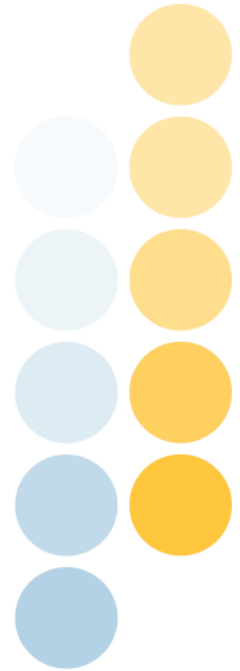


PEI Impact Evaluation Workshop

Moving Economic Inclusion to scale



Méthodes d'évaluation d'impact experimentales



Les évaluations d'impact sont un type particulier d'évaluation



Types de questions d'évaluation:

- **Questions Descriptives** pour décrire ce qui se passe, le processus de mise en oeuvre d'un programme (**SUIVI**)
 - **Questions Normatives** pour comparer ce qui se passe avec ce qui devrait se passer. (**EVALUATION DE PROCESSUS**)
 - **Questions de cause-à-effet** pour examiner les résultats et apprécier les différences de ces résultats attribuables au programme (**EVALUATION D'IMPACT**)
-





L'objectif de l'évaluation d'impact est d'estimer l'effet **causal** ou l'**impact** d'un programme sur des indicateurs de résultats

La chaîne de résultats d'un programme



Les évaluations d'impact ne sont pas pour tous les projets...

- Prioriser l'application aux programmes:
 - Novateurs
 - Stratégiquement pertinents
 - Reproductibles
- Les évaluations d'impact servent aussi pour tester des alternatives et modalités de mise en oeuvre dans un programme
 - Quelle est la modalité la plus efficace?
 - Quelles sont les innovations efficaces?

Comment répondre à la question type d'évaluation d'impact?

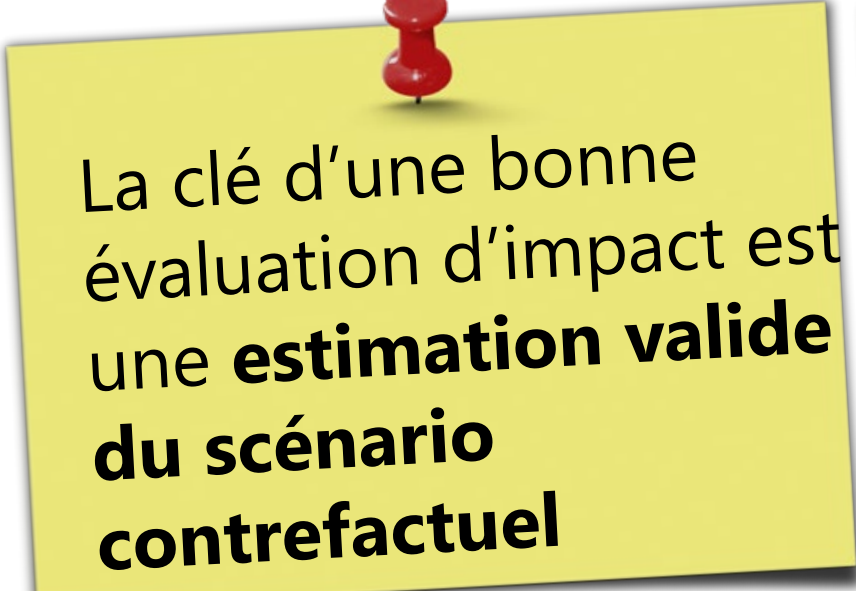
“
Évaluer l'effet causal (**l'impact**) du Programme (**P**) sur des résultats (**Y**).

Exemple: Quel est l'effet d'un programme de transferts monétaires (**P**) sur la consommation des ménages (**Y**)?




Pour estimer l'impact, nous devons estimer **le scénario contrefactuel**.

- Qu'est-ce qui arriverait en l'absence du programme?

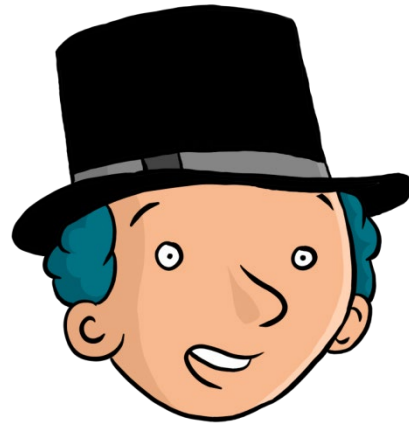
A yellow sticky note with a red pushpin at the top center. The text is written in black, with some words in bold.

La clé d'une bonne évaluation d'impact est une **estimation valide** du **scénario contrefactuel**

A yellow sticky note with a red pushpin at the top center. The text is written in black, with some words in bold.

L'estimation du contrefactuel nécessite un **groupe de comparaison**

Exemple: Quel est l'impact de...



Donner à Colin

Un transfert monétaire



(P)

Sur sa consommation

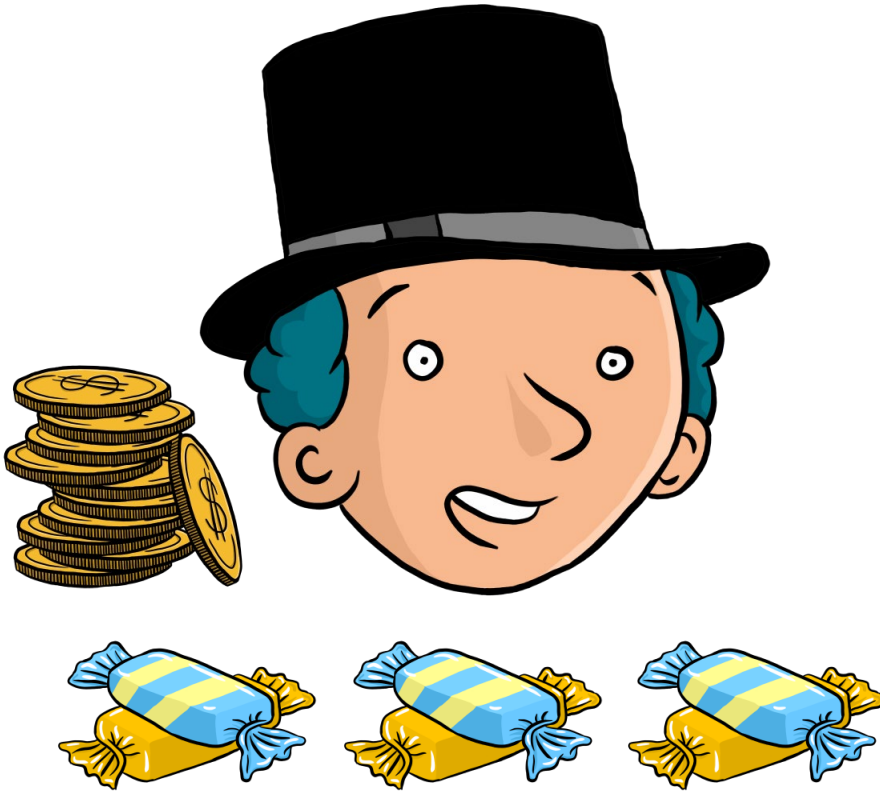


(Y)?

Le "clone" parfait

Colin

Son "clone"



6 bonbons

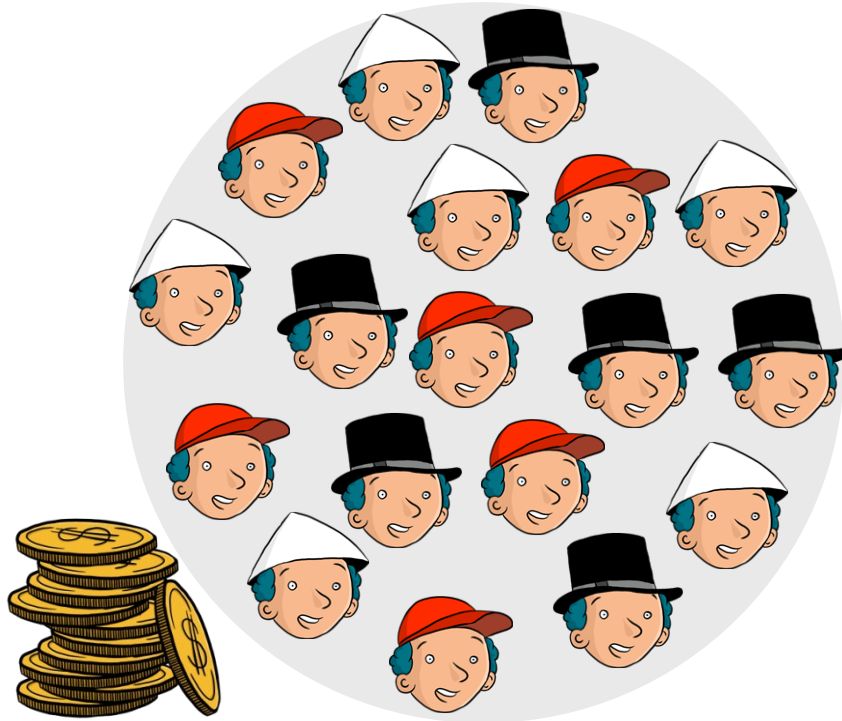


4 bonbons

IMPACT=6-4=2 bonbons

Dans la réalité, utiliser les statistiques

Le groupe de traitement



Le groupe témoin



Moyenne de $Y=6$ bonbons

Moyenne de $Y=4$ bonbons

IMPACT = $6 - 4 = 2$ bonbons

Trouver de bons groupes de comparaison

Nous voulons trouver des “**clones**” pour les Colins de nos programmes.

Les groupes traités et de comparaison doivent:

- Avoir les mêmes caractéristiques
- Excepté en ce qui a trait à leur participation au programme.

Dans la pratique, utiliser les règles d'éligibilité et d'assignation du programme pour générer un scénario contrefactuel valable.


Trouver un bon groupe de comparaison

Nous voulons trouver des “**clones**” pour les Colins de nos programmes.

Les groupe traités et de comparaison doivent:

- Avoir les mêmes caractéristiques

qui a trait à leur participation



Lorsque l'on dispose d'un scénario contrefactuel valide, la **seule chose** expliquant des résultats différents entre les groupes exposés au traitement et les groupes de comparaison est **l'intervention (P)**

en respectant les règles d'éligibilité et comme pour générer un contrefactuel valable.

Deux comparaisons problématiques pour mesurer l'impact



Avant & Après

Comparer certaines personnes **Avant et Après** leur exposition au Programme

Problème: D'autres choses peuvent se modifier dans le temps.

Inscrits & Non inscrits

Comparer le groupe de personnes **inscrit** au programme à ceux et celles qui ont **choisi** de **ne pas** s'y inscrire.

Problème: Biais de sélection. Beaucoup de facteurs peuvent contribuer le choix de s'inscrire ou non.

*Ces deux comparaisons ne sont **pas fiables** pour permettre d'estimer le contrefactuel et l'impact d'un programme*

D'où proviennent les bons groupes de comparaison?

- Les règles d'opération des programmes déterminent la stratégie d'évaluation
- Il est presque toujours possible de trouver un groupe de comparaison valable si :
 - les règles opérationnelles de sélection des bénéficiaires sont équitables, transparentes et reproductibles;
 - et la l'évaluation est prospective
- La conception du projet et la conception de l'évaluation vont ensemble.

L'évaluation d'impact se planifie dès la conception du programme

Evaluation Rétrospective: Le programme a été mis en oeuvre et on cherche à l'évaluer après-coup sur en utilisant des données existantes. *Souvent impossible!*

Evaluation Prospective, l'évaluation est planifiée dès la conception du programme *Souhaitable! Assurer qu'une enquête de référence est collectée et qu'un groupe de comparaison valide existe.*



Il existe **plusieurs méthodes** pour définir ou identifier les groupes de comparaison.

- Assignation aléatoire
- Promotion aléatoire
- Double-différence
- Régression-discontinuité
- Appariement





Choisir la meilleure méthode d'évaluation qui est faisable dans le contexte opérationnel.



Assignation aléatoire du traitement

d'éligible > # de bénéfices disponibles

- Randomiser (assigner le traitement de manière aléatoire)
- Tirer au sort ceux qui recevront le traitement
- Il s'agit d'une manière équitable, transparente et éthique d'accorder des prestations lorsque les populations cibles ne peuvent être départagées sur la base du besoin.

Sur-inscription

- Accorder à chaque groupe éligible les mêmes chances de participer
- Comparer ceux qui bénéficient de la prestation à ceux qui n'en bénéficient pas.

Déroulement aléatoire par phases

- Accorder à chaque groupe éligible les mêmes chances de recevoir le traitement en premier, en second, en troisième ...
- Comparer ceux qui ont reçu le traitement en premier à ceux qui le recevront plus tard.

Exemple: Programme THIMO en Côte d'Ivoire

- Les THIMO ont été introduits comme un instrument de stabilisation pendant une période de conflit.
- Comment décider quels individus sélectionner pour le programme dans un contexte avec forte demande mais un nombre limité de places disponibles?
- Un tirage au sort public fut choisi. Méthode considérée la plus transparente et la moins controversée dans un contexte tendu.
- L'utilisation du tirage au sort était une décision purement opérationnelle.
- L'évaluation d'impact y ajouta des collectes de données de référence et de suivi.



Randomisation et transparence

Personnel du Projet au Niger: “Les autorités politiques ne peuvent pas essayer d’interférer avec la sélection des villages. Tous les chefs de village étaient présents durant la cérémonie de tirage au sort. Personne ne peut se plaindre et essayer de changer les résultats.”



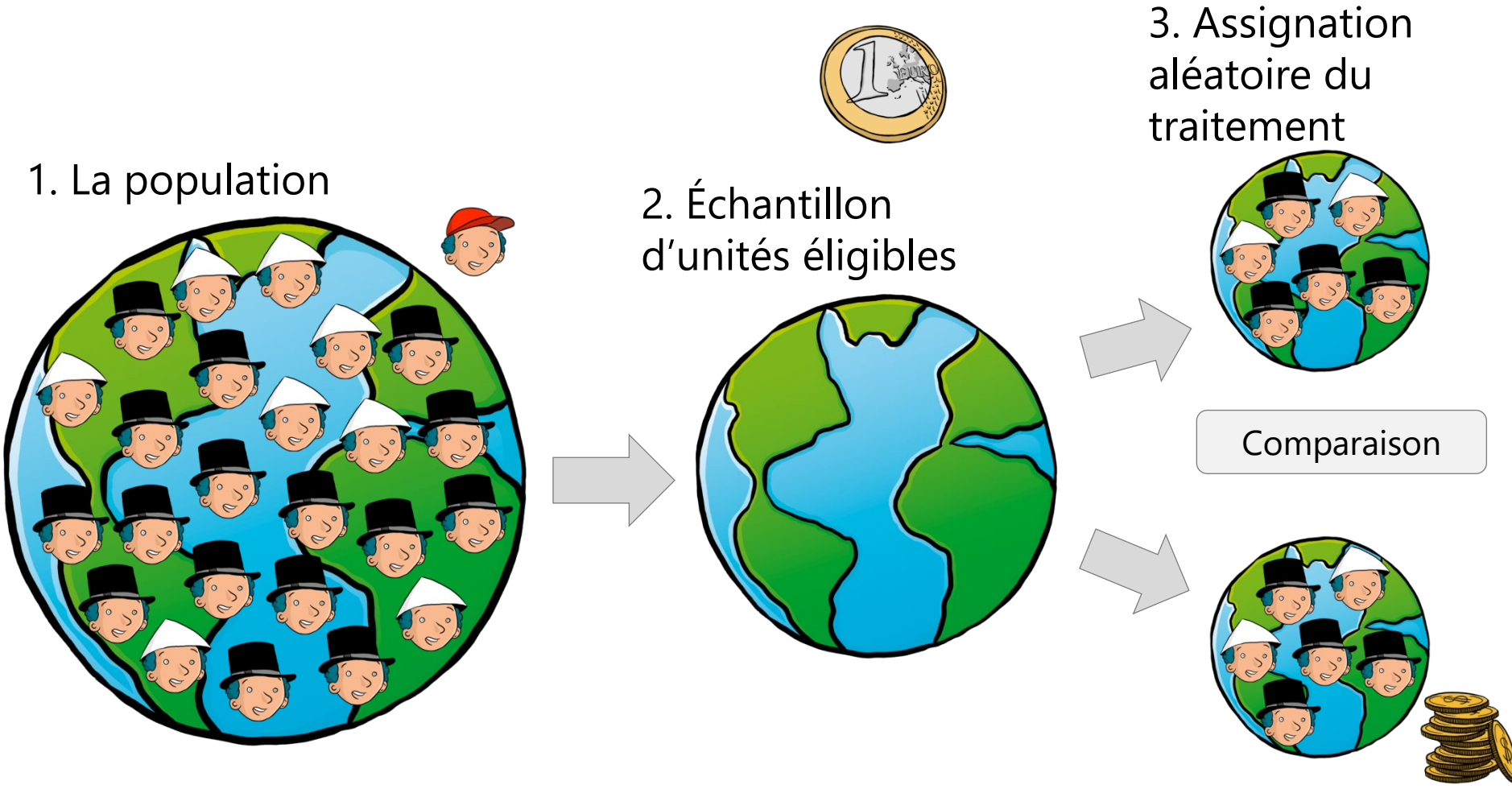
Bénéficiaires au Nicaragua: “Au moins cette fois nous savons pourquoi certains villages furent choisis pour le programme et d’autres non. Souvent ces décisions sont prises sans qu’on ne sache pourquoi un village participe et un autre non.”





Affectation aléatoire

- L'affectation aléatoire peut être utilisée dans certains contextes
- Mais beaucoup de programmes font face à un excès de demande
 - Même après l'application de critères d'éligibilité, il n'y a pas assez de ressources pour servir toute la population éligible.
 - L'affectation aléatoire est une approche transparente et éthique de distribuer les bénéfices d'un programme parmi tous les ménages ou individus éligibles.
 - Fournit la même chance de participer à toutes les unités éligibles.
- L'affectation aléatoire est aussi la méthode d'évaluation d'impact la plus fiable
 - Souvent aussi la plus simple et la moins chère
- Il y a plusieurs manières de mettre en œuvre l'affectation aléatoire

Assignation aléatoire du traitement



 = Non éligible

 = éligible

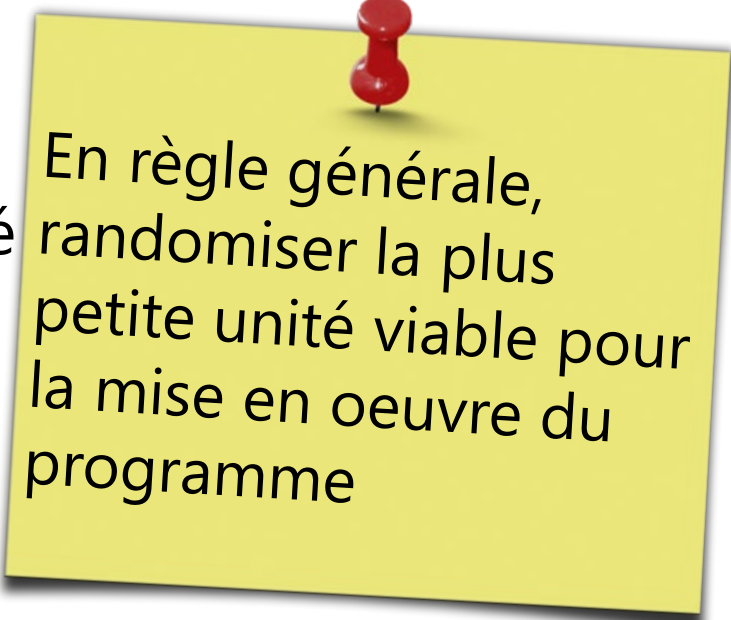
Validité externe

Validité interne

Unité d'assignation aléatoire

- À choisir en fonction du type de programme

- Individuel/Ménage
- École /Centre de santé
- Quartier/Village/Communauté
- Service/District/Région



En règle générale, randomiser la plus petite unité viable pour la mise en oeuvre du programme

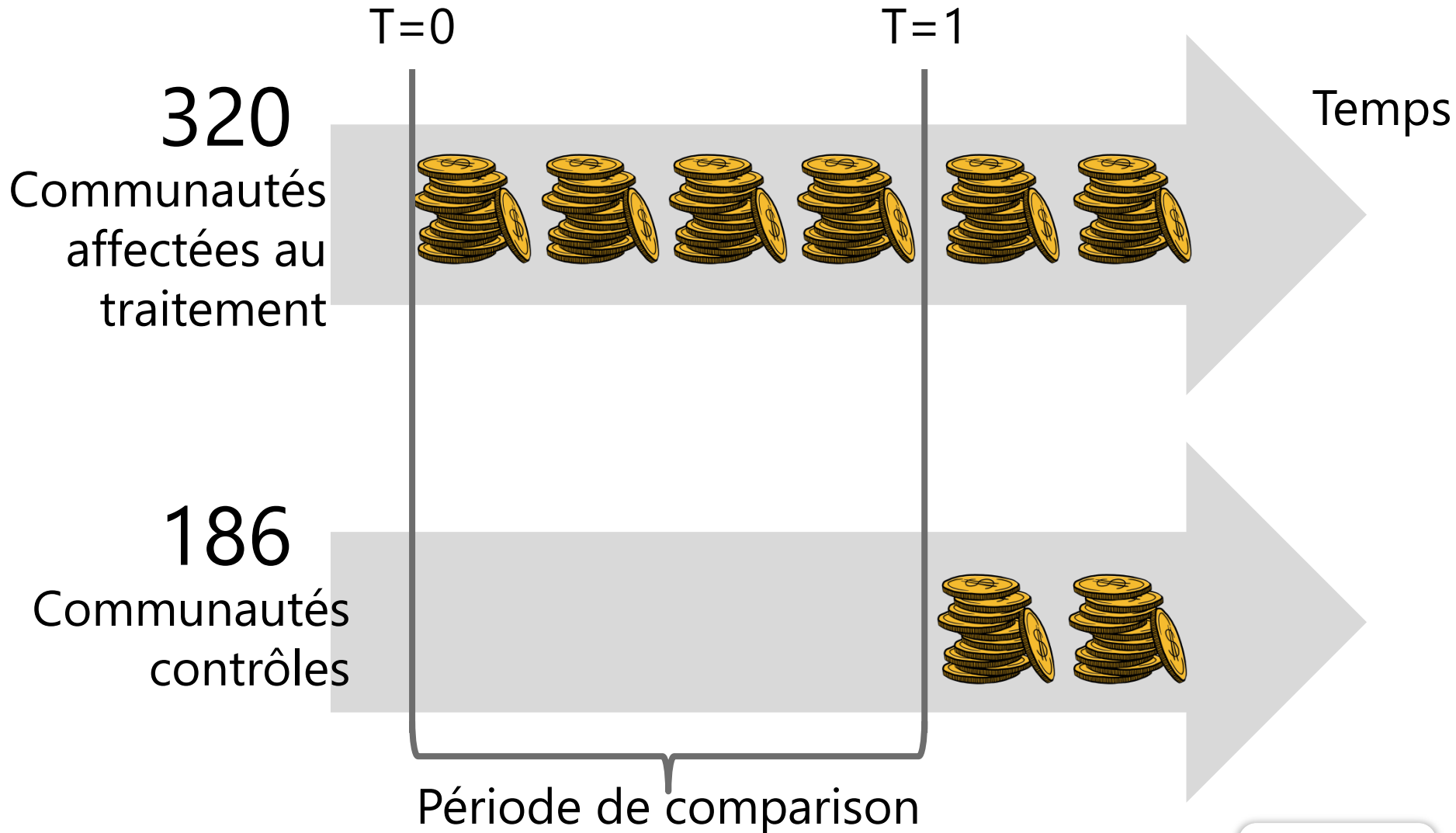
- Attention!

- Un nombre "suffisamment grand" d'unités est nécessaire pour pouvoir détecter un impact minimum désiré: **La puissance statistique.**
- "Spill-over" / contamination entre les 2 groupes

Exemple: **Assignment aléatoire**

- Un programme de transferts monétaires
- Unité de base de la randomisation: La communauté
- Échantillon d'évaluation couvrant 506 communautés
- Introduction de l'élément de randomisation
 - 320 communautés soumises au traitement (14446 ménages):
Premiers transferts effectués en année 1.
 - 186 communautés contrôles (9630 ménages):
Premiers transferts effectués en année 3.

Exemple: **Assignation aléatoire**



Exemple: **Assignation aléatoire**

Comment savoir si nous avons de bons "clones"?

En l'absence du programme, **les groupes traitement et contrôles** doivent être parfaitement identiques

Comparons leurs caractéristiques au point de départ (**T=0**)

Exemple: "Équilibre" au départ

Exemple: Randomisation

| | Groupe contrôle | Groupe traitement | <i>Stat. T</i> |
|--|--------------------|----------------------|-----------------|
| Consommation (mensuelle par personne, en \$) | 233.47 | 233.4 | -0.39 |
| Âge du chef de famille (en année) | 42.3 | 41.6 | 1.2 |
| Âge de l'époux (se) (en année) | 36.8 | 36.8 | -0.38 |
| Niveau d'instruction du chef de famille (en année) | 2.8 | 2.9 | -2.16*** |
| Niveau d'instruction de l'époux (se) (en année) | 2.6 | 2.7 | -0.006 |

NB: Effet statistiquement significatif à 1% indiqué par ***

Cas 3: "Équilibre" au départ

Exemple: Randomisation

| | Groupe contrôle | Groupe traitement | <i>Stat. T</i> |
|--|--------------------|----------------------|----------------|
| Le chef de famille est une femme =1 | 0.07 | 0.07 | 0.66 |
| Nomade=1 | 0.42 | 0.42 | 0.21 |
| Nombre de personne dans le ménage | 5.7 | 5.7 | -1.21 |
| Les toilettes =1 | 0.56 | 0.57 | -1.04 |
| Terres (en hectare) | 1.71 | 1.67 | 1.35 |
| Distance par rapport à l'hôpital (<i>en km</i>) | 106 | 109 | -1.02 |

NB: Effet statistiquement significatif à 1% indiqué par ***

Exemple: **Assignment aléatoire**

| | Groupe de traitement <i>(Assigné aléatoirement au groupe traitement)</i> | Scénario contrefactuel <i>(Assigné aléatoirement au groupe contrôle)</i> | Impact <i>(Différence)</i> |
|--|--|--|--------------------------------------|
| <i>Référence (T=0)</i> Consommation (Y) | 233.47 | 233.40 | 0.07 |
| <i>Suivi (T=1)</i> Consommation (Y) | 268.75 | 239.5 | 29.25*** |

NB: Effet statistiquement significatif à 1% indiqué par ***

À retenir



Assignment aléatoire

Lorsque les échantillons sont grands, l'assignation aléatoire produit deux groupes statistiquement équivalents.

Faisable pour les évaluations prospectives avec sur-inscription/demande excessive.

Nous avons identifié le "**clone**" parfait .

La plupart des programmes pilotes et nouveaux programmes se retrouvent dans cette catégories.

Randomisation
du bénéficiaire

Randomisation de
l'unité de contrôle



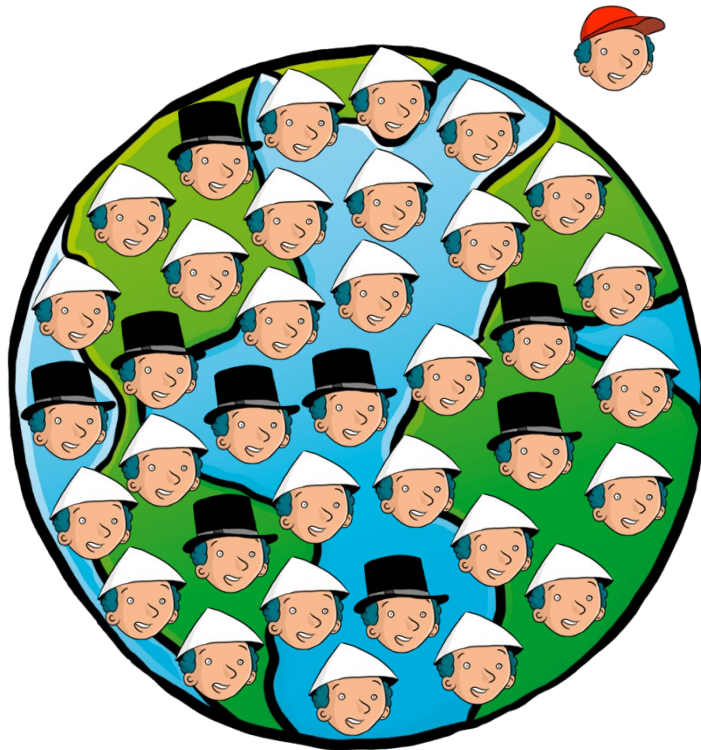
Assignation aléatoire avec différents niveaux de prestations

- Question classique d'évaluation:
 - Quel est l'impact d'un programme sur un résultat?
- Autres types de questions dignes d'intérêt:
 - Quel est le niveau optimal de prestation d'un programme?
 - Quel est l'impact d'un traitement à "forte intensité" comparativement à un autre de "moins forte intensité"?
- Randomisation de 2 niveaux de prestations

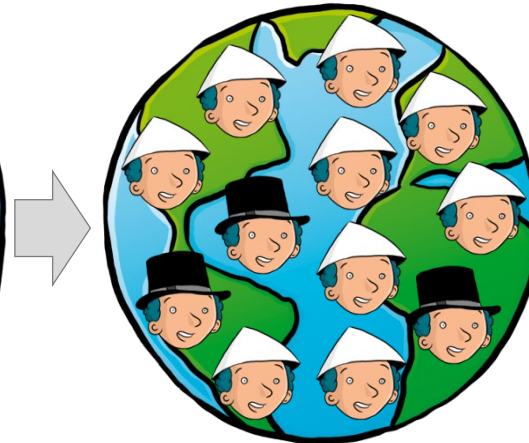
| Comparaison | Faible niveau de prestation | Niveau élevé de prestation |
|---|--|---|
|  |  |  |

Assignment aléatoire avec différents niveaux de prestations

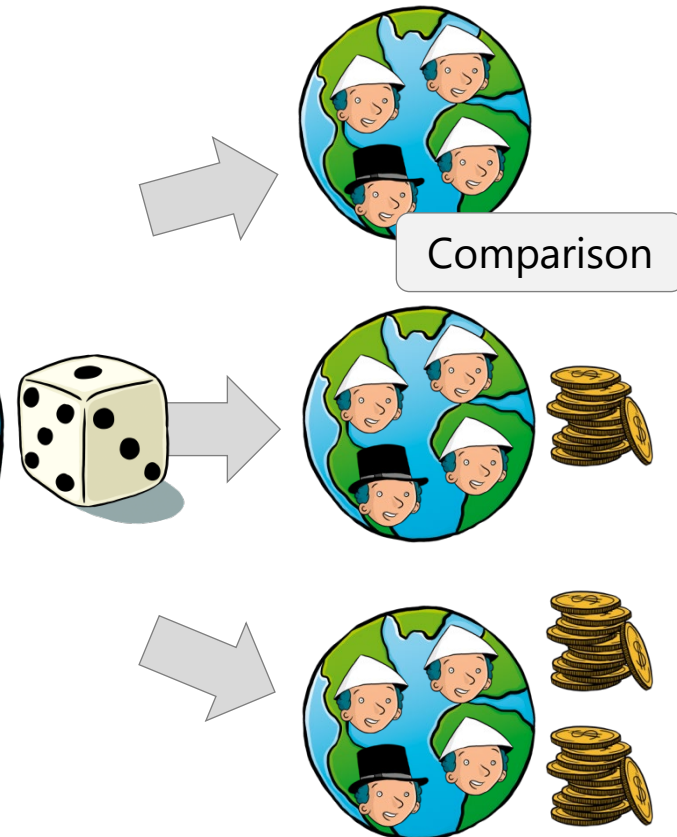
1. Population éligible






2. Échantillon d'unités éligibles



3. Assignment aléatoire des 2 niveaux de prestation.







 = Les non éligibles

  = Les éligibles

Assignation aléatoire: Le scénario d'interventions multiples

- Autre type de questions clés dans le cas d'un programme offrant différentes prestations :
 - Quel est l'impact de chaque prestation comparé à un autre?
 - Y a t-il des complémentarités entre les différentes prestations ?
- Assignation aléatoire avec 2 prestations:

| | | Intervention 1 | |
|----------------|-------------|---|--|
| | | Comparaison | Traitement |
| Intervention 2 | Comparaison | Group A  | Group C  |
| | Traitement | Group B  | Group D  |

Assignation aléatoire

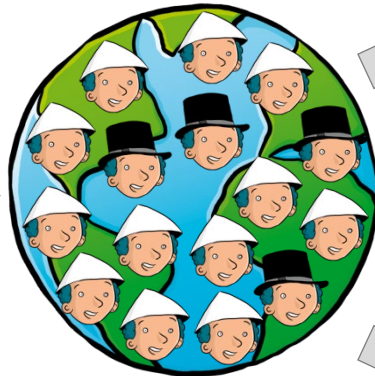
Le scénario d'interventions multiples

4. Assignation aléatoire de l'intervention 2

3. Assignation aléatoire de l'intervention 1

2. Échantillon d'unités éligibles

1. Population éligible



 = Non admissible   = Admissible



L'objectif de l'évaluation d'impact est d'estimer l'effet **causal** ou l'**impact** d'un programme sur des indicateurs de résultats



Pour estimer l'impact, nous devons estimer **le scénario contrefactuel.**

- Qu'est-ce qui arriverait en l'absence du programme?
 - Utiliser des groupes de comparaison
-





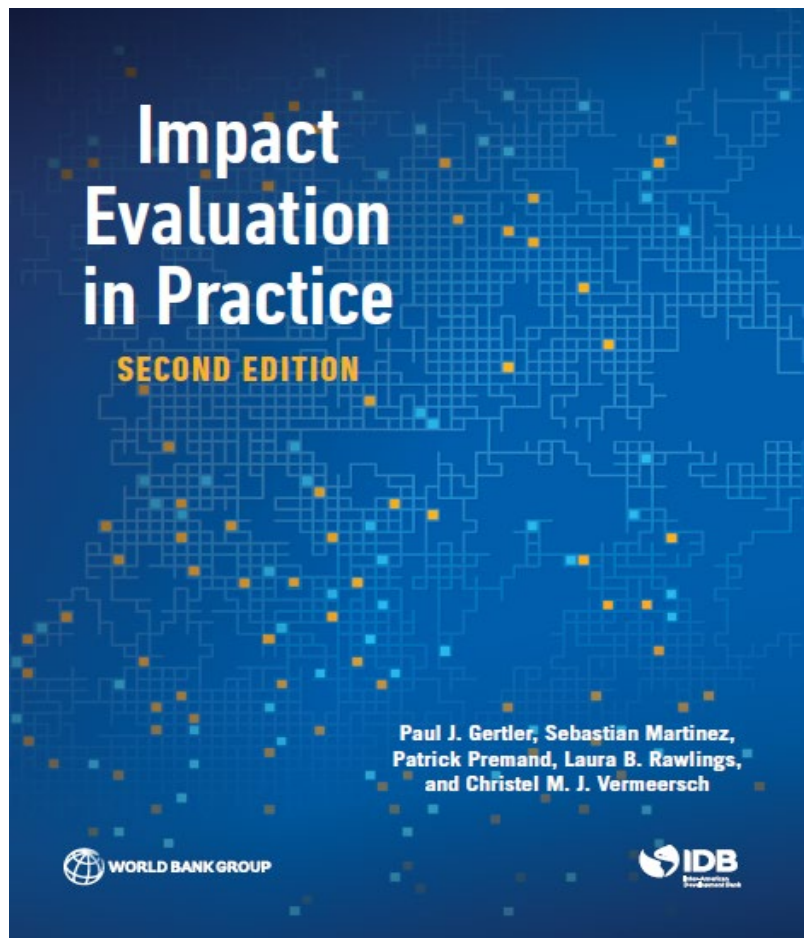
Il existe **plusieurs méthodes** pour définir ou identifier les groupes de comparaison.





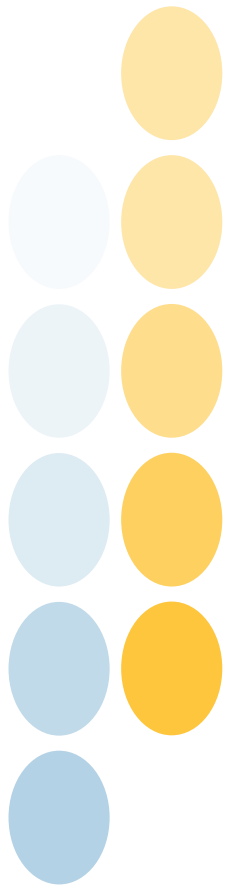
Les méthodes expérimentales
fournissent un groupe de
comparaison similaire au
groupe d'intervention





Référence en Anglais,
Espagnol, Portuguais,
et en Français
(1ère édition)

www.worldbank.org/ieinpractice



Patrick Premand
ppremand@worldbank.org

Merci!



PEI FUNDING PARTNERS



Implemented by



Impact Evaluation Collaborative Workshop | Hosted by PEI and DIME