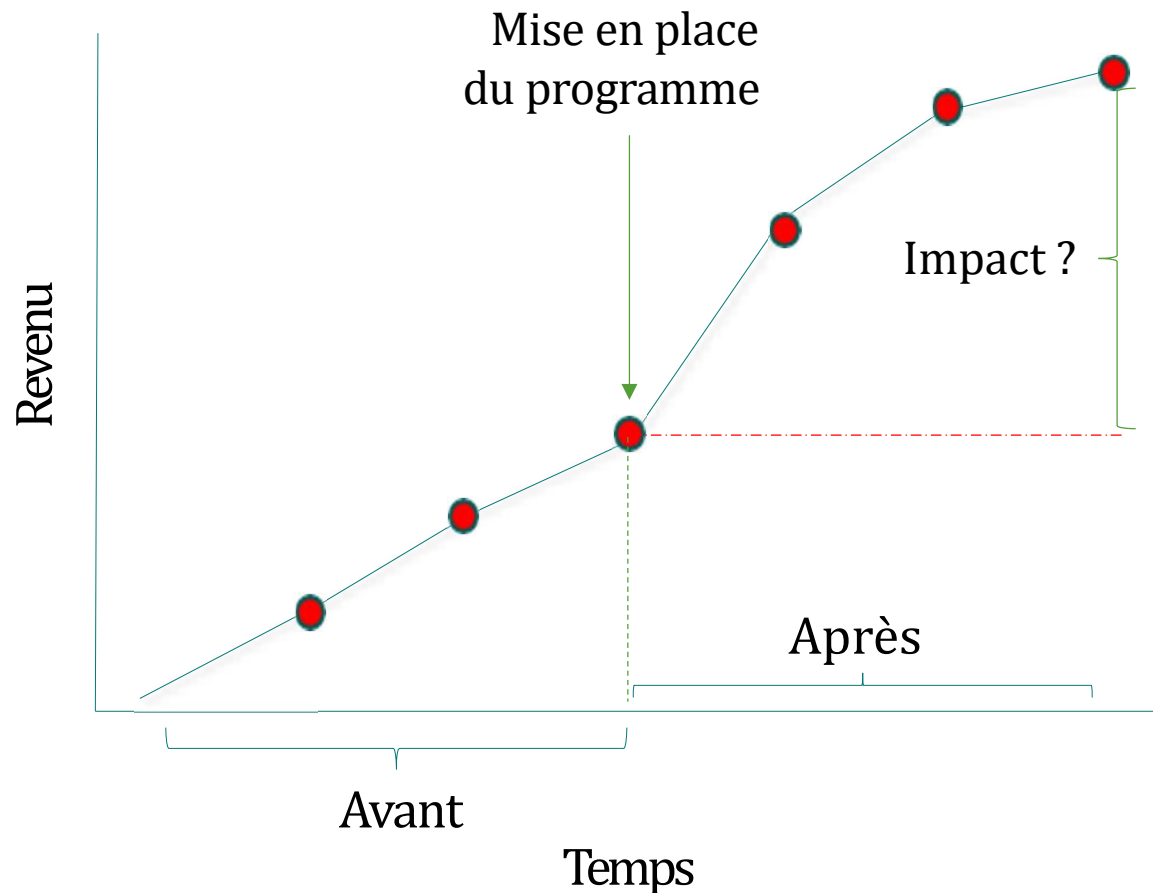


# Comparaison Avant-Après

## EXEMPLE

Programme: formation professionnelle pour les jeunes

Groupe de traitement: groupe des jeunes inscrits dans le programme



Contrefactuel:  
Que serait-il  
passé sans le  
programme?

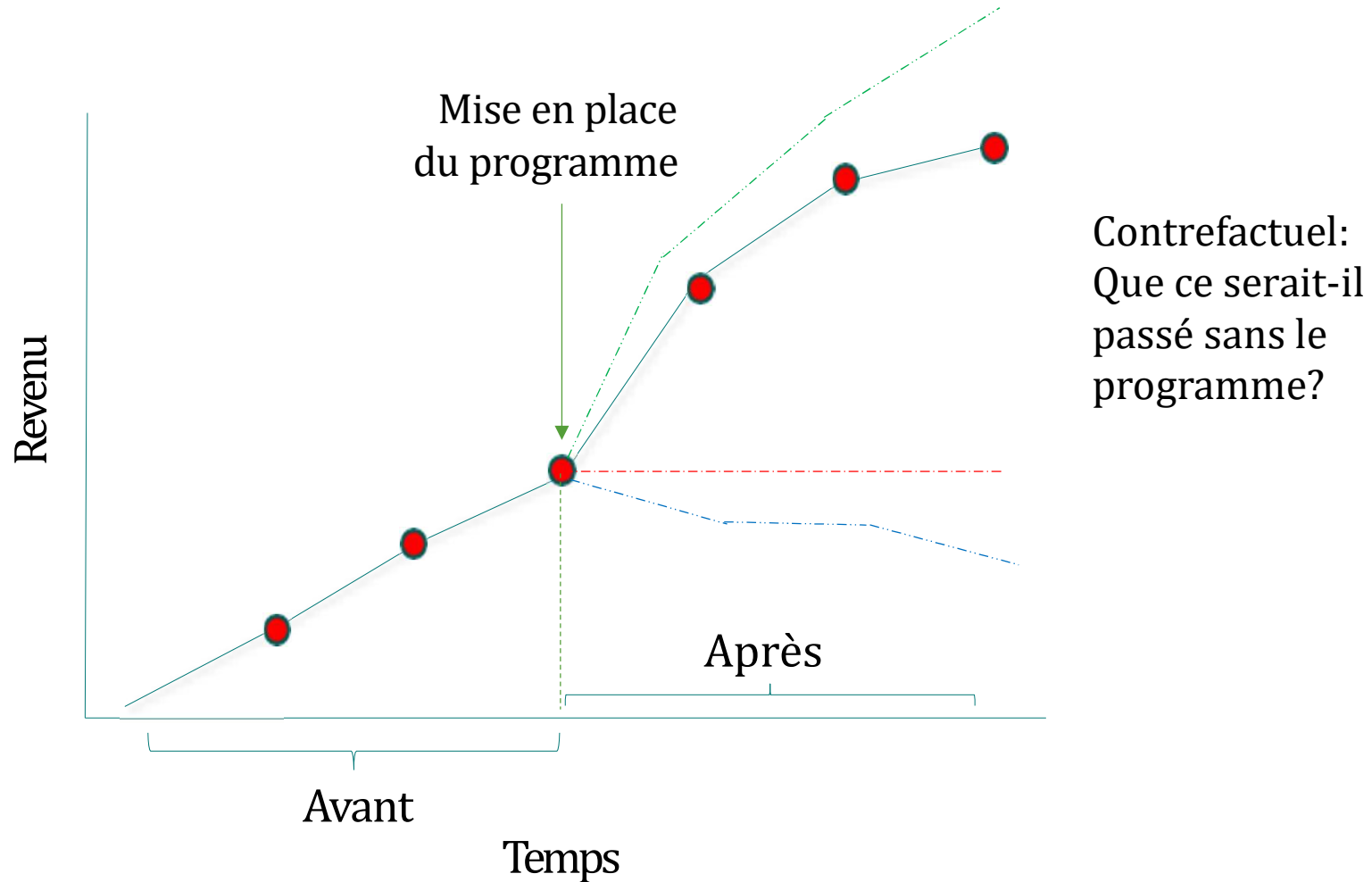


# Comparaison Avant-Après

---

- Il est probable que – en raison de facteurs **externes** qui influencent le revenu (macroéconomiques, géographiques, climatiques, etc.) – le niveau de revenu ne soit pas resté le même (ou ne soit pas resté sur la même tendance) **sans** le programme de formation
  - Les résultats de référence (« baseline ») ne sont probablement pas une bonne estimation du contrefactuel

# Comparaison Avant-Après



# Comparaison Avant-Après

---

- **Hypothèses:**

- Le niveau de la variable de résultat d'intérêt n'aurait pas change sans l'intervention
- Il n'y a pas d'autres facteurs (que le programme) qui ont influence la variable de résultat au cours du temps

- **Difficultés:** On ne sait pas ce qui ce serait passé sans l'intervention, car il n'y a pas de groupe de comparaison

# Le recours à un groupe de comparaison

---

... On renforce la crédibilité des résultats en mesurant le contrefactuel au moyen d'un groupe de comparaison...

# Le choix d'un groupe de comparaison

---

Dans une évaluation d'impact, la notion de « **RIGUEUR** » se résume principalement à déterminer **le meilleur groupe de comparaison possible**



# Un bon groupe de comparaison

---

- Présente les mêmes caractéristiques (en moyenne) que le groupe de traitement
- Ne bénéficie pas du programme
- Réagirait au programme de la même façon que le groupe de traitement (s'il devait en bénéficier)

# L'impact a souvent été mesuré par...

---

**Différence Simple**

# Différence Simple

---

- On peut comparer (au moins) deux groupes, c'est-à-dire les participants par rapport aux non-participants, à un instant donné
- Il faut disposer des données d'au moins deux groupes à un moment donné

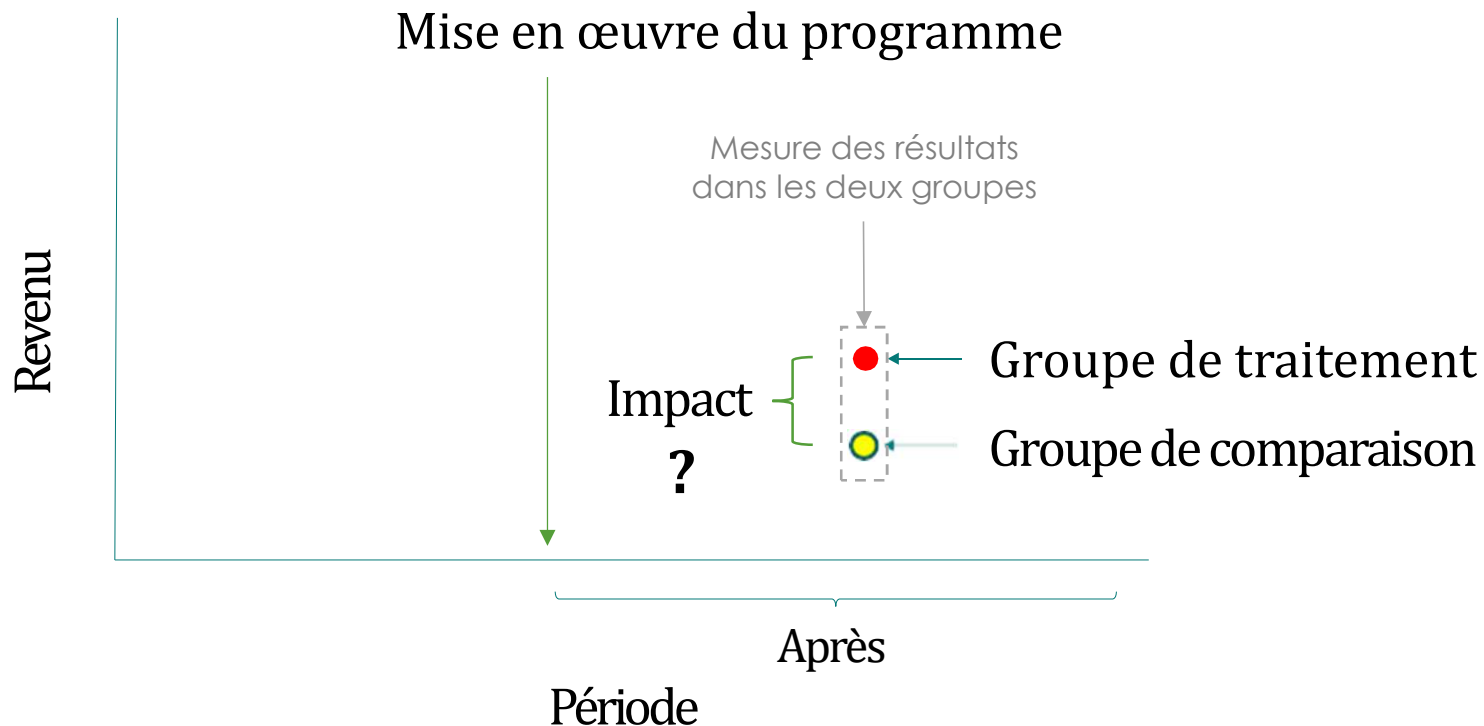
# Différence Simple

## EXEMPLE:

Programme : programme de formation professionnelle destiné aux jeunes

Groupe de traitement : le groupe des jeunes inscrits

**Groupe de comparaison :** le groupe des jeunes qui, bien qu'éligibles, ont choisi de ne pas s'inscrire



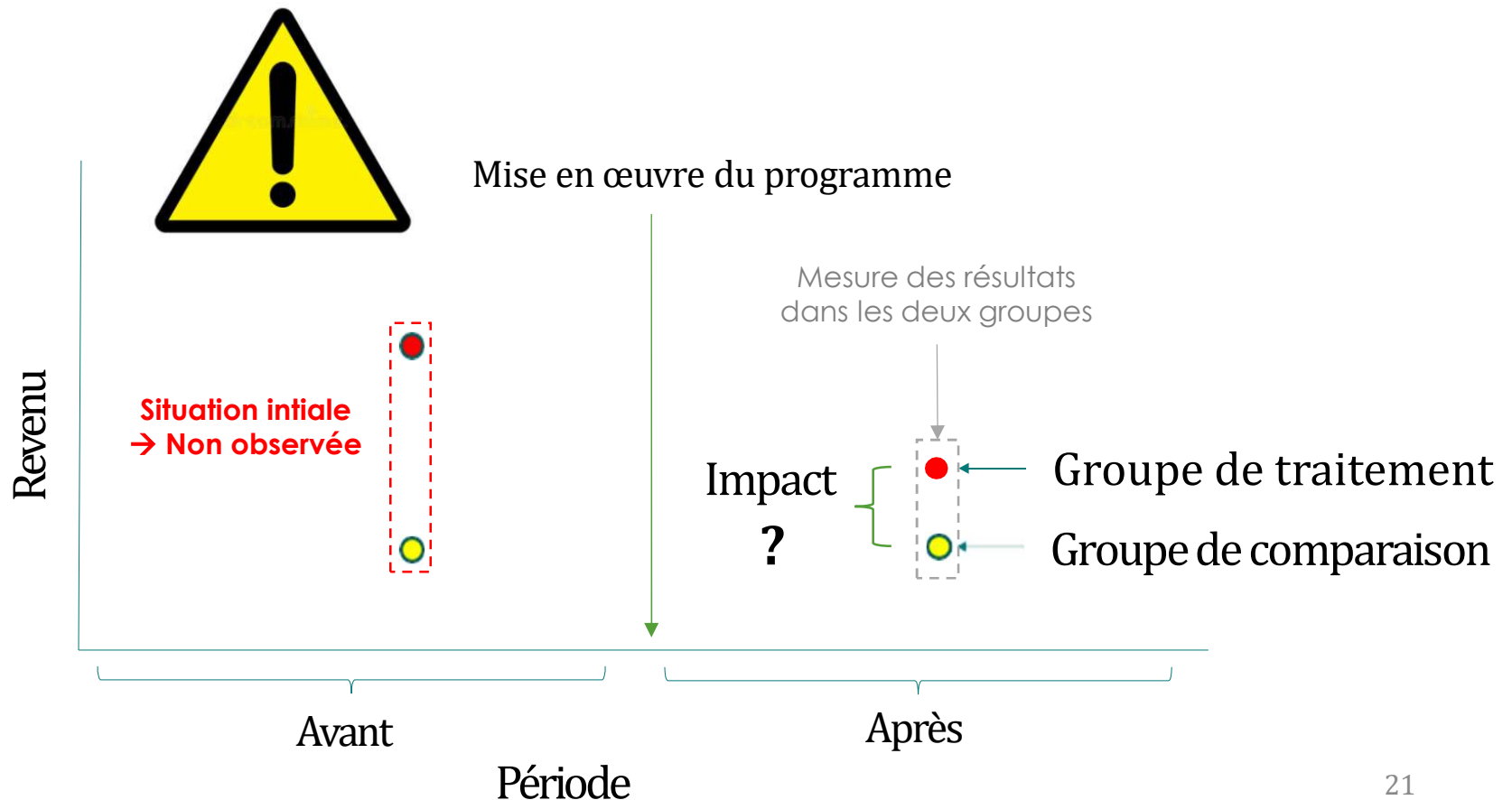
# Différence Simple

## EXEMPLE:

Programme : programme de formation professionnelle destiné aux jeunes

Groupe de traitement : le groupe des jeunes inscrits

**Groupe de comparaison :** le groupe des jeunes qui, bien qu'éligibles, ont choisi de ne pas s'inscrire



# Différence Simple

---

- Le groupe de comparaison dans l'exemple précédent ne fournit pas une bonne estimation du contrefactuel
- Si l'on observe une différence de revenus après la formation entre les 2 groupes, on ne peut pas distinguer si cela est dû à la formation ou à des différences sous-jacentes en termes de motivation, de capacités ou d'autres facteurs qui existent entre les 2 groupes
- Cela crée un biais dans l'estimation de l'impact: le ***biais de sélection***

# Différence Simple

---

- Les jeunes qui décident de participer sont peut-être très motivés et anticipent de larges bénéfices de la formation
- Les jeunes qui décident de ne pas participer sont peut-être peu motivés et ne pensent pas que la formation soit bénéfique

→ Il est probable que ces 2 groupes auraient eu des comportements différents sur le marché du travail et auraient connus des niveaux de revenus différents **même en l'absence du programme** de formation professionnelle

# Méthodes non expérimentales

**1. Double  
Différence**

**2. Appariement  
("Matching")**

Méthodes non  
expérimentales

**3. Variables  
Instrumentales  
(VI)**

**4. Régression sur  
Discontinuité  
(RSD)**



# Double Différence

---

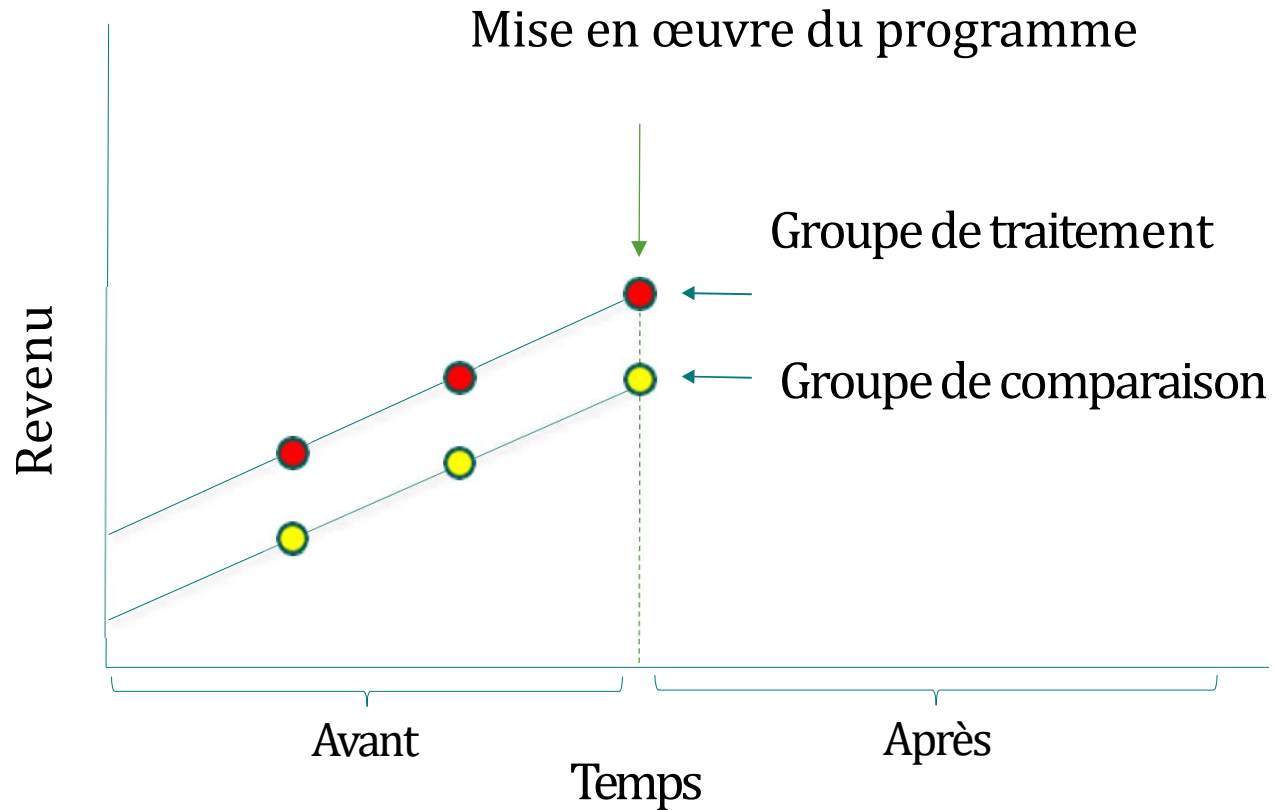
- Programme : programme de formation professionnelle destiné aux jeunes
  - Certains conseils locaux décident d'adhérer au programme, d'autres non
  - La comparaison simple des revenus des jeunes entre les districts inscrits et non inscrits pose problème (biais de sélection)
  - Idée : Et si nous combinions les deux méthodes précédemment discutées ?
    - La différence avant-après (pour chaque groupe) permet d'éliminer les biais dus aux facteurs externes qui sont **constants** dans le temps
    - La différence entre les groupes (à chaque période) permet de contrôler les différences de conditions initiales
- **Double Différence**

# Double Différence

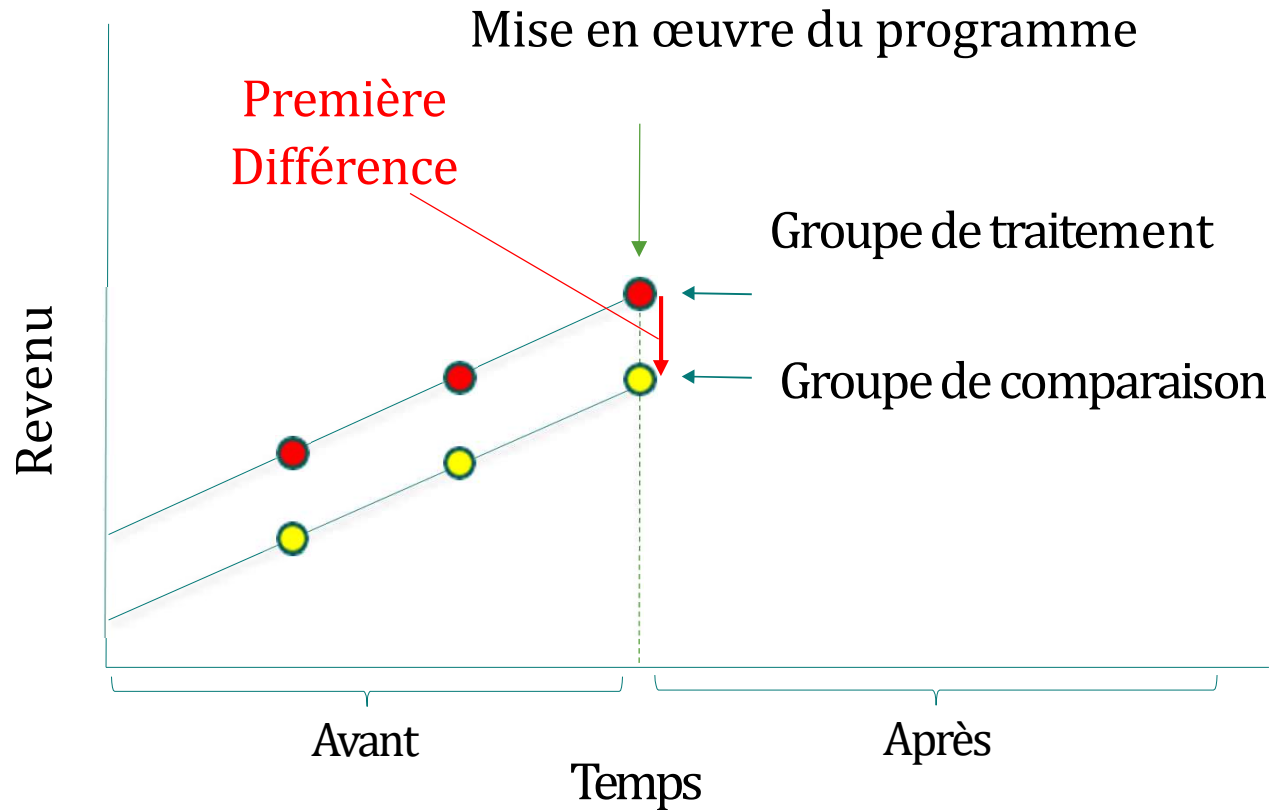
---

- Comparez (au moins) deux groupes, c'est-à-dire les participants par rapport aux non-participants, dans le temps
- **Hypothèses:**
  - Deux groupes qui sont comparables au niveau de la variable de résultat d'intérêt résultat visé et qui présentent les ***des tendances d'évolution parallèles.***
  - Des tendances d'évolution similaires observées avant le début du programme peuvent servir de preuve suggestive que cela soit aussi le cas en l'absence du programme

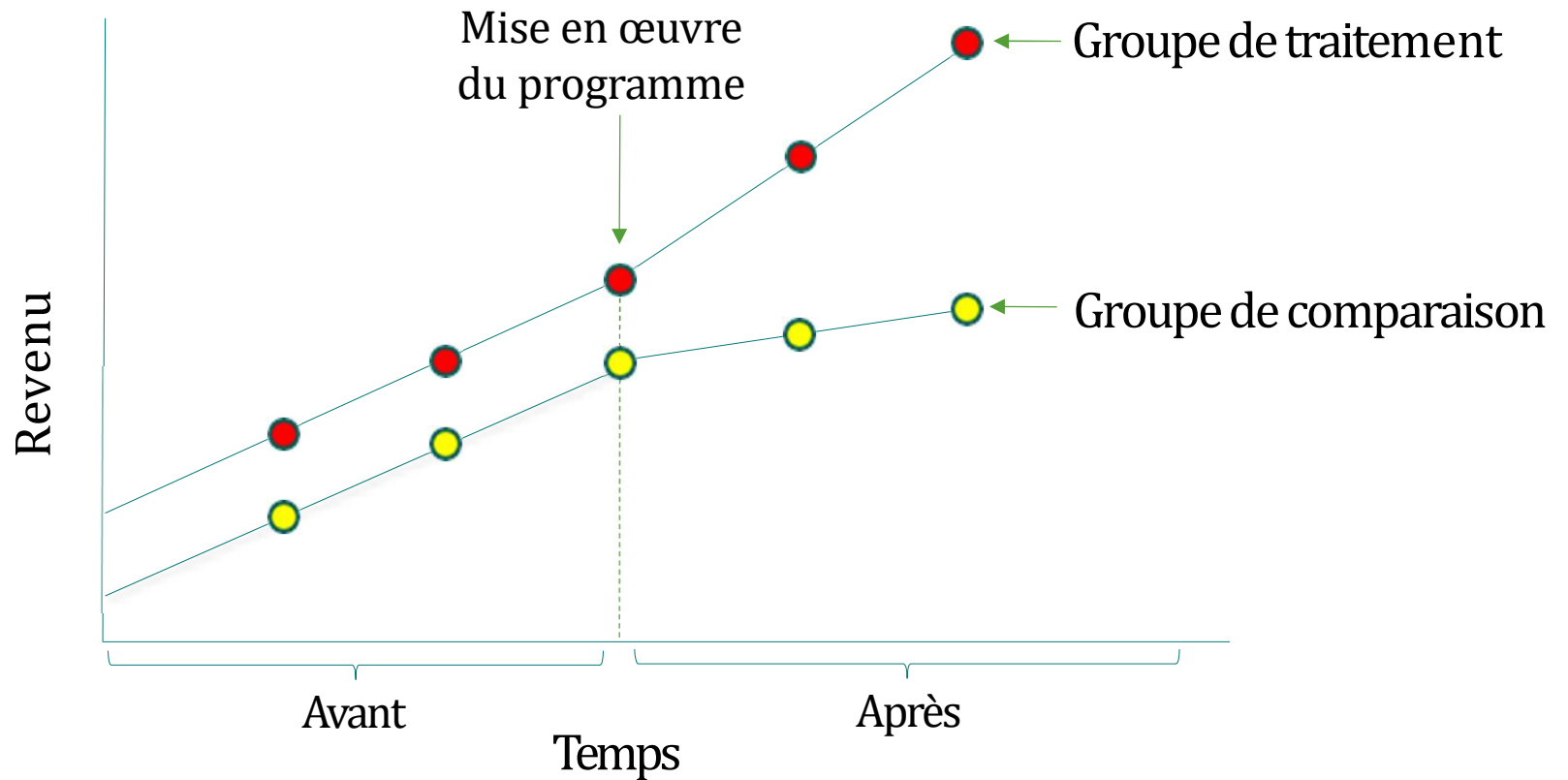
# Double Différence



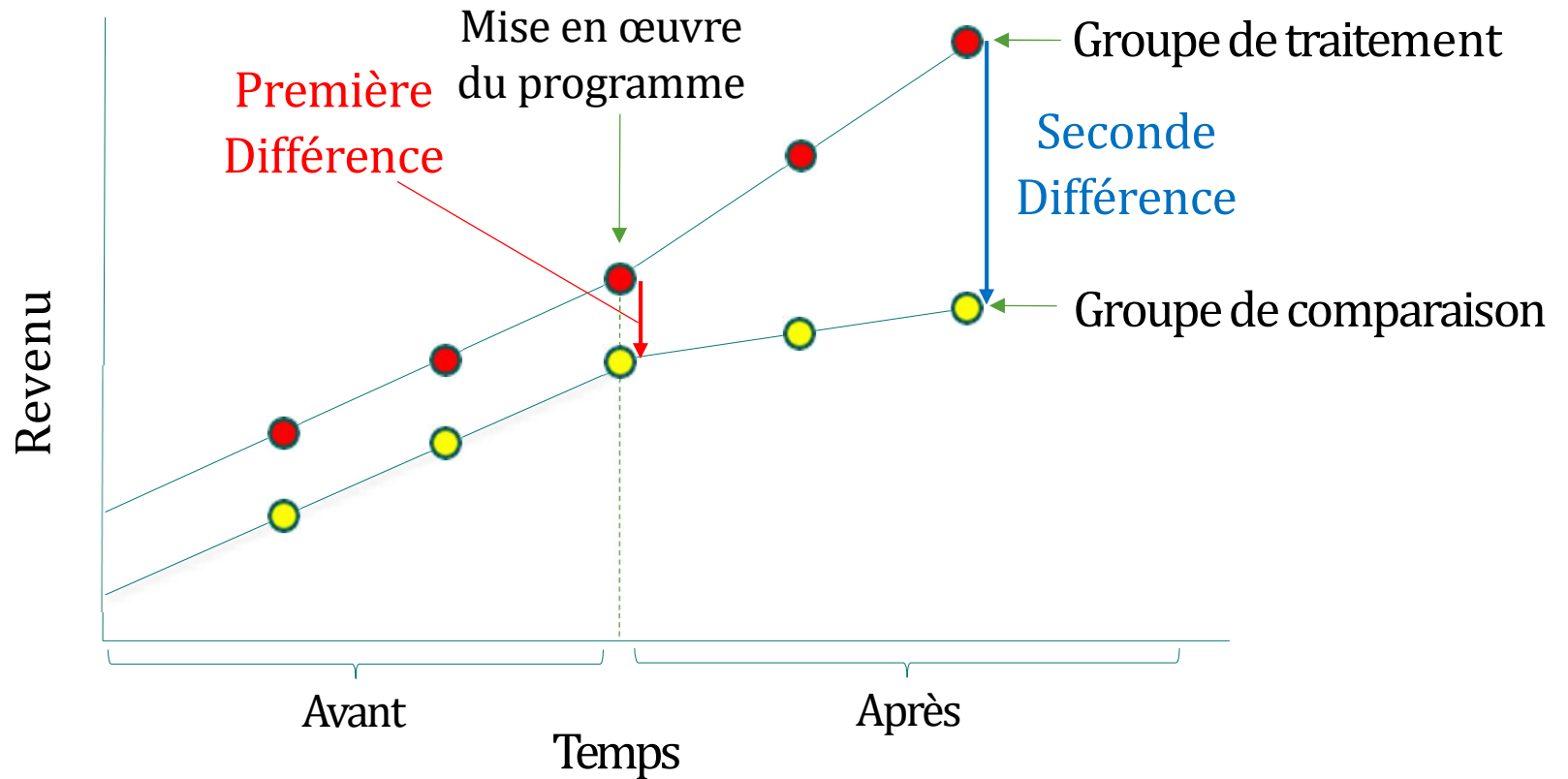
# Double Différence



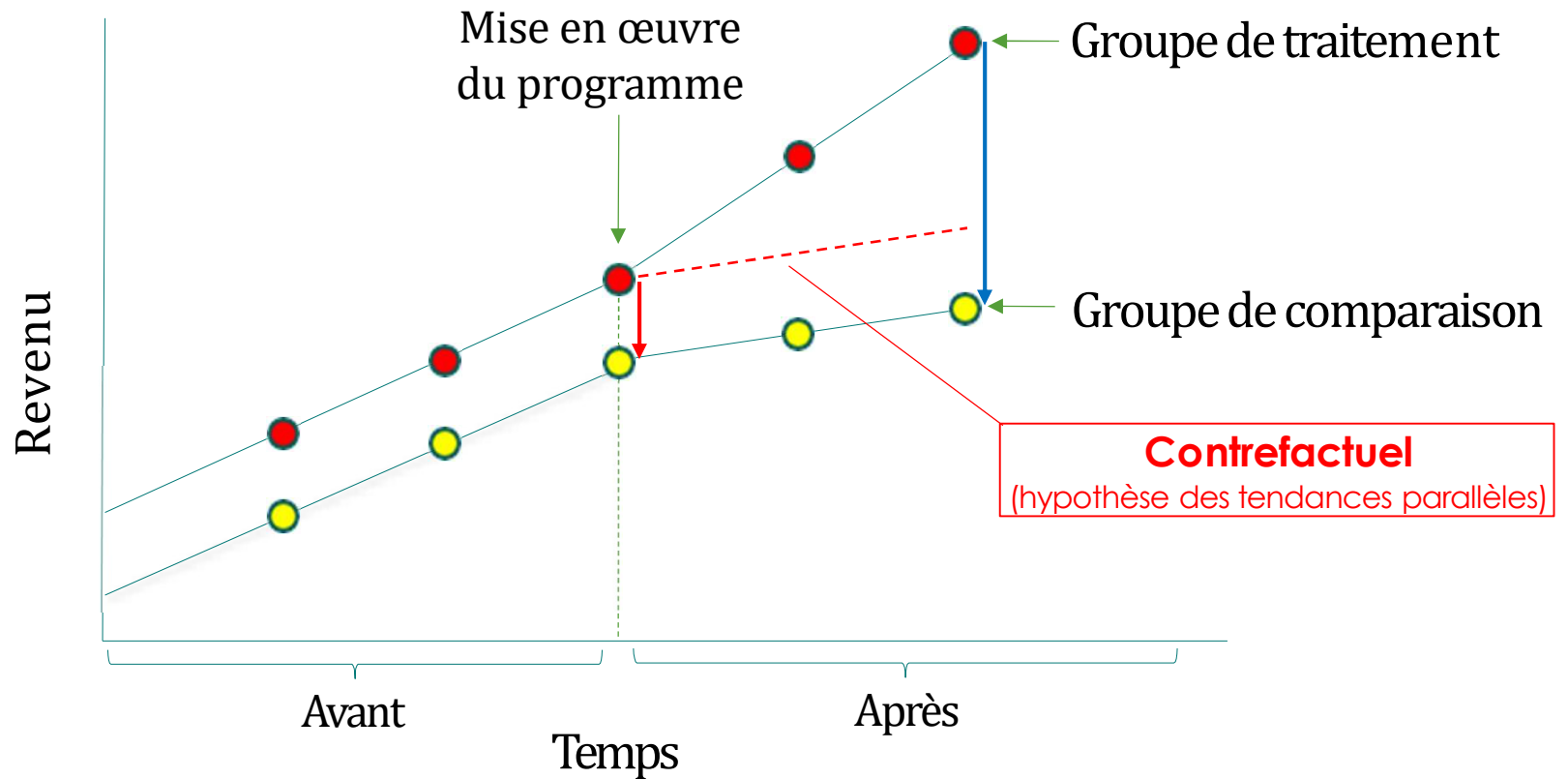
# Double Différence



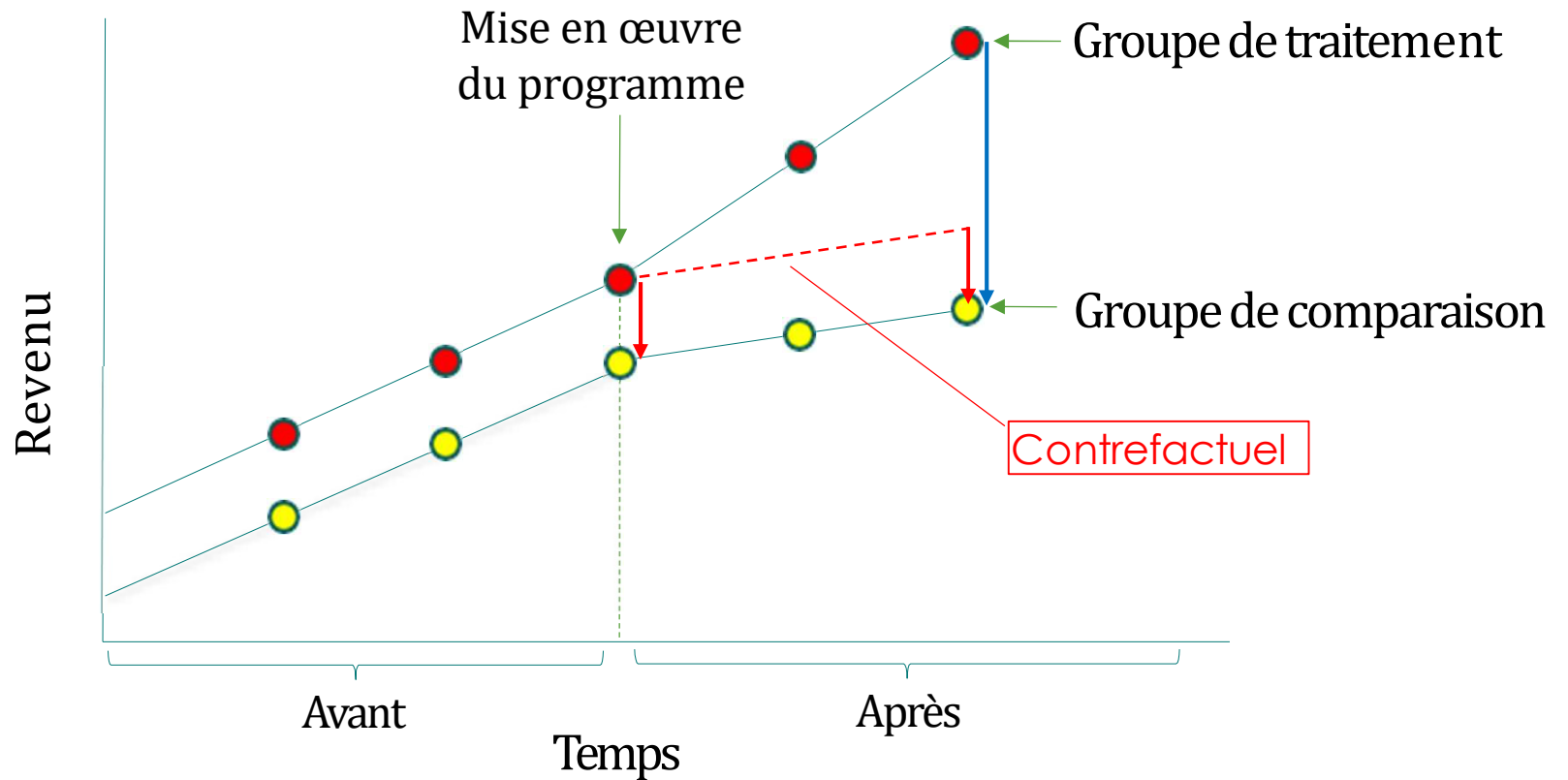
# Double Différence



# Double Différence

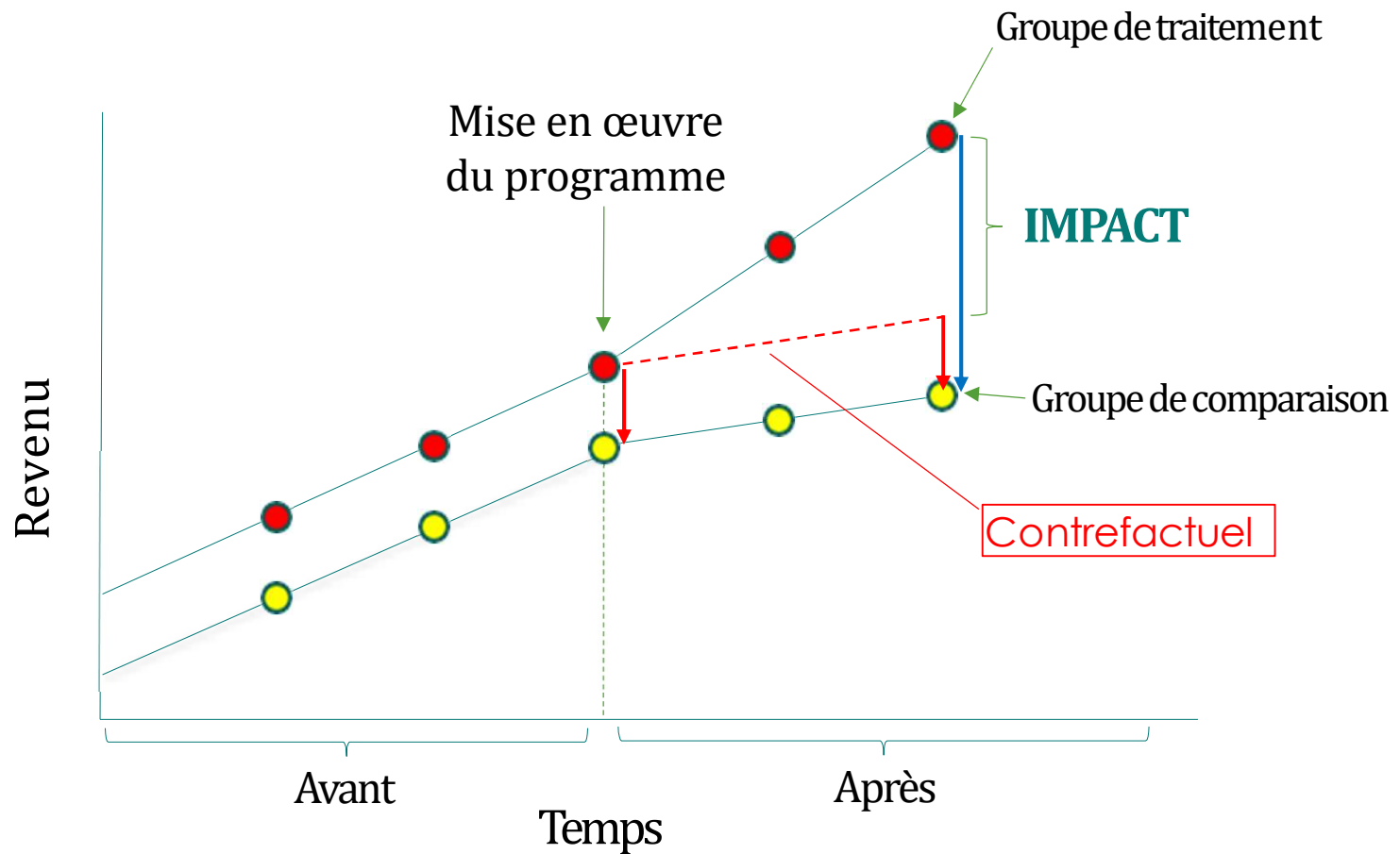


# Double Différence



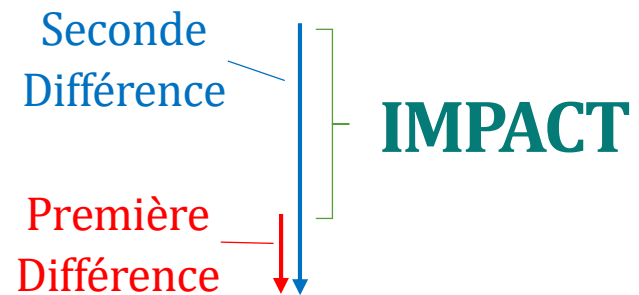


# Double Différence



# Double Différence

---



# Double Différence: Exemple

---

## Evaluation de PROGRESA (Schultz, 2004)

- Le premier programme de transfert monétaire conditionnel
- (3 ans de) Prime de scolarité accordée aux mères d'un enfant scolarisé éligibles en termes de pauvreté : conditionnée à la présence de l'enfant à l'école pendant 85 % des jours d'école
- Schultz s'appuie sur la Double Différence pour comparer la scolarisation des enfants entre les localités PROGRESA (localités de Traitement) et les localités non-PROGRESA (localités de Contrôle)

# Double Différence: Exemple

<b>Résultat= Années cumulées de scolarisation après la première année de scolarité</b>	<b>Avant le Programme</b>	<b>Après le programme</b>
Groupe de traitement (localités Progresa)	6.80	6.95
Groupe de comparaison (localités non Progresa)	6.66	6.14

Source: Results from Schultz (2004) – Table 7

# Double Différence: Exemple

Comparaison avant-après



<b>Résultat= Années cumulées de scolarisation après la première année de scolarité</b>	<b>Avant le Programme</b>	<b>Après le programme</b>	<b>Différence</b>
Groupe de traitement (localités Progresa)	6.80	6.95	0.15
Groupe de comparaison (localités non Progresa)	6.66	6.14	-0.52

Source: Results from Schultz (2004) – Table 7

# Double Différence: Exemple

Comparaison avant-après



Résultat= Années cumulées de scolarisation après la première année de scolarité	Avant le Programme	Après le programme	Différence
Groupe de traitement (localités Progresas)	6.80	6.95	0.15
Groupe de comparaison (localités non Progresas)	6.66	6.14	-0.52
<i>Différence</i>	0.14	0.81	

Simple différence



# Double Différence: Exemple

Comparaison avant-après



Résultat= Années cumulées de scolarisation après la première année de scolarité	Avant le Programme	Après le programme	Différence
Groupe de traitement (localités Progresas)	6.80	6.95	0.15
Groupe de comparaison (localités non Progresas)	6.66	6.14	-0.52
<i>Différence</i>	0.14	0.81	0.67

Simple différence



Double  
Différence

Source: Results from Schultz (2004) – Table 7

# Double Différence

---

- **Hypothèses:**

- Tendances d'évolution parallèles → prend en compte les différences non observées qui sont constantes au cours du temps  
→ Pas de différences changeant au cours du temps, par ex. à cause d'un choc/événement extrême

- **Difficultés:**

- Prouver que les deux groupes auraient connu la même évolution en l'absence du programme.
- Il est nécessaire de démontrer que ces hypothèses sont valides, ce qui peut être très difficile à faire sans preuve réelle.
- Le groupe de comparaison est-il vraiment comparable au groupe de traitement ? Sur l'ensemble des caractéristiques pertinentes ?



# Méthodes non expérimentales

1. Double  
Différence

2. Appariement  
("Matching")

Méthodes non  
expérimentales

3. Variables  
Instrumentales  
(VI)

4. Régression sur  
Discontinuité  
(RSD)

# Appariement (« Matching »)

---

- Programme : programme de formation professionnelle destiné aux jeunes
- Groupe de traitement : le groupe des jeunes inscrits
- Groupe de comparaison : **groupe de jeunes non inscrits**

## **Idée de l'appariement (« matching »):**

- Associer chaque individu du groupe T avec un individu du groupe de comparaison (former une paire)
- Trouver de bonnes paires nécessite une approximation aussi bonne que possible des caractéristiques qui expliquent la décision des jeunes de s'inscrire dans le programme
- Estimer l'impact par la différence des résultats observées entre les groupes T et C une fois appariés

# Appariement (« Matching »)

---

- Programme : programme de formation professionnelle destiné aux jeunes
- Groupe de traitement : le groupe des jeunes inscrits
- Groupe de comparaison : **groupe de jeunes non inscrits**

Méthode possible: **Appariement exact**

→ l'unité de traitement et l'unité de comparaison doivent avoir ***exactement*** les mêmes caractéristiques

# Appariement (« Matching »)

- Programme : programme de formation professionnelle destiné aux jeunes
- Groupe de traitement : le groupe des jeunes inscrits
- Groupe de comparaison : **groupe de jeunes non inscrits**

## Appariement exact

Unités du groupe de traitement				Unités du groupe de comparaison			
Age	Genre	Mois sans emploi	Diplômé du secondaire	Age	Genre	Mois sans emploi	Diplômé du secondaire
19	1	3	0	24	1	8	1
35	1	12	1	38	0	1	0
41	0	17	1	58	1	7	1
23	1	6	0	21	0	2	1
55	0	21	1	34	1	20	0
27	0	4	1	41	0	17	1
24	1	8	1	46	0	9	0
46	0	3	0	41	0	11	1
33	0	12	1	19	1	3	0
40	1	2	0	27	0	4	0

Source: Gertler et al. (2016)

# Appariement (« Matching »)

- Programme : programme de formation professionnelle destiné aux jeunes
- Groupe de traitement : le groupe des jeunes inscrits
- Groupe de comparaison : **groupe de jeunes non inscrits**

## Appariement exact

Unités du groupe de traitement				Unités du groupe de comparaison			
Age	Genre	Mois sans emploi	Diplômé du secondaire	Age	Genre	Mois sans emploi	Diplômé du secondaire
19	1	3	0	24	1	8	1
35	1	12	1	38	0	1	0
41	0	17	1	58	1	7	1
23	1	6	0	21	0	2	1
55	0	21	1	34	1	20	0
27	0	4	1	41	0	17	1
24	1	8	1	46	0	9	0
46	0	3	0	41	0	11	1
33	0	12	1	19	1	3	0
40	1	2	0	27	0	4	0

Source: Gertler et al. (2016)

# Appariement (« Matching »)

- Programme : programme de formation professionnelle destiné aux jeunes
- Groupe de traitement : le groupe des jeunes inscrits
- Groupe de comparaison : **groupe de jeunes non inscrits**

## Appariement exact

L'appariement basé sur le score de propension peut servir si l'appariement exact n'est pas possible

Unités du groupe de traitement				Unités du groupe de comparaison			
Age	Genre	Mois sans emploi	Diplômé du secondaire	Age	Genre	Mois sans emploi	Diplômé du secondaire
19	1	3	0	24	1	8	1
35	1	12	1	38	0	1	0
41	0	17	1	58	1	7	1
23	1	6	0	21	0	2	1
55	0	21	1	34	1	20	0
27	0	4	1	41	0	17	1
24	1	8	1	46	0	9	0
46	0	3	0	41	0	11	1
33	0	12	1	19	1	3	0
40	1	2	0	27	0	4	0

Source: Gertler et al. (2016)

# Appariement basé sur le score de propension (Propensity Score Matching - PSM)

---

Idée:

- On estime la **probabilité qu'un individu participe au programme (le « score de propension »)** en fonction d'un **ensemble de caractéristiques observées**
- On apparie ensuite les groupes de traitement et de contrôle avec des valeurs de score de propension similaires

# Appariement basé sur le score de propension (Propensity Score Matching - PSM)

---

## Hypothèses:

- Excellente compréhension des caractéristiques observables qui influencent la participation à une intervention
- Résoud seulement le biais de sélection *observable* → appariement fait seulement sur des caractéristiques observables (suppose qu'il n'y a pas de différences non observables !)
- Socle commun (nécessite que la plupart des scores de propension des participants et non participants se trouvent dans le même intervalle)

## Difficultés:

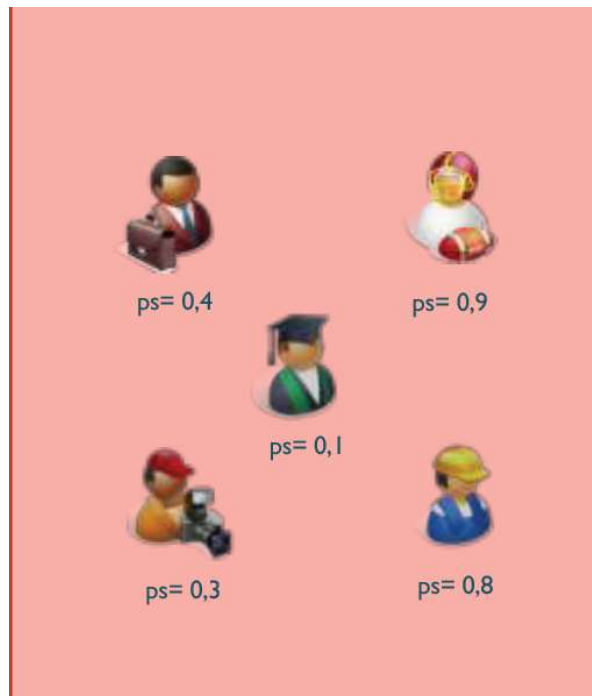
- Nécessite un grand nombre d'unités dans le groupe de comparaison
- Le manque de «socle commun» peut affecter la validité externe des résultats



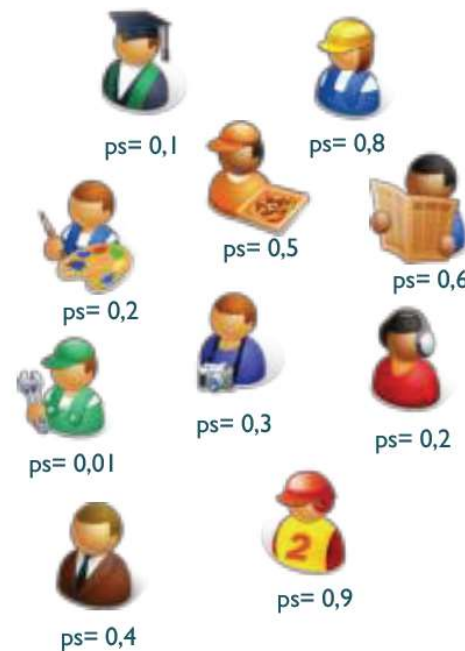
# Appariement basé sur le score de propension (Propensity Score Matching - PSM)

- De nombreux algorithmes d'appariement existent
- Exemple de matching 1-pour-1:

Participants



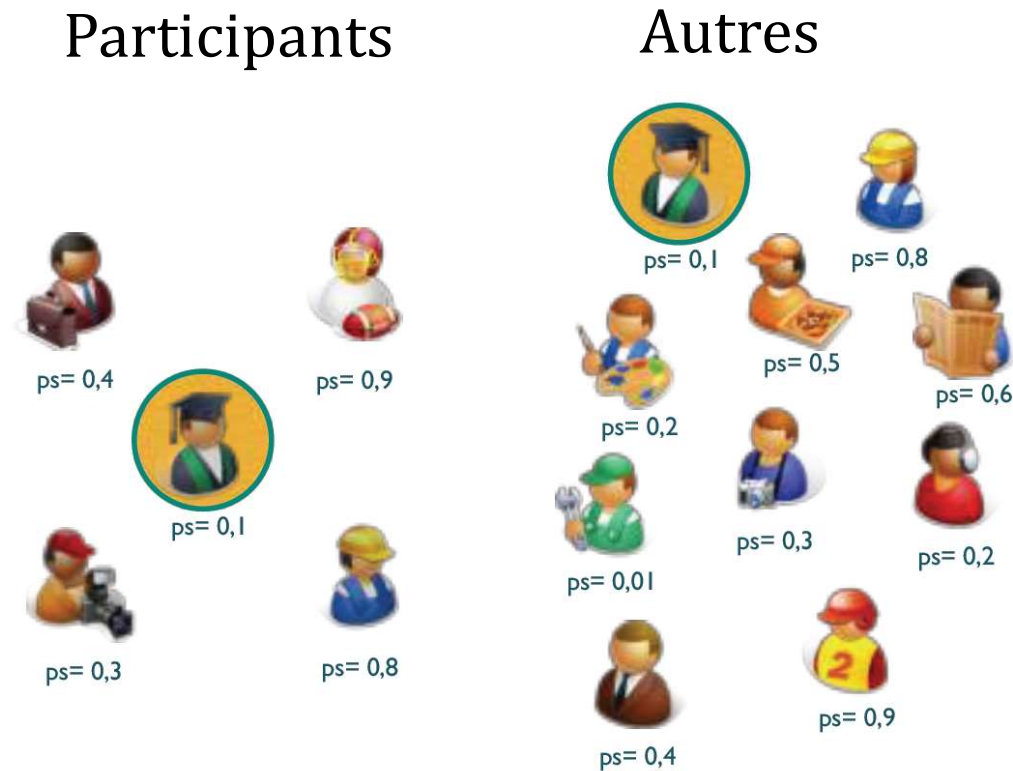
Autres



Source: Image adjusted from Trycinski (2011)

# Appariement basé sur le score de propension (Propensity Score Matching - PSM)

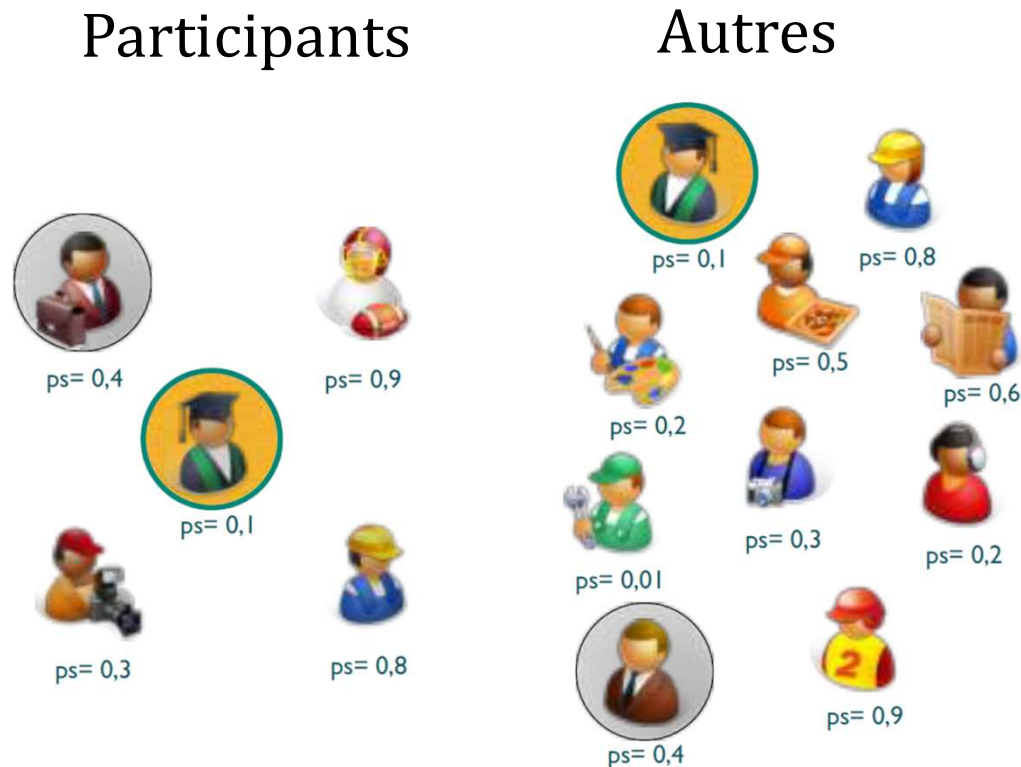
- De nombreux algorithmes d'appariement existent
- Exemple de matching 1-pour-1:



Source: Image adjusted from Trycinski (2011)

# Appariement basé sur le score de propension (Propensity Score Matching - PSM)

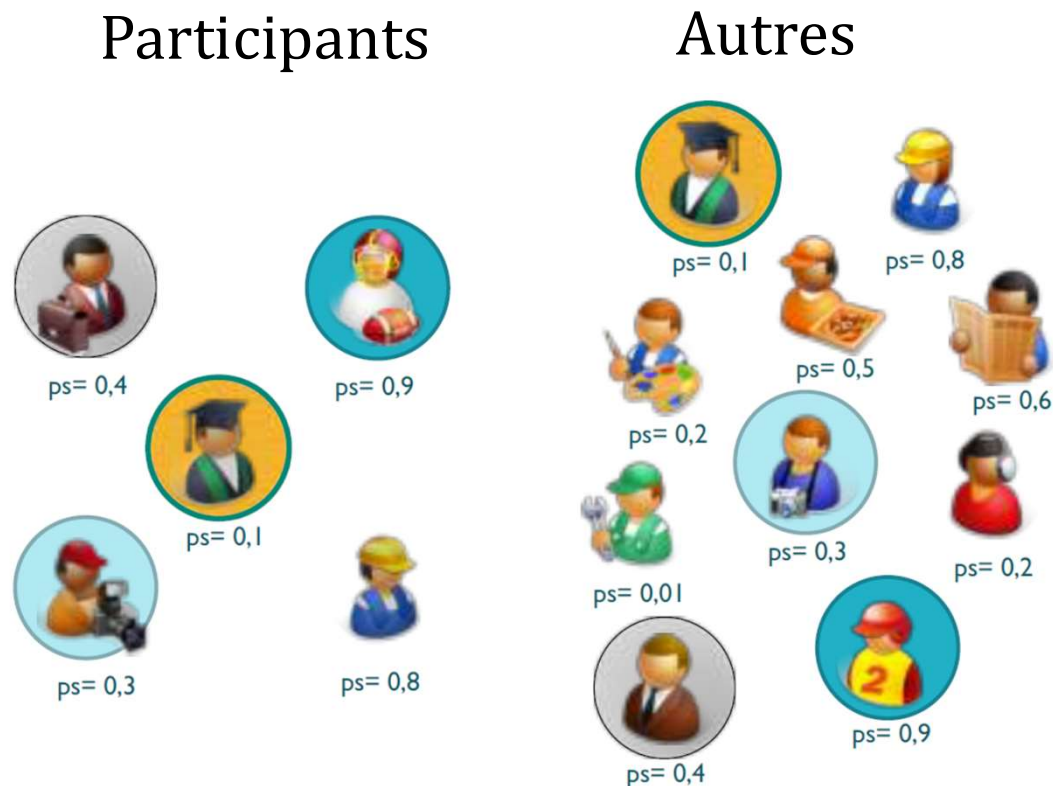
- De nombreux algorithmes d'appariement existent
- Exemple de matching 1-pour-1:



Source: Image adjusted from Trycinski (2011)

# Appariement basé sur le score de propension (Propensity Score Matching - PSM)

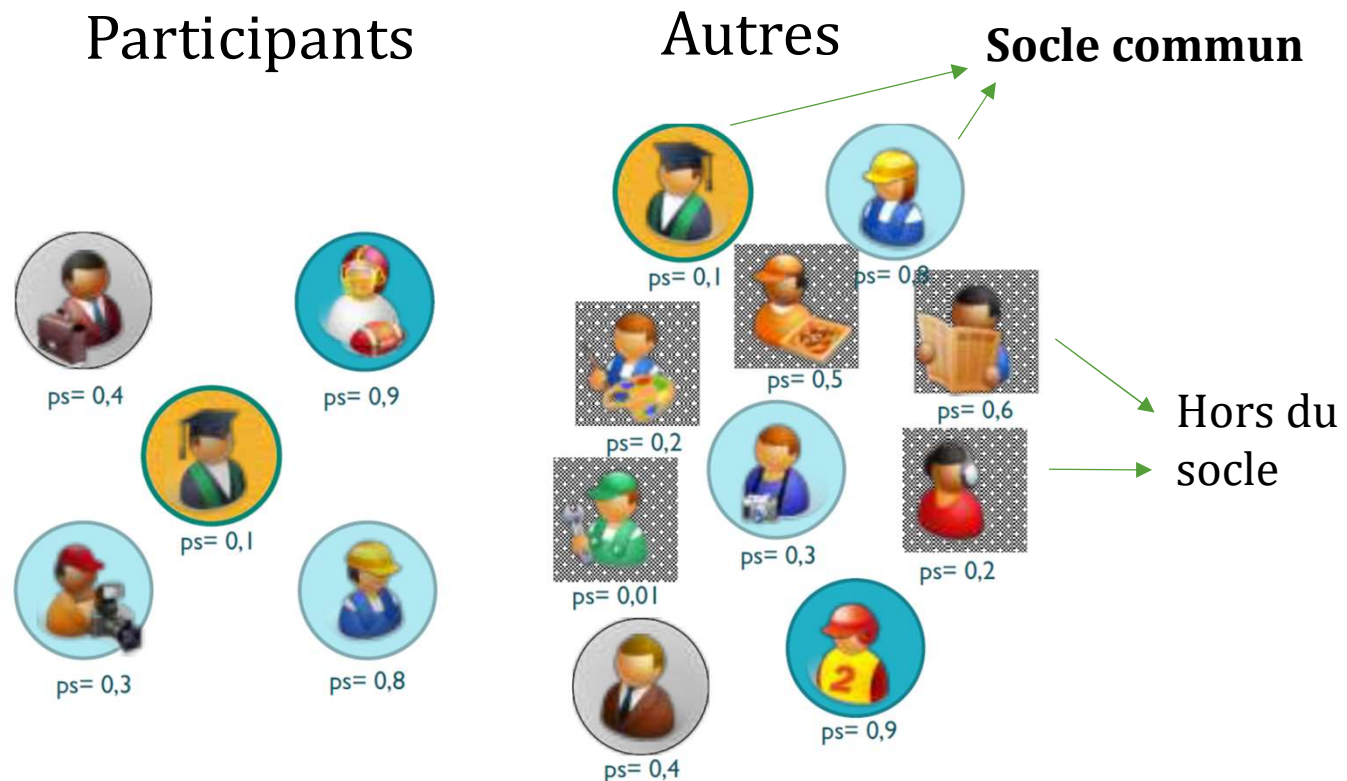
- De nombreux algorithmes d'appariement existent
- Exemple de matching 1-pour-1:



Source: Image adjusted from Trycinski (2011)

# Appariement basé sur le score de propension (Propensity Score Matching - PSM)

- De nombreux algorithmes d'appariement existent
- Exemple de matching 1-pour-1:



Source: Image adjusted from Trycinski (2011)

# Appariement basé sur le score de propension (PSM): Exemple

---

Évaluation du programme intégré de sécurité alimentaire (IFSP) en Éthiopie (Abebaw, Fentie et Kassa, 2010)

- **Programme :** réhabilitation de l'environnement, promotion de l'agriculture et de l'élevage, construction et entretien des infrastructures (dans la région d'Ahmara)
- **Problème:** placement non aléatoire du programme sur la base de critères de vulnérabilité des *kebele* (villages) et des ménages
  - Les ménages du groupe de traitement sont extrêmement différents des ménages du groupe de comparaison, à l'intérieur d'un même *kebele* ainsi qu'entre *kebele*

# Appariement basé sur le score de propension (PSM): Exemple

## Estimation du score de propension

Variables	Specification (3) Coefficient (std. error)
SEX	0.239 (0.606)
PAGE	0.015 (0.0175)
PEDU	0.201 (0.368)
PFAMNO	0.053 (0.115)
PLAND	-0.261* (0.147)
PTLU	-0.069 (0.057)
PHOUSTAT	1.688*** (0.539)
PHHDURAB	-0.015*** (0.004)
PMKTDIST	-0.084** (0.039)
PEXTDIST	-0.147* (0.075)
PFAMNO squared	
PAGE squared	
PLAND squared	
PTLU squared	
PHHDURAB squared	
PMKTDIST squared	
PEXTDIST squared	
Constant	1.735* (0.967)
Sample size (N)	200
Pseudo-R <sup>2</sup>	0.19
LR $\chi^2$ value	34.78***
Log-likelihood	-111.82

\*\*\*, \*\* and \* stand for significance at the 1%, 5% and 10% levels, respectively.

- Variable dépendante = 1 si le ménage participe au programme
- Variables d'appariement mesurées *avant* le programme :
  - Sexe du chef de ménage
  - Âge du chef de ménage
  - Niveau d'éducation du chef de ménage
  - Nombre de membres au sein du ménage
  - Possession de terres
  - Détention de bétail
  - Type de logement
  - Valeur des biens durables
  - Distance par rapport au marché
  - Distance par rapport au bureau de l'agent de développement



# Appariement basé sur le score de propension (PSM): Exemple

## Estimation du score de propension

Variables	Specification (3) Coefficient (std. error)
SEX	0.239 (0.606)
PAGE	0.015 (0.0175)
PEDU	0.201 (0.368)
PFAMNO	0.053 (0.115)
PLAND	-0.261* (0.147)
PTLU	-0.069 (0.057)
PHOUSTAT	1.688*** (0.539)
PHHDURAB	-0.015*** (0.004)
PMKTDIST	-0.084** (0.039)
PEXTDIST	-0.147* (0.075)
PFAMNO squared	
PAGE squared	
PLAND squared	
PTLU squared	
PHHDURAB squared	
PMKTDIST squared	
PEXTDIST squared	
Constant	1.735* (0.967)
Sample size (N)	200
Pseudo-R <sup>2</sup>	0.19
LR $\chi^2$ value	34.78***
Log-likelihood	-111.82

\*\*\*, \*\* and \* stand for significance at the 1%, 5% and 10% levels, respectively.

- Variable dépendante = 1 si le ménage participe au programme

- Variables d'appariement

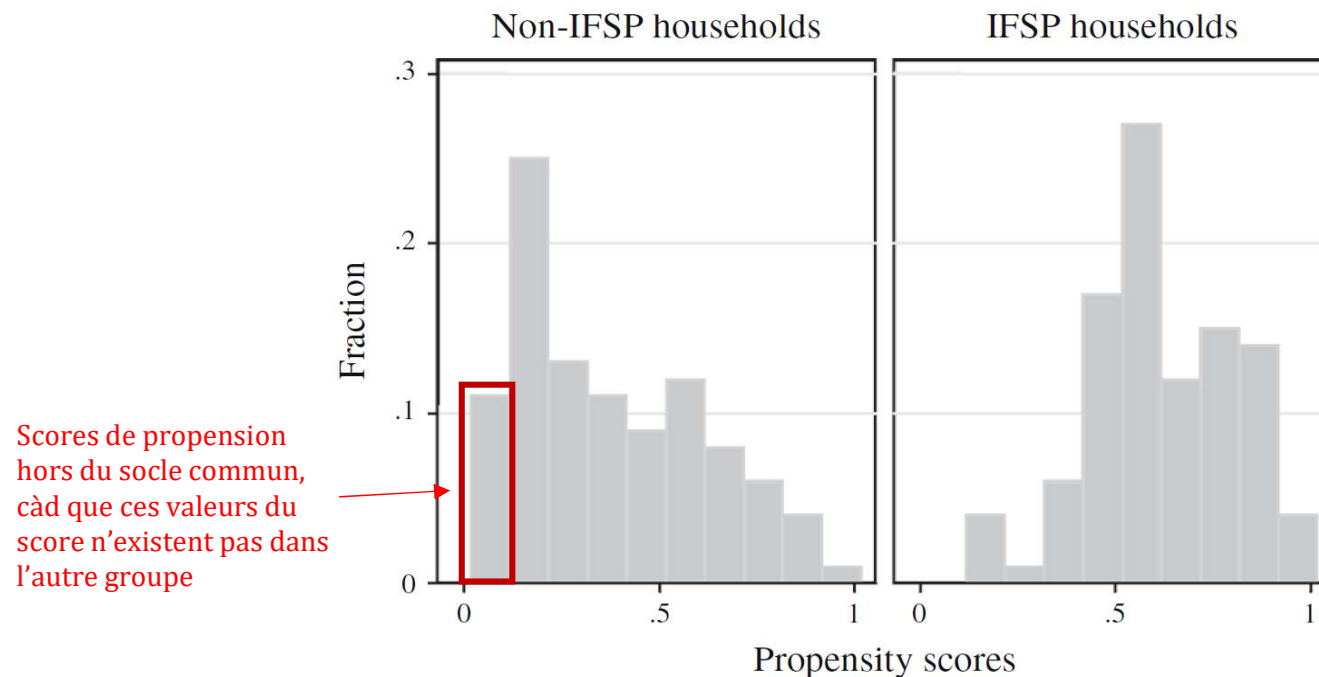
**IMPORTANT** : l'appariement doit se faire sur la base de caractéristiques mesurées *avant* le programme OU qui *ne varient pas dans le temps*, afin d'éviter d'introduire un biais dans l'analyse (la participation au programme affecte les comportements !)



# Appariement basé sur le score de propension (PSM): Exemple

## Évaluation du socle commun

Avant l'appariement



Source: Abebaw, Fentie and Kassa (2010)

- Scores entre 0,02 et 0,95 dans le groupe de comparaison (non IFSP)
  - Scores entre 0,17 et 0,96 dans le groupe de traitement (IFSP)
- exclusion de 23 ménages

# Appariement basé sur le score de propension (PSM): Exemple

## Estimations de l'impact

	Specification (3)
Household food calorie intake	694.96*** (4.77)
Balancing property satisfied	Yes
Common support imposed	Yes
Number of observations	
IFSP households	99
Non-IFSP households	79

\*\*\* indique une significativité de niveau 1%. Entre parenthèses les erreurs standard basées sur le bootstrap (50 répétitions).

Source: Abebaw, Fentie and Kassa (2010)

- De nombreux algorithmes d'appariement sont disponibles
  - Algorithme « kernel »: chaque unité traitée est appariée à toutes les unités de comparaison dans un intervalle donné, pondérée par la proximité de la valeur du score de propension (→ plus l'unité de comparaison a un score proche, plus son poids est important)
- L'impact est mesuré par la différence de moyenne entre les unités appariées

Résultat : + 696 calories par jour par unité d'équivalent adulte, soit environ 30 % de plus que dans le groupe de comparaison

# Méthodes non expérimentales

---

1. Double  
Différence

2. Appariement  
("Matching")

Méthodes non  
expérimentales

3. Variables  
Instrumentales  
(VI)

4. Régression sur  
Discontinuité  
(RSD)

## Variable(s) Instrumentale(s) (VI)

---

- La VI ne crée pas de groupe de comparaison mais elle se sert d'un modèle de **régression statistique** pour estimer l'impact d'une intervention (à partir de données transversales ou de panel)
- La VI contrecarre les biais de sélection, notamment en ce qui concerne **les caractéristiques non observables**
- Si ces caractéristiques non observables sont corrélées à la participation au programme ainsi qu'à la variable de résultat, les estimations de l'impact du programme seront **biaisées**

# Variable(s) Instrumentale(s) (VI)

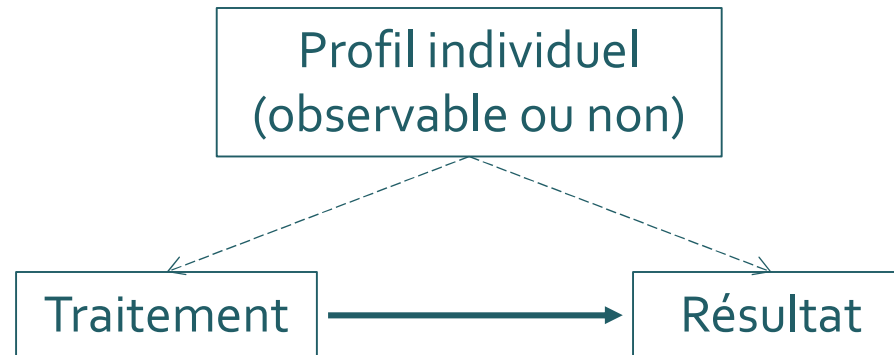
---

- Cette approche a recours à une **variable supplémentaire** (la VI) qui est **fortement corrélée avec la participation au programme**, mais qui **n'est pas corrélée avec les caractéristiques non observables qui affectent les résultats**
- On se sert de cette variable supplémentaire pour "apurer" la variable de traitement en séparant et en éliminant la partie qui est affectée par des caractéristiques non observables

# Variable Instrumentale (IV)

## Intuition

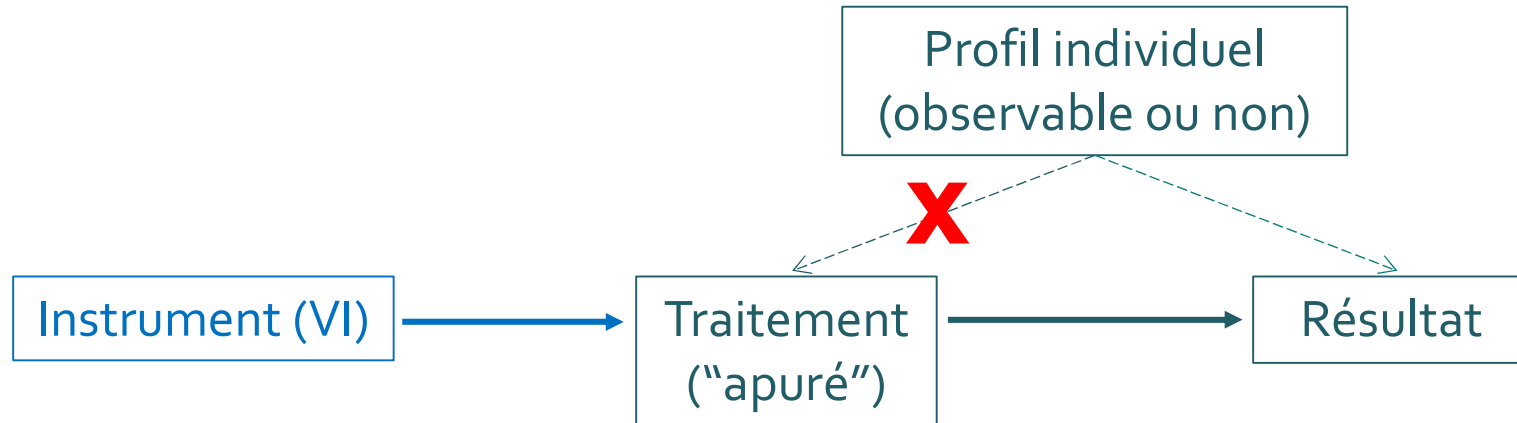
---



### Intuition:

- Certains facteurs influencent à la fois la participation au programme **et** le résultat.
- Si on ne les prend pas en compte, on ne peut pas savoir si la relation qui existe entre la participation au programme et le résultat est causale, ou bien si elle est influencée (toute ou en partie) par lesdits facteurs.

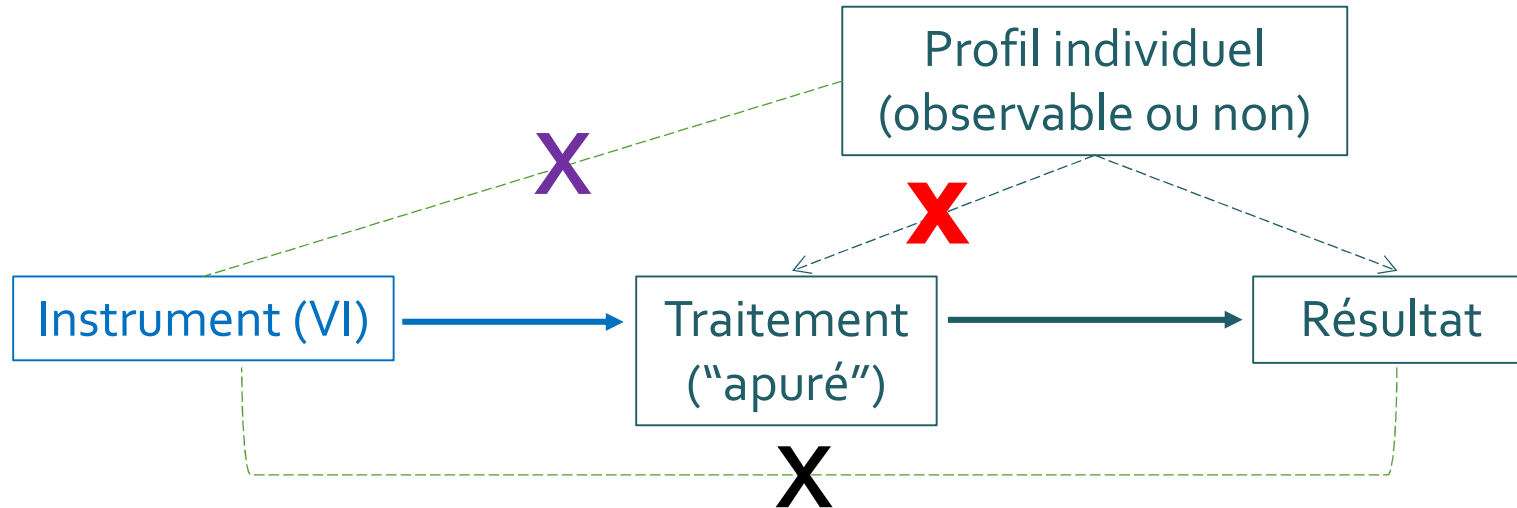
# Variable Instrumentale (IV) Intuition



## Intuition:

- La VI est corrélée avec le résultat ***seulement à travers*** la participation au programme
- Elle est utilisée pour “apurer” la variable de traitement en enlevant les corrélations problématiques (croix rouge sur la figure), ce qui permet d’isoler la relation causale entre la participation au programme et le résultat

# Variable Instrumentale (IV) Intuition



## Conditions nécessaires:

- 1) L'instrument affecte la probabilité de participer au programme (« **Pertinence** » - flèche bleue sur la figure)
- 2) L'instrument n'affecte pas le résultat à travers des canaux autres que via la participation au programme (« **Exclusion** » - la VI n'est pas corrélée directement aux résultats [**X**], ni aux facteurs non observables [**X**])



# Variable Instrumentale (IV)

---

## **Hypothèse principales:**

- VI fortement corrélée avec le traitement réel
- VI n'affecte la variable de résultat qu'à travers le traitement

## **Difficulté:**

- Trouver une VI est très difficile

## **Exemples de VI:**

- Allocation aléatoire du programme
- Allocation aléatoire des incitations (approche par « encouragement »)

## VI - Exemple

---

### Evaluation de l'effet d'une formation pour les entreprises sur les profits de microentreprises (Berge et al., 2011)

- Cible: clients d'un programme de microcrédit → formation sur l'entrepreneuriat
  - Méthode: allocation aléatoire des groupes d'emprunteurs (de microcrédit) aux sessions de formation
  - Participation réelle au programme → 70% de présence aux formations
- Conformité partielle à l'allocation aléatoire (« partial compliance »)

## VI - Exemple

---

- L'allocation aléatoire peut être utilisée comme instrument pour la participation réelle au programme
  - ✓ Forte corrélation avec la participation aux formations
  - ✓ VI présumée affecter les profits seulement via la formation et ne pas être corrélée avec les facteurs non observables qui influencent la non-conformité (« non compliance »)
- Important d'enregistrer la présence des participants!
- L'approche VI mesure l'impact pour les « conformistes » (**compliers**), c-à-d ces clients de microcrédit qui participent à la formation à cause de l'allocation au programme, mais n'auraient pas participé autrement

## VI - Exemple

Nomenclature des groupes d'étude dans l'approche VI

		Assigné au <b>groupe de contrôle</b>	
		0 = Ne participe pas au programme	1 = Participe au programme
Assigné au <b>groupe de traitement</b>	0 = Ne participe pas au programme	Jamais preneurs ("never takers")	Provocateurs ("defiers")
	1 = Participe au programme	Conformistes ("compliers")	Toujours preneurs ("always takers")

L'approche VI mesure  
l'impact pour ce groupe

## VI - Exemple

### Estimations des impacts sur les profits et le CA

	(1) Profit Margin ITT	(2) Profit Margin ATET	(3) Sales ITT	(4) Sales ATET
Training	-0.014 (0.028)	-0.015 (0.031)	0.257** (0.123)	0.288** (0.137)
Grant	-0.004 (0.016)	-0.004 (0.015)	0.038 (0.073)	0.036 (0.072)
Training*Female	0.003 (0.033)	0.003 (0.037)	-0.262* (0.157)	-0.295* (0.177)
Female	-0.013 (0.024)	-0.013 (0.024)	0.044 (0.110)	0.044 (0.109)
Sum Female	-0.010 (0.018)	-0.012 (0.021)	-0.006 (0.089)	-0.007 (0.103)
Observations	494	494	494	494

*Note: The table reports regressions of profit margin and sales on treatment status, all regressions controlling for gender and covariates. Covariates include age, gender, education, number of businesses, PRIDE branch, PRIDE loan size, marketing index, religion, and the lagged dependent variable. Cluster-robust standard errors in parentheses; \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .*

Source: Berge et al. (2011)

Impact de l'exposition au programme (impact moyen sur les "compliers" et "non-compliers") aka "Intention de Traitement" ("Intent To Treat")

Impact de la participation effective au programme pour les "compliers" (par VI), aka "Effet Moyen du Traitement sur les Traités" ("Average Treatment Effect on the Treated")

# Méthodes non expérimentales

---

1. Double  
Différence

2. Appariement  
("Matching")

Méthodes non  
expérimentales

3. Variables  
Instrumentales  
(VI)

4. Régression sur  
Discontinuité  
(RSD)

# Régression sur Discontinuité (RSD)

---

- La participation au programme dépend parfois d'une règle transparente avec des critères basés sur un seuil limite clair, par ex.:
  - Admission à l'université avec un score  $>$  au 80<sup>ème</sup> percentile
  - Lois qui s'appliquent aux entreprises avec plus de 50 employés
  - Transfert d'argent pour les ménages avec un score de pauvreté  $< 50\%$
  - Un côté donné d'une frontière géographique

# Régression sur Discontinuité (RSD)

---

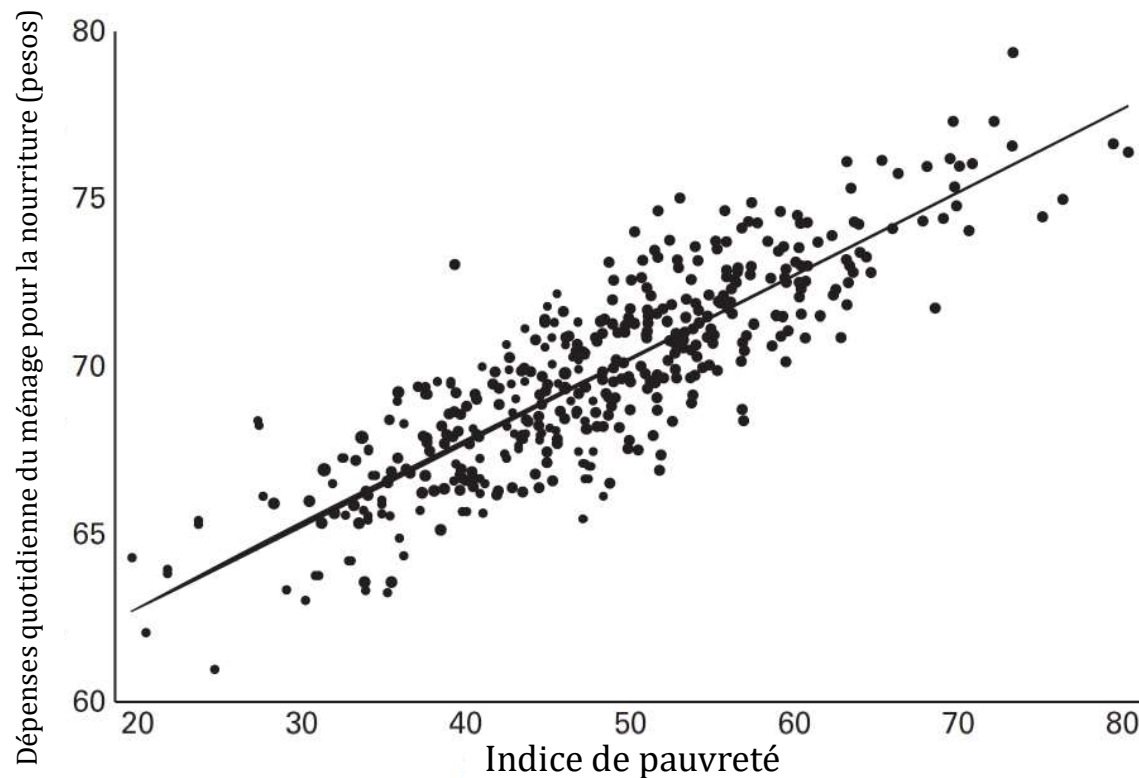
- Les individus avec un score/classement similaire sont comparables en termes de caractéristiques observables et non observables
- Mesure des impacts autour du seuil limite seulement, par ex.:  
Juste au-dessus du seuil limite (pas de programme = groupe de comparaison) vs. Juste au-dessous du seuil limite (programme = groupe de traitement)



# Régression sur Discontinuité (RSD)

- Programme: Transfert monétaire aux ménages
- Règle d'éligibilité: ménages avec un score de pauvreté < 50% (ici le score de pauvreté s'appelle la *variable d'assignation* - « running variable »)

*Dépenses du ménage et pauvreté (Avant l'intervention)*

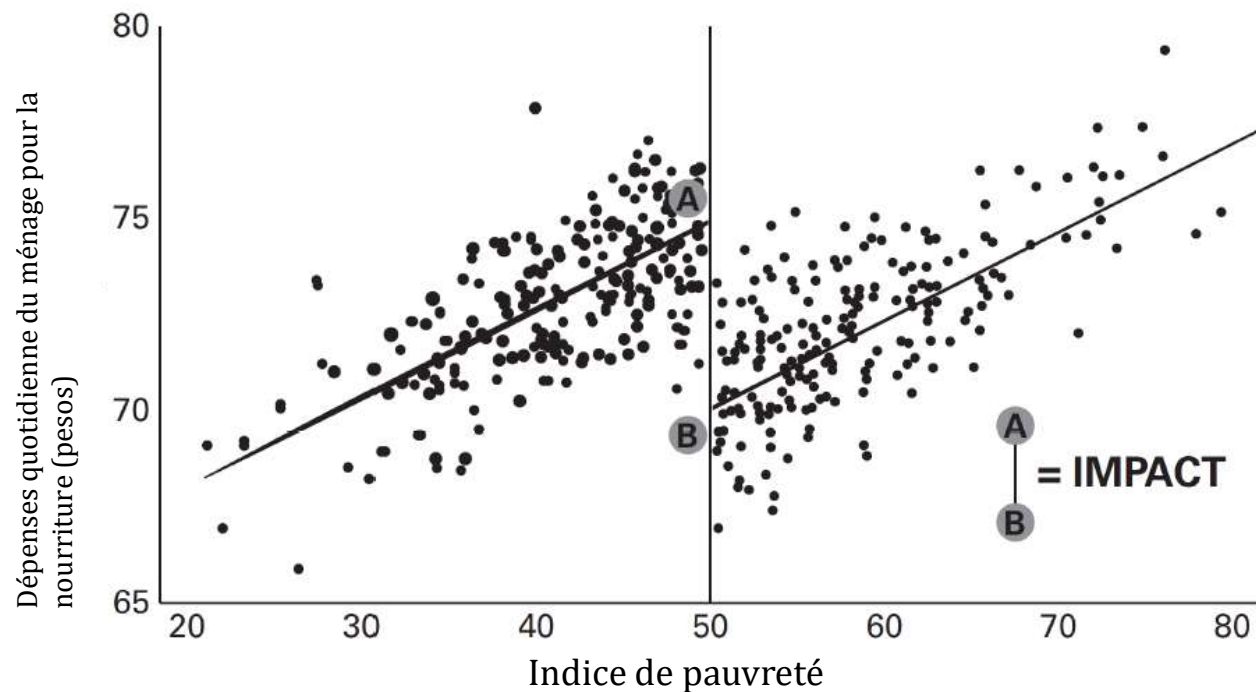


Source: Gertler et al., 2011

# Régression sur Discontinuité (RSD)

- Programme: Transfert monétaire aux ménages
- Règle d'éligibilité: ménages avec un score de pauvreté < 50% (ici le score de pauvreté s'appelle la *variable d'assignation* - « running variable »)

*Dépenses du ménage et pauvreté (Après l'intervention)*

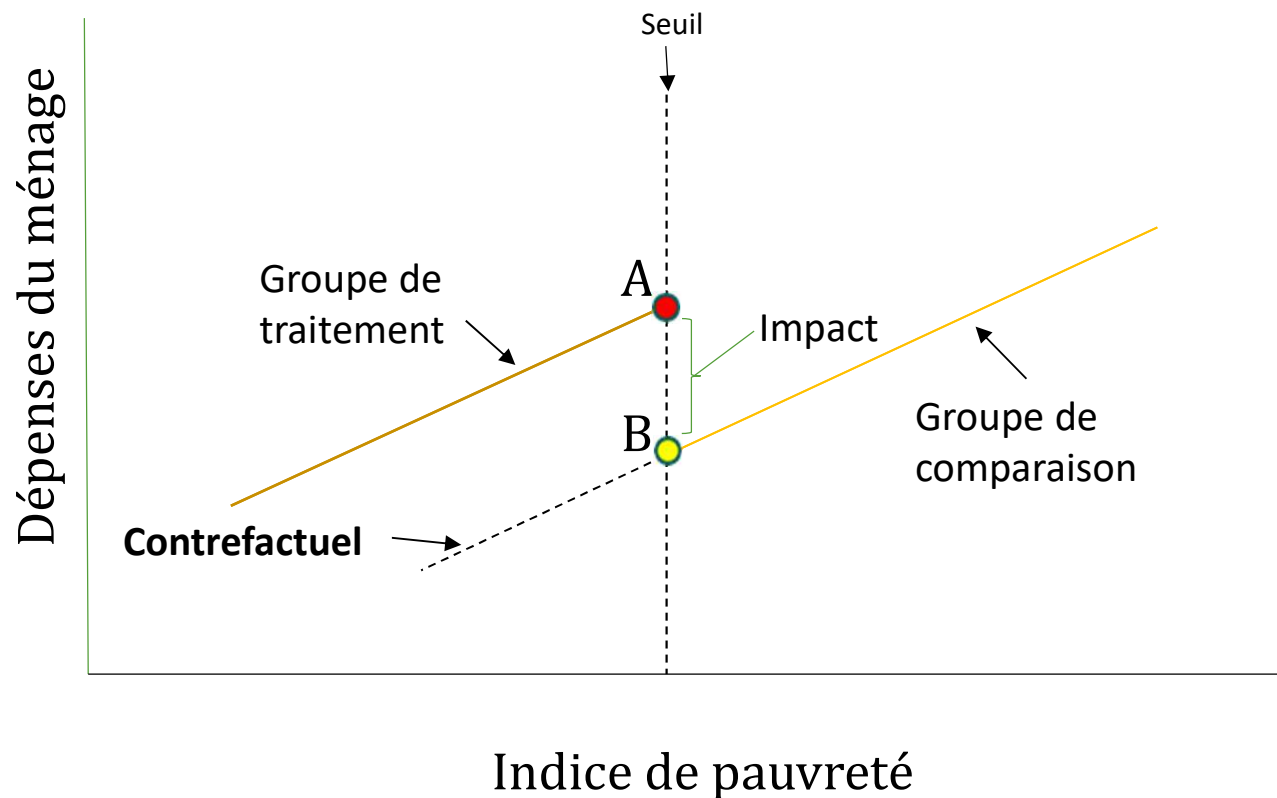


Source: Gertler et al., 2011

# Régression sur Discontinuité (RSD)

- Programme: Transfert monétaire aux ménages
- Règle d'éligibilité: ménages avec un score de pauvreté < 50% (ici le score de pauvreté s'appelle la *variable d'assignation* - « running variable »)

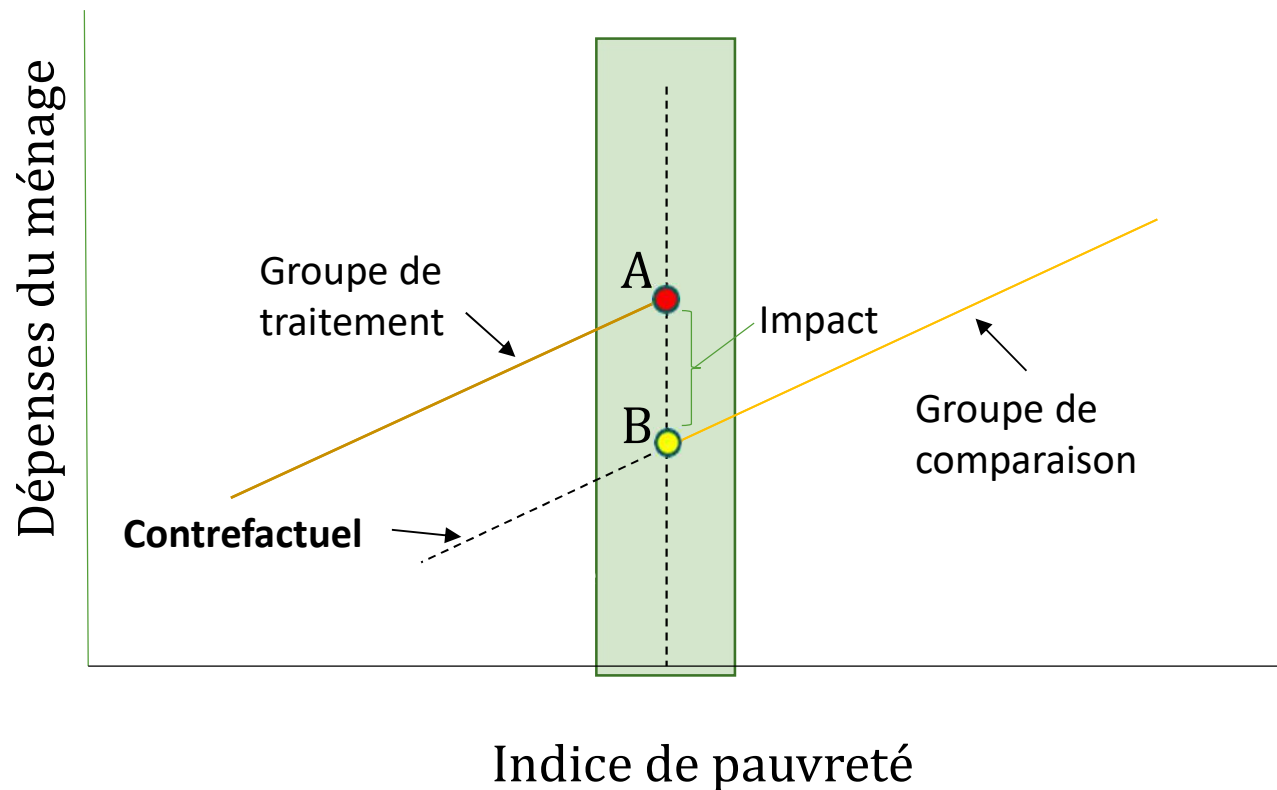
*Dépenses du ménage et pauvreté (Après l'intervention)*



# Régression sur Discontinuité (RSD)

- Programme: Transfert monétaire aux ménages
- Règle d'éligibilité: ménages avec un score de pauvreté < 50% (ici le score de pauvreté s'appelle la *variable d'assignation* - « running variable »)

*Dépenses du ménage et pauvreté (Après l'intervention)*



→ RSD mesure l'effet **local** du traitement autour du seuil

→ On peut comprendre la RSD comme un « ECR local » autour du seuil

# Régression sur Discontinuité (RSD)

---

- RSD « net »: Le traitement est déterminé par la variable d'assignation → **tous** les ménages éligibles participent au programme (et vice versa)
- RSD « flou »: Dans certains cas, la participation n'est pas complètement déterminée par la variable d'assignation. Le seuil limite génère une *discontinuité* dans la **probabilité** de participer
  - Par ex: les ménages en-dessous du seuil limite de pauvreté **ont plus de chance de participer** au programme, ménages au-dessus du seuil limite de pauvreté **ont moins de chance de participer** au programme  
→ Conformité partielle ! → On peut utiliser une VI !

# Régression sur Discontinuité (RSD)

---

## Hypothèses:

- Les participants potentiels ne peuvent manipuler précisément leur score
- Seule la variable de résultat a une discontinuité au seuil limite (pas les autres variables pertinentes)
- Il existe un nombre suffisant d'observations dans un intervalle donné autour du seuil
- L'intervalle autour du seuil peut être changé afin de tester la robustesse des résultats

# Régression sur Discontinuité (RSD)

---

- **Difficultés:**

- Il faut qu'aucun autre programme n'utilise le même seuil d'éligibilité
- Résultats valides seulement pour les unités autour du seuil: effet **local** moyen du traitement. On ne peut pas connaître l'effet du programme sur les ménages très pauvres (avec un score très inférieur au seuil)

## Quelle méthode requiert ABSOLUMENT d'avoir des données de référence sur la variable de résultat?

---

---

Appariement  
(« matching »)

---

ECR

---

Double Différence

---

Variable Instrumentale



## Quelle méthode requiert ABSOLUMENT d'avoir des données de référence sur la variable de résultat?

---

---

Appariement  
(« matching »)

---

ECR

---

Double Différence

---

Variable Instrumentale

## Récapitulatif: Méthodes non expérimentales

Méthode	Description	Hypothèses	Remarques
Double différence	Mesure le changement avant-après pour les deux groupes T et C. La différence des deux mesure le changement des résultats dans le groupe T relatif au groupe C. Prend en compte les différences entre T et C constantes au cours du temps.	<ul style="list-style-type: none"> <li>Tendances parallèles: Résultats pour groupes T et C auraient connu la même évolution en l'absence du programme</li> <li>Aucun choc non observé n'affecte les deux groupes différemment ni les résultats</li> </ul>	<ul style="list-style-type: none"> <li>Requiert données de référence sur la variable de résultat</li> <li>Preuve suggestive que l'hypothèse des tendances parallèles tient en se basant sur des données historiques (pré-programme)</li> </ul>
Matching	Apparie chaque participant du groupe T avec un ou plusieurs non-participants du groupe C. On compare ensuite les résultats entre les groupes T et C appariés.	<ul style="list-style-type: none"> <li>Aucune différence non observable entre groupes T et C corrélée avec les résultats et la participation au programme</li> <li>Socle commun suffisant pour la probabilité de participation (score de propension) entre les groupes T et C</li> </ul>	<ul style="list-style-type: none"> <li>Requiert une bonne compréhension de la sélection dans le programme</li> <li>Nécessite un grand groupe C</li> <li>Appariement fait sur des variables non affectées par le traitement (idéalement données de référence)</li> <li>Peut être combiné avec DD</li> </ul>
Variables Instrumentales (VI)	Résout le biais dû aux variables inobservables. VI ne crée pas de groupe de comparaison mais utilise une régression statistique pour estimer l'impact. La VI, corrélée avec le traitement mais pas avec les observables, est utilisée pour « apurer » la variable de participation.	<ul style="list-style-type: none"> <li>La VI est un prédicteur important de la participation au programme</li> <li>La VI affecte la variable de résultat seulement à travers le traitement</li> </ul>	<ul style="list-style-type: none"> <li>Trouver un VI valide est très difficile</li> <li>Les effets mesurés sont « locaux » : valides pour les « compliers », c'est-à-dire les individus influencés par l'instrument</li> </ul>
Regrression Sur Discontinuité (RSD)	Eligibilité dépend d'une règle claire avec un seuil prédéfini en fonction du classement des individus selon une variable d'assignation. Les résultats des individus juste en-dessous (par ex. groupe T) et juste au-dessus (par ex. groupe C) du seuil sont comparés.	<ul style="list-style-type: none"> <li>Aucun autre facteur que le programme ne crée la discontinuité dans la variable de résultat au seuil</li> <li>Un nombre suffisant d'observations autour du seuil</li> <li>Les participants ne peuvent pas manipuler leur éligibilité au programme</li> </ul>	<ul style="list-style-type: none"> <li>Grande taille d'échantillon nécessaire autour du seuil</li> <li>Effet mesuré est « local » : valide seulement autour du seuil</li> <li>Utiliser une VI en case de conformité partielle (RSD « flou »)</li> </ul>

# Exercice

- Un récent sondage national dans le pays Fantasia a révélé que le taux de chômage chez les femmes jeunes était de 65% et tend à augmenter.
- Le gouvernement a décidé d'engager des ONG locales pour mettre en place des formation professionnelles pour promouvoir l'employabilité et l'entrepreneuriat des femmes jeunes.
- Les districts les plus vulnérables seront ciblés dans une première phase (2022-2024), et la vulnérabilité est définie par un comité gouvernemental sur la base du niveau de pauvreté, le taux moyen de chômage, la sécurité alimentaire, et les taux d'émigration.
- Après la première phase, le programme sera peut-être mis à l'échelle au niveau national.
- Les femmes sans emploi entre 18 et 35 ans d'âge vivant dans les districts identifiés comme vulnérables seront éligibles au programme. Le programme n'a pas encore démarré.

Le gouvernement a besoin de votre aide: il désire en savoir plus sur l'impact du programme sur l'emploi et l'entrepreneuriat des femmes jeunes.

*Suggérez une méthode d'EI non expérimentale qui pourrait être utilisée.*

- *Décrivez comment vous choisirez les groupes de traitement et de comparaison.*
- *Quels sont les points forts et les limitations de cette méthode?*

## Exercice - Discussion

---

Groupe de traitement: jeune femmes (18-35) sans emploi dans les districts ciblés qui participent au programme

A) DD

Groupe de comparaison : jeune femmes (18-35) sans emploi dans les districts **non** ciblés

Groupe de comparaison alternatif: jeune femmes (18-35) sans emploi dans les districts ciblés ne participant pas au programme

- Hypothèses:
  - Tendances parallèles: Résultats pour groupes T et C auraient connu la même évolution en l'absence du programme
  - Aucun choc non observé n'affecte les deux groupes différemment ni les résultats
- Données: de référence (baseline) et finales (endline)

B) DD + Matching: Appariement peut être combiné avec la DD

## Exercice - Discussion

---

### C) Matching

Groupe de comparaison : jeune femmes (18-35) sans emploi dans les districts **non** ciblés

Groupe de comparaison alternatif: jeune femmes (18-35) sans emploi dans les districts ciblés ne participant pas au programme

- Hypothèses:
  - Aucune différence non observable entre groupes T et C corrélée avec les résultats et la participation au programme
  - Socle commun suffisant pour la probabilité de participation (score de propension) entre les groupes T et C
- Données : de référence (idéalement) + finales (endline)

### D) RSD ?

Y a-t-il une règle avec un seuil clair?



Center for Evaluation  
and Development

---

# GÉRER LES ATTENTES

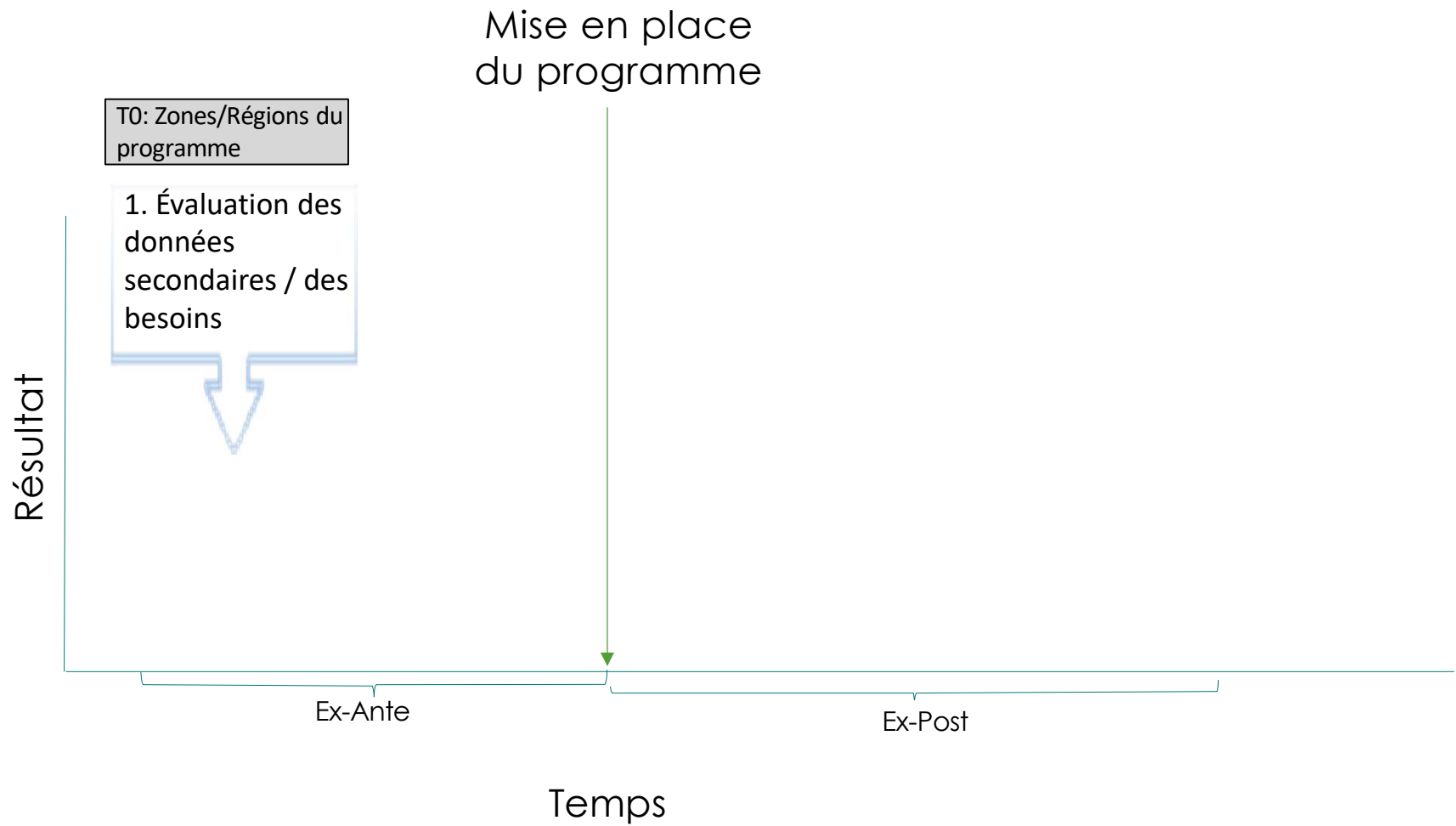
- **Calendrier / Echéances**
- **Qualité des données et éthique de recherche**
- **Besoins en données et sources de données**
- **Budget**



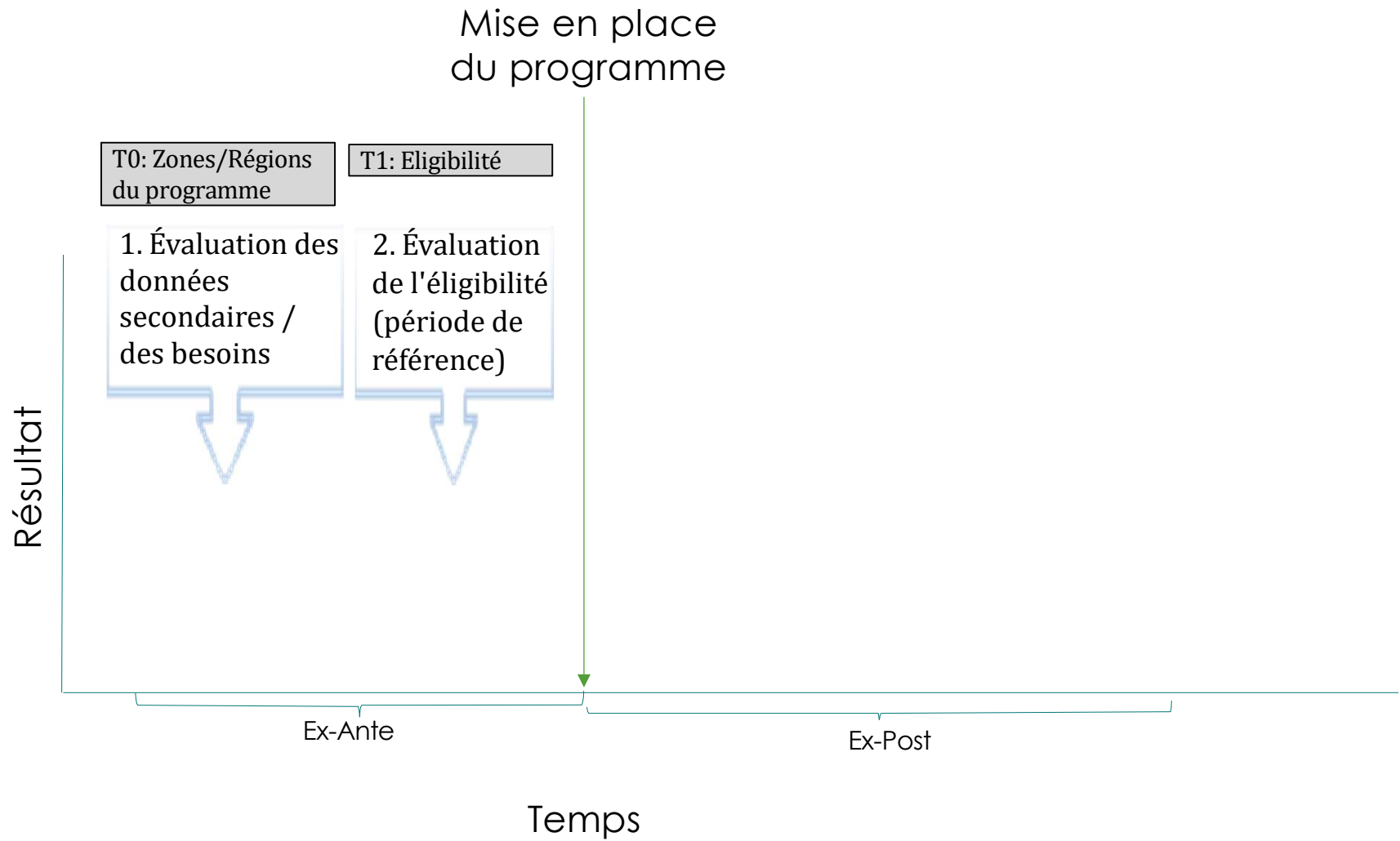
# CALENDRIER / ECHÉANCES



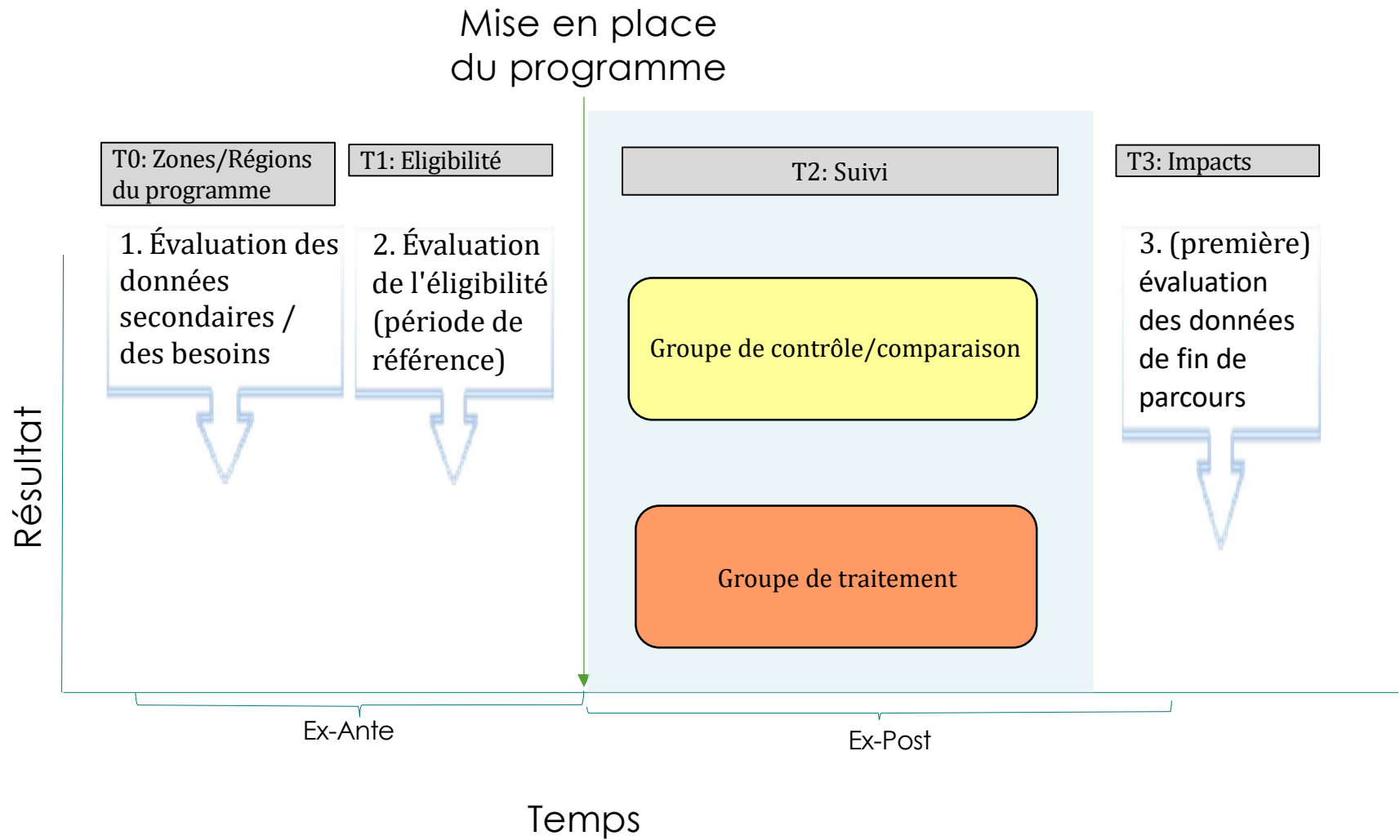
# Calendrier



# Calendrier



# Calendrier



## Remarques générales:

- Evaluation d'Impact et mise en place du projet sont *entrelacés*
- Les évaluations d'impact les plus robustes doivent être planifiées au **début** du projet, avant le début de sa mise en œuvre
- Phases d'évaluation:
  - I. Enquête de référence** (si nécessaire): **AVANT** la mise en place du projet
  - II. Enquête de mi-parcours** (optionnelle)
  - III. Enquête de fin**: Après la fin du projet


- La décision pour les enquêtes de référence et de mi-parcours dépend du ***cadre d'évaluation choisi*** ainsi que des intérêts et des ressources du projet.
  - ECR → la collecte de données de référence est hautement souhaitable mais pas strictement nécessaire
  - DD → enquêtes de référence et de fin obligatoires
- Cette décision soit être prise ***en consultation*** avec un spécialiste d'EI

# Calendrier d'une phase de l'évaluation



8 – 12 mois

1. Formation de l'équipe d'évaluation
2. Examen de la documentation
3. Mission de cadrage
4. Conception de l'évaluation
5. Échantillonnage
6. Outils de collecte de données (par exemple, questionnaire)
7. Collecte des données
8. Nettoyage et analyse des données
9. Communication des résultats
10. Diffusion des résultats



Répété pour  
**chaque**  
phase  
d'évaluation

# Calendrier

## Exemple pour une enquête de référence

Enquête de référence –Année 2022 - Mois												
	Jan	Fev	Mar	Avr	Mai	Juin	Juil.	Août	Sep	Oct	Nov	Dec
Préparation Mission de Cadrage												
Mission de Cadrage												
Revue des documents												
Rapport de démarrage												
Préparation des outils d'enquête												
Préparation collecte de données												
Pré-test et formation												
Collecte de données												
Mise en place du projet (au plus tôt)												
Nettoyage des données												
Analyse des données												
Rapport de l'enquête de référence												
Diffusion des résultats												

# Difficultés et problèmes éventuels

---

## 1. Défis prévisibles



- Autorisation éthique et autorisations de recherche locales
- Les appels d'offre et passation de marchés prennent du temps
- Vacances/festivals/élections
- Données manquantes/incomplètes

1. Prévoir suffisamment de temps pour les activités !!
2. Les connaissances locales sont importantes en matière de timing !!
3. Demandez les coordonnées des personnes interrogées !



# Difficultés et problèmes éventuels

---

## 2. Défis imprévisibles



- Catastrophes naturelles, pandémies, conflits locaux
- Retards dans la mise en œuvre du projet
- Changement d'équipe de projet/de personne de contact du partenaire local

4. Soyez prêt à faire face aux changements et à prendre des marges de sécurité !!
5. Soyez flexibles et innovants !!
6. Documentez tout !!



# QUALITÉ DES DONNÉES ET ÉTHIQUE DE RECHERCHE

# Stratégies pour améliorer la qualité des données

---

- **Améliorer** la façon de collecter, stocker et gérer les données au cours du programme.
- Envisager des **directives** (à l'échelle du secteur) **liées à la propriété, à la protection et à la sécurité des données** (définir un cadre institutionnel interne pour la gouvernance des données).
- Tirer parti des innovations du secteur privé en matière de science des données
- Envisager des contrôles de qualité des données

## Ressource pertinente

---

- Le **Rapport sur le développement dans le monde 2021** (World Development Report for 2021) est sur le sujet des **Données pour le développement** (Data for Development)
- Ce rapport va influencer la recherche et la pratique, et lancer une discussion entre les différents acteurs du secteur

# Plan d'analyse des données

---

- Un plan d'analyse est volontairement élaboré par les chercheurs afin de montrer un niveau d'engagement élevé
- Engagement contre le "data mining" et la sélection de résultats statistiquement significatifs, positifs ou négatifs.
- Idée : Si un chercheur peut choisir les résultats à communiquer, il est facile de voir comment les résultats peuvent être manipulés.

Published on **Development Impact**

## A pre-analysis plan checklist

DAVID MCKENZIE | OCTOBER 28, 2012

This page in: [English](#)



A **pre-analysis plan** is a step-by-step plan setting out how a researcher will analyze data which is written in advance of them seeing this data (and ideally before collecting it in cases where the researcher is collecting the data). They are recently starting to become popular in the context of randomized experiments, with [Casey et al.](#) and [Finkelstein et al.](#)'s recent papers in the QJE both using them. There is also some discussion in political science – see this [recent paper](#) by Macarten Humphrey's and co-authors.

There are several goals in specifying an analysis plans, but one important reason is to avoid many of the issues associated with data mining and specification searching by setting out in advance exactly the specifications that will be run and with which variables. This is particularly important for interventions which have a whole range of possible different outcomes, like the CDD programs looked at by Casey et al. They look at 334 different outcomes, and illustrate that they could have picked 7 outcomes that made their program look like it strengthened institutions, or alternatively

# Plan d'analyse des données

---

- **Un peu avant la collecte de données pour l'enquête de fin** le document est enregistré (sécurisé/archive) en ligne (AEA RCT Registry; RIDIE - 3IE) et peut être demandé ultérieurement par la communauté scientifique à titre de référence

# Pour qui un plan d'analyse est-il utile ?

---

- Il ne s'agit pas d'un document d'autorisation éthique ou gouvernementale, bien qu'il puisse être utilisé comme tel
- Transparence et engagement d'une évaluation d'impact



# Comment ça marche?

---

- Dans un scénario idéal, le plan d'analyse correspond à la rédaction du rapport avant de voir les résultats (données finales).
- Toutefois, dans la plupart des cas, les données elles-mêmes peuvent révéler des tendances intéressantes qui méritent d'être explorés. C'est pourquoi le document est (moralelement) contraignant pour l'analyse principale, mais pas forcément pour des analyses complémentaires, tant que celles-ci sont bien documentées.

# Comment ça marche?

Spécifier les variables, les procédures de nettoyage des données, les spécifications de régression, etc.

*Table 1*  
**Pre-Analysis Plan Checklist**

<i>Item</i>	<i>Brief description</i>
Primary outcome variable	The key variable of interest for the study. If multiple variables are to be examined, one should know how the multiple hypothesis testing will be done.
Secondary outcome variable(s)	Additional variables of interest to be examined.
Variable definitions	Precise variable definitions that specify how the raw data will be transformed into the actual variables to be used for analysis.
Inclusion/Exclusion rules	Rules for including or excluding observations, and procedures for dealing with missing data.
Statistical model specification	Specification of the precise statistical model to be used, hypothesis tests to be run.
Covariates	List of any covariates to be included in analysis.
Subgroup analysis	Description of any heterogeneity analysis to be performed on the data.
Other issues	Other issues include data monitoring plans, stopping rules, and interim looks at the data.

Source: Olken (2015)



# **BESOINS EN DONNÉES ET SOURCES DE DONNÉES**



# **COLLECTES DE DONNÉES PRIMAIRES**

**(recueillies directement par les chercheurs auprès des bénéficiaires  
et des sources principales)**

# Collecter les bonnes données

## Les principes “CART”

### Crédible

Collecter des données de haute qualité et les analyser correctement



### Credible

Collect high quality data and analyze the data accurately



### Actionable

Commit to act on the data you collect

### Exploitable

Engagement à exploiter les données collectées

### Responsable

S'assurer que les avantages de la collecte de données l'emportent sur les coûts



### Responsible

Ensure the benefits of data collection outweigh the costs



### Transportable

Collect data that generate knowledge for other programs

### Transportable

Collecter des données qui génèrent des enseignements pour d'autres programmes

Source: Gugerty, Mary Kay and Dean Karlan (2018)

## Collecter des données de haute qualité et les analyser correctement

Possible lorsque les données sont :

- Valides : elles doivent capturer l'essence de ce qu'elles cherchent à mesurer.
- Fiables : la même procédure de collecte de données doit produire les mêmes données.
- Sans biais : il ne doit pas y avoir de différences systématiques entre la façon dont une personne répond à une question et la véritable réponse.

## Engagement à exploiter les données collectées

- Plan pour l'utilisation des données ?
- Ne collecter que les données qui seront utilisées
  - Y a-t-il une mesure spécifique que vous allez entreprendre à partir des résultats ?
  - Disposez-vous des ressources et de l'engagement nécessaires pour prendre cette mesure ?
- Mettez en place les bons systèmes pour traiter les données que vous collectez

# Responsable

---

S'assurer que les avantages de la collecte de données l'emportent sur les coûts

- Collecter trop de données est inefficace
- Collecter trop peu de données ou ne pas collecter de données sur ce qui s'est passé n'est pas responsable.
  - Le manque de données peut cacher les défauts d'un programme et conduire à une mauvaise prise de décision.



## Compromis:

- Méthodes de collecte des données : existe-t-il des méthodes moins coûteuses/plus efficaces sans diminuer la qualité ?
- Utilisation des ressources : Le budget est-il justifié compte tenu des résultats attendus par rapport au reste du budget du programme ?
- Utilisation du temps des enquêtés : L'information recherchée justifie-t-elle le temps demandé aux répondants ?
- Le moment est-il bien choisi pour une évaluation d'impact ?  
Qu'espère-t-on apprendre ? Les décisions futures seront-elles influencées par les résultats ?

# Transportable

---

Collecter des données qui génèrent des enseignements pour d'autres programmes

- L'objectif est de générer des leçons qui peuvent aider à concevoir/investir dans des programmes et des politiques efficaces.
- Nécessité d'une théorie sous-jacente pour expliquer les résultats : Votre TdC peut-il être reproduite ?
  - Des TdC claires et complètes aideront à générer des travaux similaires, en évaluant si votre TdC peut fonctionner dans d'autres contextes.



# TAILLE D'ÉCHANTILLON

# Taille d'échantillon

---

- Question d'évaluation : Quel est l'impact ou l'effet causal d'un programme sur le résultat qui nous intéresse ?  
→ En d'autres termes : L'impact du programme mesuré est-il différent de zéro ?
- Alors, quelle doit être la taille de l'échantillon ? Quelle est la taille minimale de l'échantillon requise pour mener une étude qui répondra de manière convaincante aux questions d'intérêt ?
  - Plus l'échantillon est grand, plus l'estimation est précise MAIS ...
  - Collecter plus de données est coûteux !
- Principale préoccupation pratique pour l'EI : compromis entre le coût de la collecte des données et la précision des estimations  
→ Les calculs de puissance sont un outil permettant d'aider cette décision

# Calculs de puissance: les paramètres de la formule

---

## 1. Effet minimum détectable (EMD):

Quel est la magnitude d'impact en dessous de laquelle une intervention doit être considérée comme infructueuse ?

Intuition : il est plus difficile de détecter de petits impacts que de grands impacts.

Petits impacts → grande taille d'échantillon

## 2. Précision statistique

Précision avec laquelle nous pouvons mesurer l'EMD, étant donnée la taille de l'échantillon.

2 composantes : **niveau de significativité** (probabilité d'une erreur de type I) et **puissance statistique**

## 3. Informations de référence

Valeur (moyenne) du résultat d'intérêt à la période de référence

Écart-type de référence du résultat d'intérêt

# Calculs de puissance: les paramètres de la formule

Information Disponible	On peut calculer
EMD Niveau de significativité Puissance statistique	Taille de l'échantillon
Taille de l'échantillon Niveau de significativité Puissance statistique	EMD

# Calculs de puissance – Précision statistique

	Aucun impact détecté par l'EI	Impact significatif détecté par l'EI
L'intervention n'a pas d'impact	Pas d'erreur (conclusion correcte)	<b>Erreur de type I</b> (faux positif)
L'intervention a un impact	<b>Erreur de type II</b> (faux négatif)	Pas d'erreur (conclusion correcte)

- **Niveau de significativité** = probabilité d'erreur de type I
  - Valeur typique en sciences sociales = 5%
- **Puissance statistique** = probabilité de détecter un impact lorsqu'il existe réellement → en fait égal à 1 moins la probabilité d'erreur de type II
  - Valeur typique en sciences sociales = 80%

# Calculs de puissance – Informations de référence

---

1. Où pouvons-nous trouver des informations sur les valeurs de référence des résultats ? Ces informations sont-elles disponibles pour votre projet ?
2. Comment pouvons-nous déterminer l'EMD ? Y a-t-il un impact attendu clairement défini pour votre projet ?



# Plus l'effet attendu du projet est grand, plus la taille d'échantillon nécessaire pour le détecter est...

---

---

petite

---

grande

# Plus l'effet attendu du projet est grand, plus la taille d'échantillon nécessaire pour le détecter est...

---

---

petite

---

grande

**Pour un effet donné, plus la taille d'échantillon est petite, plus la précision statistique est...**

---

---

élevée

---

faible

**Pour un effet donné, plus la taille d'échantillon est petite, plus la précision statistique est...**

---

---

**élevée**

---

**faible**

# Calculs de puissance – Compromis

Etant donné(e)...	Si...	Alors...
Taille de l'échantillon	EMD ↑	Précision ↑
Taille de l'échantillon	EMD ↓	Précision ↓
EMD	Taille de l'échantillon ↑	Précision ↑
EMD	Taille de l'échantillon ↓	Précision ↓
Précision	Taille de l'échantillon ↑	EMD ↓
Précision	Taille de l'échantillon ↓	EMD ↑

# Calculs de puissance – Compromis

---

## ATTENTION

- Les compromis dans les calculs de puissance **ne sont pas linéaires**

### Example:

- Revenu de référence = 1 000€, Ecart-type = 1 000
- Significativité = 5%; puissance = 80%
- Effect attendu (EMD): +50% (c.à.d +500€)
  - Taille d'échantillon requise pour détecter l'impact: **128** (64 dans le groupe de traitement, 64 dans le groupe de contrôle)
- Effect attendu (EMD): +25% (c.à.d +250€)
  - Taille d'échantillon requise: **506** (253 par groupe)

EMD divisé par 2, mais la taille d'échantillon a presque **quadruplé !!!**

# Pourquoi une grande taille d'échantillon?

---

Toutes choses égales par ailleurs, un grand échantillon est nécessaire...

- ...si vous voulez capturer un petit EMD
- ...si vous prévoyez une adhésion au programme imparfaite
- ...si vous prévoyez une forte attrition
- ...si vous n'avez pas de valeurs de référence
- ...si votre variable de résultat présente une variance élevée
- ...si vous avez une approche par grappes
- ...si les unités (par exemple, les personnes) d'un groupe (par exemple, un village) sont très semblables
- ...si vous utilisez une méthode d'évaluation non expérimentale (par ex. appariement).



Center for Evaluation  
and Development

---

# **COLLECTES DE DONNÉES SECONDAIRES**

**(données administratives)**

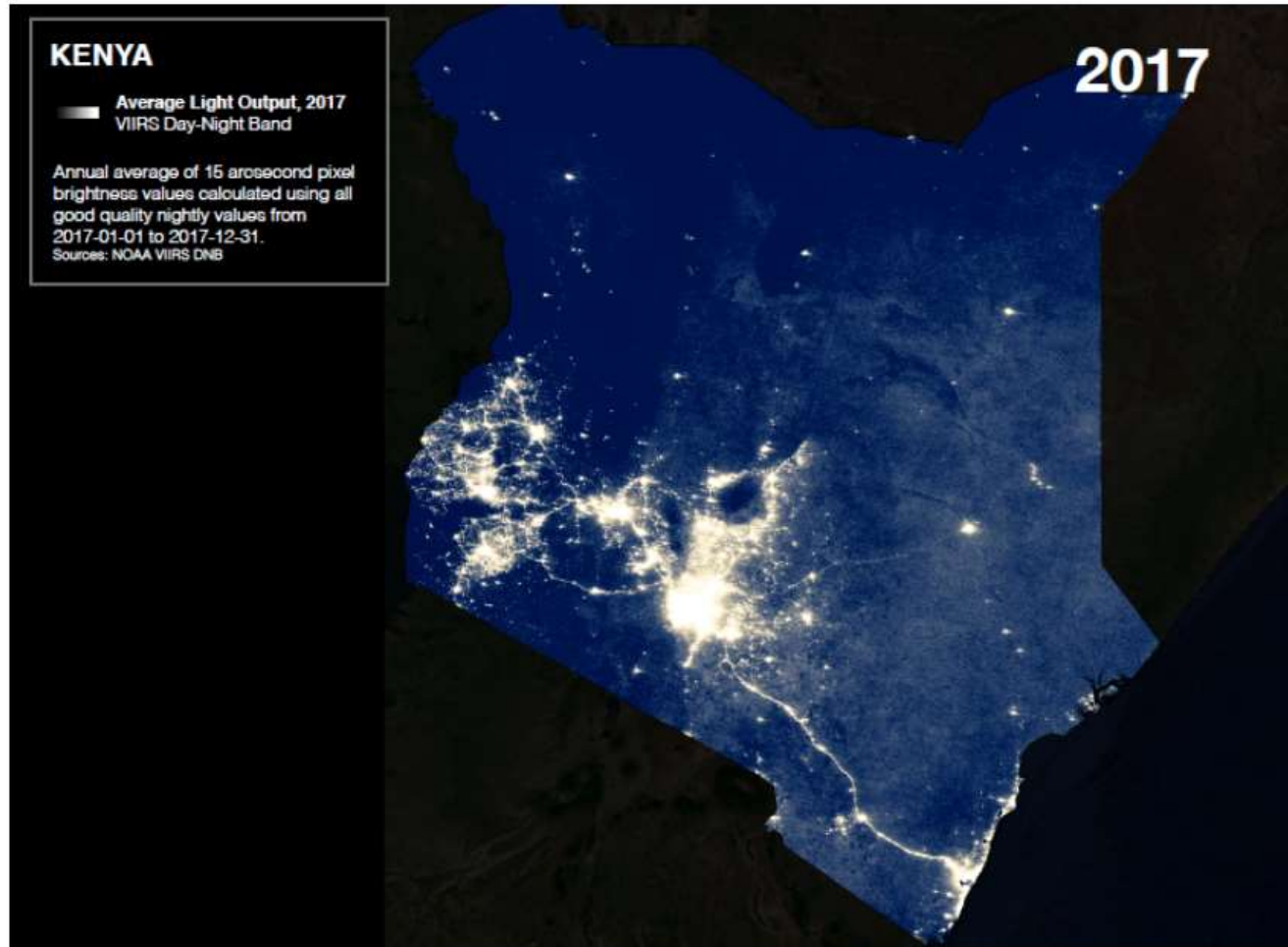


# Les sources de données secondaires

---

- Données administratives (anonymisées mais désagrégées)
- Données du secteur privé (par exemple sur la consommation)
- Données géo-référencées
- Données de suivi du programme !!!

## Exemple de données secondaires: Intensité lumineuse moyenne



Source: [Link to World bank blog «Innovations in satellite measurements for development»](#)

# Exemple de données secondaires: Zones construites



Source: [Link to World bank blog «Innovations in satellite measurements for development»](#)

# Avantages des données secondaires

---

- **Plus rapide** que les enquêtes qui prennent beaucoup de temps
- **Large bases de données** administratives
- Données **moins coûteuses**
- Estimations d'impact **plus précises**
- Données plus **précises avec moins d'attrition**
- Plus **inclusives** que les enquêtes

# Données de suivi

---

**Le contrôle/suivi** consiste généralement à suivre les progrès réalisés par rapport aux plans ou aux objectifs préalablement identifiés, en utilisant des données faciles à saisir et à mesurer de manière continue.

L'évaluation de l'impact **ne doit pas** se faire sans données solides sur la mise en œuvre du projet.

## Suivi: Objectif

---

Le suivi est effectué à des fins très diverses, qui ont généralement peu à voir avec l'évaluation d'impact.

Par exemple :

- Utilisation interne par les gestionnaires de projet pour déterminer si le projet est en phase avec ses objectifs ou non (par exemple, quels services sont réellement fournis ; qui est servi).
- Répondre aux exigences des donateurs en matière de rapports et de responsabilité
- Servir de *système d'alerte précoce* et, en cas de résultats négatifs ou inattendus, suggérer la nécessité d'envisager un changement d'approche alors que le projet ou le programme est toujours en cours.

# Suivi - Exemple

---

Le suivi pourrait permettre de « tracker »:

- **quelles** nouvelles modalités de traitement (compléments) sont introduites.
- quelles nouvelles modalités de traitement (compléments) sont introduites **ensemble** (combinaisons). Ces informations nous permettraient de savoir si les protocoles de randomisation sont respectés.
- **quand** les nouvelles modalités de traitement (compléments) sont introduites. L'identification d'un moment précis permettrait de mesurer l'intensité de l'exposition au traitement (en termes de durée).



# BUDGET



# Budget

---

- Le budget pour une EI (en particulier la collecte de données) doit être estimé de manière **réaliste** et **alloué clairement** au début du projet.

Facteurs déterminants :

- Niveau général du **coût de la vie/des prix** dans un pays.
- **Taille de l'échantillon** et nombre de collectes de données
- **Transport**
- **Sécurité**
- Nombre de **langues** parlées dans la région du projet
- **Sous-traitance** de la collecte de données à une entreprise externe

# Lignes de budget pour collecte de données

## Exemples de lignes de budget

<b>Coût du personnel</b>	Coordinateur de terrain, superviseur, enquêteur, modérateur (qualitatif), traducteur....
<b>Coût de la formation</b>	Salle de formation, restauration, allocation de formation pour les participants, hébergement....
<b>Transport</b>	Location de voiture, carburant, chauffeur, ticket de bus, moto pendant la formation et la collecte des données.
<b>Autre</b>	Tablettes, incitations monétaires, impression du matériel de formation, coûts de communication/internet, lieu des discussions de groupe (qualitatif).

# Budget

## Pours/Contres l'embauche d'une firme externe pour la collecte et/ou l'analyse de données



- Les entreprises spécialisées produisent généralement des données de meilleure qualité
- Libère le temps de l'équipe de projet
- Souvent, il n'y a pas d'autre choix puisqu'un appel d'offre est obligatoire et qu'il s'agit d'une bonne pratique
- Assure l'indépendance de l'évaluation d'impact



- Les appels d'offre prennent du temps
- Le coût est généralement plus élevé (y compris pour la coordination)
- Moins flexible et peut être risqué
- Il est toujours nécessaire de vérifier la qualité et l'analyse des données

Pour les appels d'offre:  
important d'avoir quelqu'un de qualifié pour juger  
de la qualité des offres techniques

Avez-vous des questions?

Vous pouvez aussi m'écrire à l'adresse email suivante:

[m.gillaizeau@c4ed.org](mailto:m.gillaizeau@c4ed.org)

---

Merci pour votre attention!

- Abebaw Ejigie, D., Fentie, Y. and Kassa, B. (2010), The impact of a food security program on household food consumption in Northwestern Ethiopia: A matching estimator approach, *Food Policy*, 35, issue 4, p. 286-293
- Berge, L., I. Oppedal, K. Bjorvatn, Tungodden, B. (2011) Human and Financial Capital for Microenterprise Development: Evidence from a Field and Lab Experiment. NHH Discussion Paper Sam 1 2011. Norwegian School of Economics, Bergen, Norway
- Gertler, P. J., Martinez, S., Premand, P., Rawlings L., Vermeersch, C. M. J. (2011) Impact Evaluation In Practice, First Edition, The World Bank.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings L., Vermeersch, C. M. J. (2016) Impact Evaluation In Practice, Second Edition, The World Bank.
- Gugerty, M. K., Karlan, D. (2018) The Goldilocks Challenge: Right-Fit Evidence for the Social Sector, Oxford Scholarship Online.
- MacPherson, C. Sterck, O. (2019) Humanitarian vs. Development Aid for Refugees: Evidence from a Regression Discontinuity Design, CSAE Working Paper Series 2019-15, Centre for the Study of African Economies, University of Oxford.
- Olken, B. A. (2015) Promises and Perils of Pre-analysis Plans, *Journal of Economic Perspectives*, 29 (3): 61-80.
- Schultz, T. (2004) School subsidies for the poor: evaluating the Mexican Progresa poverty program, *Journal of Development Economics*, 74(1): 199-250.
- Tan, B. (2018) Prioritizing the Learning Agenda: the CART Principles, Practitioners Forum, Adaptive Programming and Monitoring, Evaluation and Learning, Philippines.
- Trzcinski, R. (2011) Active labour market measures and entrepreneurship in Poland, Impact Evaluation Spring School, Hungary.
- White, H., & Raitzer, D. A. (2017). Impact evaluation of development interventions: A practical guide. Asian Development Bank.