

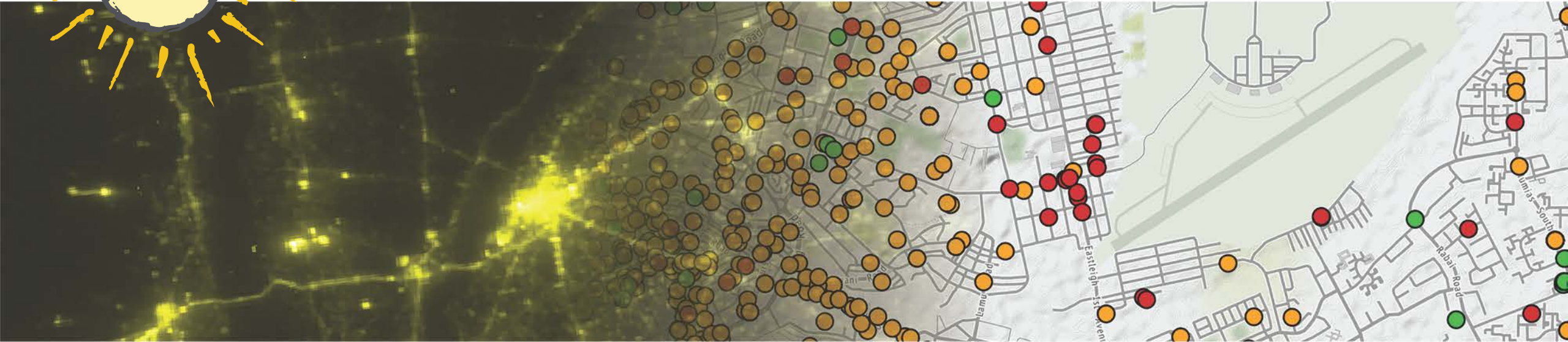


IE CONNECT FOR IMPACT

Transforming the Growth Potential
of Transport Investments

Méthodes Expérimentales et Cas Pratique

Théophile Bougna, Economiste -- DIME

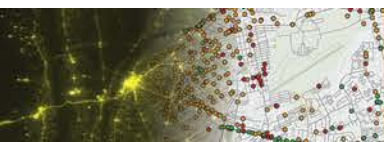


Marrakech, le 4 Décembre 2019



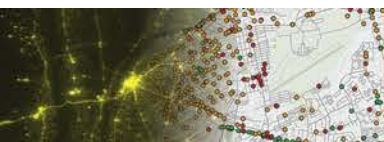
Introduction

- **Objectif principal:**
 - Guider les décisions politiques et programmatiques sur base des preuves solides
- **Avocat du diable**
 - Plusieurs types d'évidence (simple et moins-couteuse)
- **Argument**
 - En cas de doute sur l'efficacité d'un programme important, l'évaluation d'impact se justifie
 - Si nous n'avons pas de doute nous pouvons nous passer de l'évaluation d'impact



Pourquoi Est-ce important?

- **Permet de *mesurer* les impacts directs des politiques**
 - Est-ce que les conditions des bénéficiaires se sont améliorées suite au programme?
 - Comprendre les avantages nets du programme et comparer les différentes options du programme
 - Comprendre la répartition des effets du programme entre différents sous-groupes
- **Permet d'*obtenir* les impacts des politiques désirées**
 - Comprendre comment les résultats peuvent changer en modifiant certains aspects des politiques élaborées
- **Autres avantages opérationnels**
 - Décaissement plus rapide et moins de retards dans la mise en œuvre du projet



Deux approches

Suivi et
évaluation
(traditionnel)

Evaluation
d'impact

Suivi et évaluation (traditionnel)

- Mesurer les progrès des indicateurs au fil du temps
 - *Est-ce que le projet est sur une bonne voie?*
 - L'accent est mis sur les activités du projet et les bénéficiaires
 - Permet d'observer le changement, mais ne nous *dit pas pourquoi* ou *comment il* est arrivé.



mmm... QUAND J'UTILISE
MON PARAPLUIE, MES
SOULIERS SE MOUILLENT

Est-ce le parapluie qui
mouille mes souliers?

JE PENSE QUE JE
VAIS ARRETER
D'UTILISER MON
PARAPLUIE.

POURQUOI
COMPRENDRE LE LIEN
DE CAUSALITE EST-IL
IMPORTANT?

AFIN D'EVITER DE
ERREURS QUI PEUVENT
CAUSER DES DEGATS
ET GASPILLAGES

Il est très facile de confondre corrélation et causalité

Evaluation d'Impact

- **Impact**

- C'est la différence entre les résultats avec programme et sans programme
- *Qu'est-ce qui arriverait en l'absence du programme ou projet?*
- L'objectif de l'évaluation d'impact est de mesurer cette différence de manière à l'attribuer au programme, et **uniquement** le programme



Méthodes pour construire un contrefactuel valide

Faux Contrefactuels

Avant – Après

Participants – Non-participants

Impact causal sous certaines hypothèses & fortes limitations

Double difference (Diff-in-diff)

Appariement (matching, appariement de score de résultat)

Modèle de discontinuité

Impact causal

Méthodes expérimentales (randomization)

L'expérience parfaite

C'est l'impact de donner de l'argent de poche supplémentaire à Kami ?

Deux clones **identiques** à l'extérieur (observables) et **identiques** à l'intérieur (non observables)



Kami

Nous sommes tous les deux des marionnettes



Tami



Nous aimons tous les deux les bonbons



La seule différence est le projet ou traitement: *Donner de l'argent de poche supplémentaire à Kami*

L'expérience parfaite

Mettre en œuvre le programme (argent de poche supplémentaire à un des clones: Kami)



Kami

IMPACT = 6 – 4 = 2 Bonbons



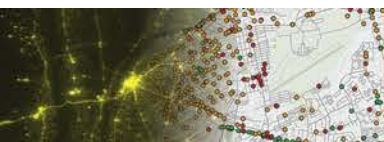
Tami



Comme Tami et Kami sont identiques (à l'interne comme à l'externe), la différence des résultats est due au programme

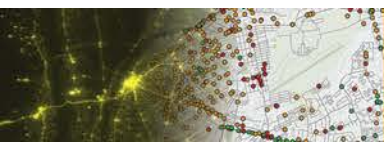
Rappel

- Besoin d'un contrefactuel valide pour estimer l'impact
 - Similarité des groupes dans les caractéristiques interne et externe
- Eviter les faux contrefactuels — ils produisent toujours des effets biaisés
- Les méthodes non-expérimentales reposent sur des hypothèses et des exigences en données fortes
- Or...les résultats de l'évaluation d'impact sont crédibles si nous utilisons des méthodes rigoureuses
- **Y-at-il une alternative plus efficace?**

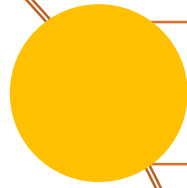


Méthodes Expérimentales

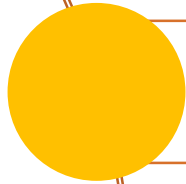
- **Autres appellations:**
 - Randomisation Expérience contrôlée (RCT) ou Assignation Aléatoire
- **C'est “l'étalon d'or”**
 - Comparé à d'autres techniques, les résultats des études expérimentales sont plus fiables et sont traités différemment par les bailleurs de fonds, les chercheurs, les décideurs politiques



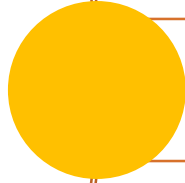
Menu



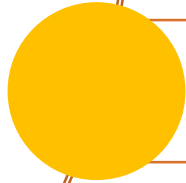
Comment randomiser?



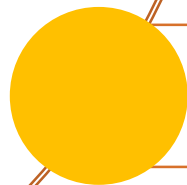
Autres opportunités de randomiser



Defis potentiels



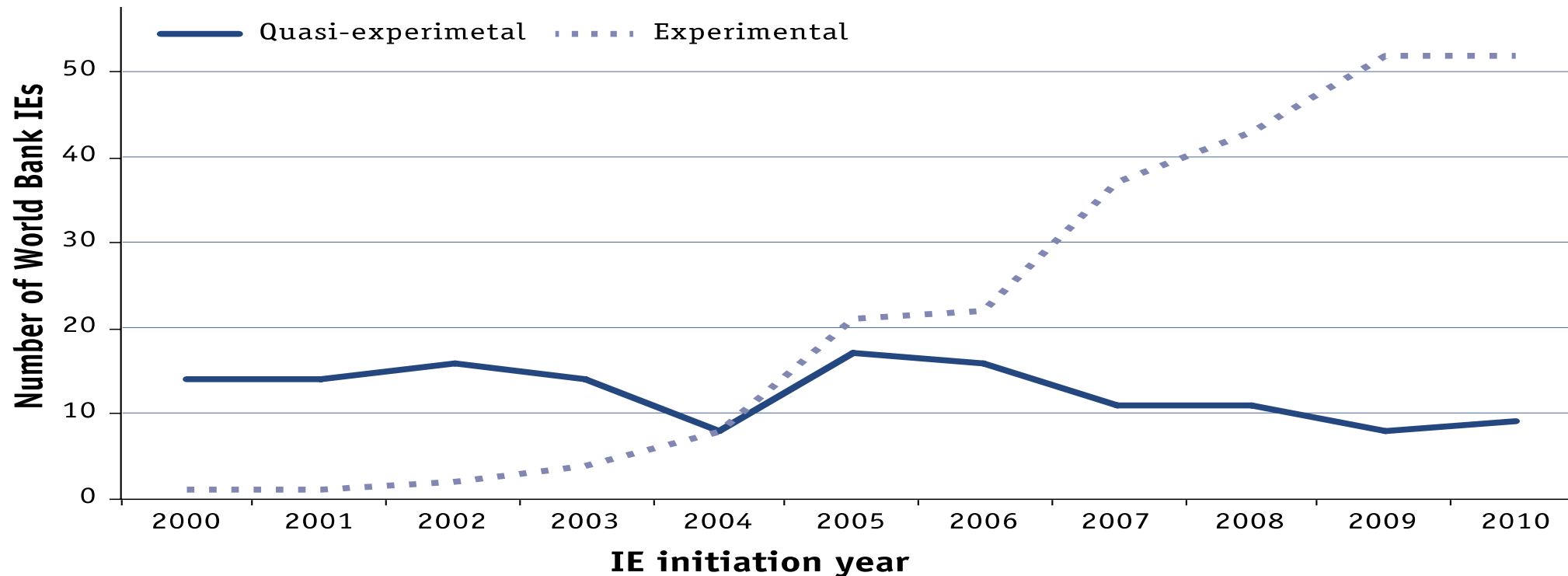
Pourquoi randomiser?



Modèle d'encouragement

Regain d'intérêt

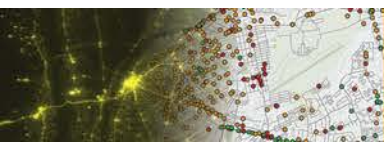
Impact Evaluations at the World Bank Using Experimental versus Quasi-Experimental Design by Initiation Year



Source: IEG.

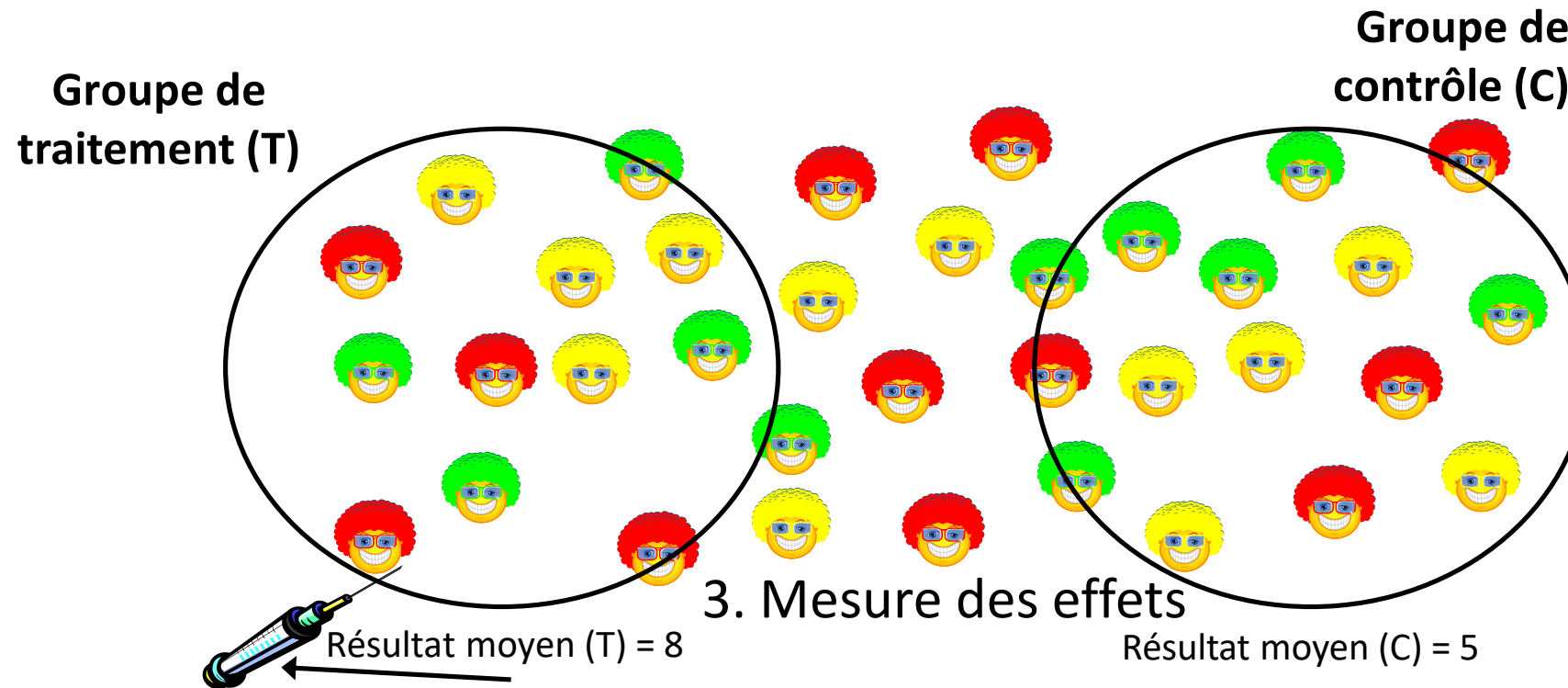
Comment randomiser (en pratique)?

- Identifiez votre population cible (ou groupe éligible au programme)
- Sélectionnez deux ou plusieurs groupes similaires (en moyenne) à travers la technique de randomisation
- Assignment aléatoire du programme à l'un des groupes
- Mesurer les résultats auprès des groupes de traitement et de contrôle pour dégager l'impact

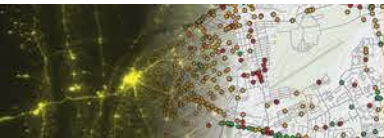


Comment randomiser (en pratique)?...

Population cible: tout le monde est éligible

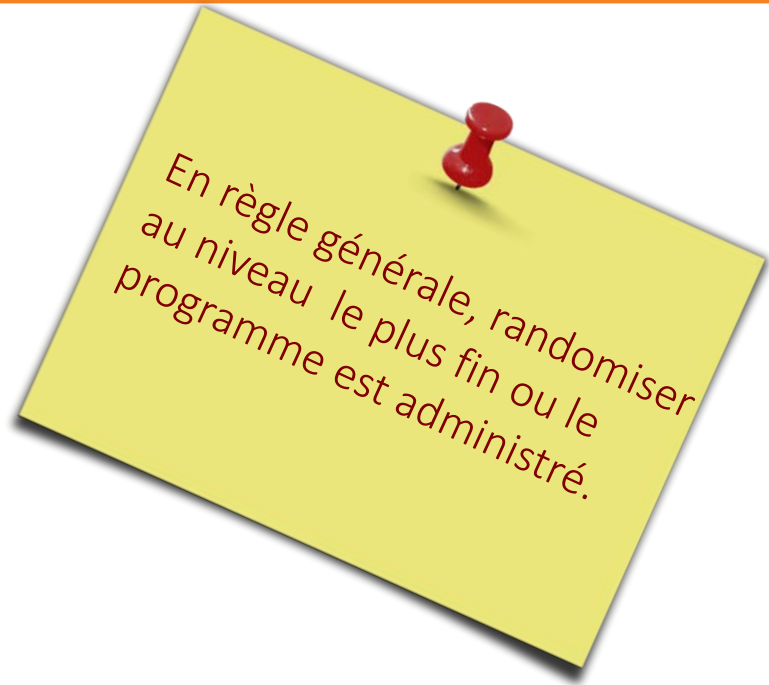


Impact Moyen (T - C): $8 - 5 = 3$



Unité de randomization

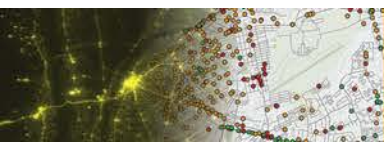
- **Choisissez l'unité d'observation selon le type de programme**
 - Individus/ménages
 - Ecoles/Cliniques/Zone de desserte ``catchment area``
 - Villages/communautés/quartiers
 - Régions/provinces/communes, etc.
- **A garder à l'esprit:**
 - Besoin d'un nombre d'unité ``*suffisamment large*`` pour détecter l'impact minimum désiré : **Puissance**.
 - Débordements (``Spillovers``) / Contamination
 - Coûts d'enquêtes et d'opérations



En règle générale, randomiser au niveau le plus fin ou le programme est administré.

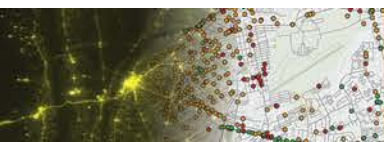
Modèle d'encouragement

- Il n'est pas toujours possible de randomiser
 - Raisons d'éthique et raisons politiques
 - La participation est volontaire, et tous sont éligibles
- Dans ce modèle, tout le monde peut bénéficier immédiatement du programme.
- Mais tout le monde n'est pas forcément au courant de l'existence du programme
 - Ex: Certaines municipalités ne savent pas qu'elles peuvent faire une demande pour bénéficier du programme d'entretien ou de modernisation de leur réseau routier
 - Incitatifs (petite rémunération ou récompense)
 - Transport (subvention des billets/tickets)
- Choisissez certaines municipalités/villages au hasard et encouragez-les à candidater
- Utilisez les autres comme contrôle/témoin
- **Est-ce éthique?**

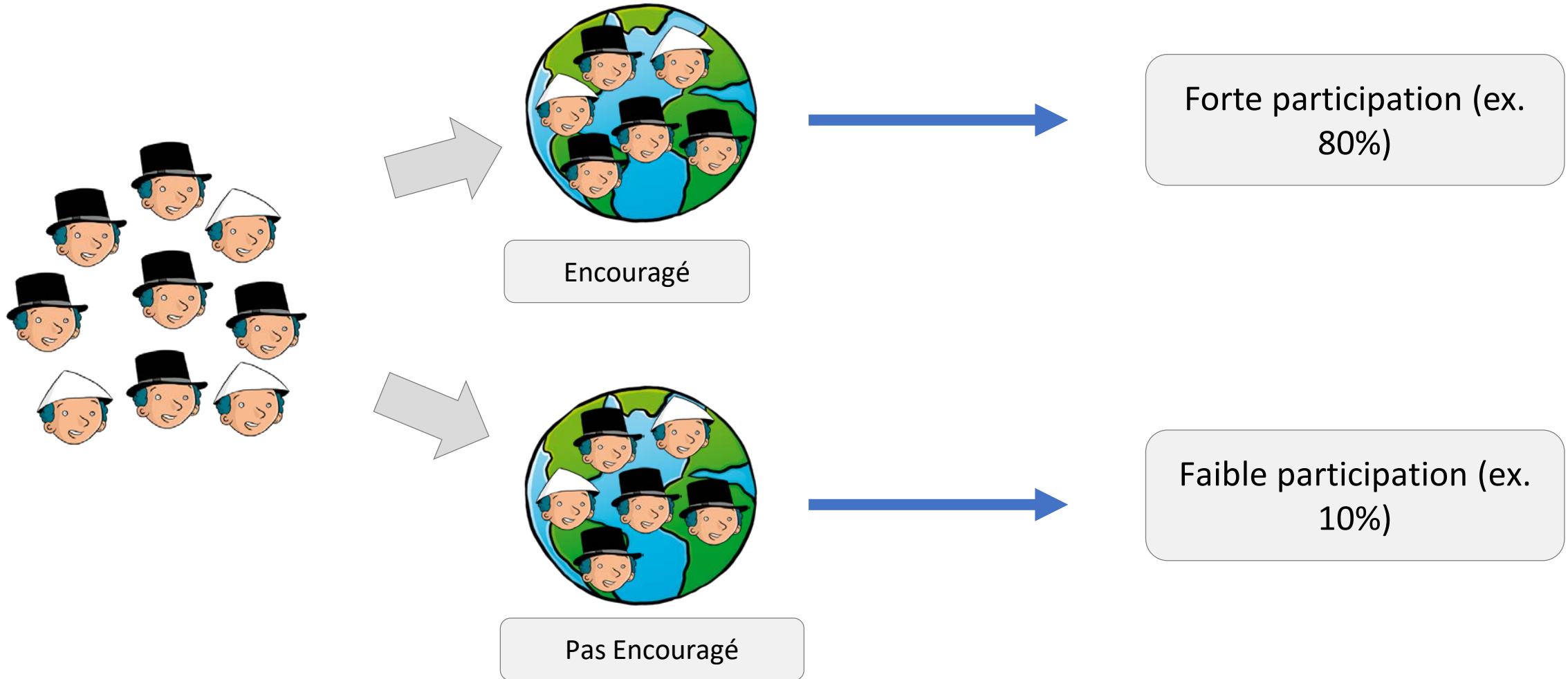


Modèle d'encouragement (incitation)

- Qu'est-ce qu'un « **encouragement** » ?
 - Un élément qui rend certaines personnes (groupe de traitement) plus susceptibles de participer que d'autres (groupe de contrôle)
- **Principaux critères:**
 - Ne doit pas être, en soi, un « traitement »
 - Mauvaise idée : programme de formation intense à l'entrepreneuriat, comme encouragement au crédit
 - Bonne idée : marketing ciblé vers certains, mais pas tous ; s'assurer que le marketing n'est pas trop informatif
- **Réfléchissez à ceux qui répondent à l'encouragement**
 - **Sont-ils différents ?**

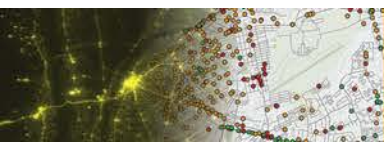


Modèle d'encouragement (1)



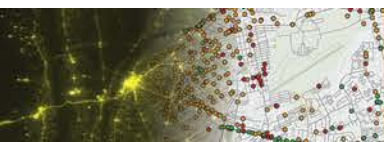
Etude de cas: Routes Rurales de l'Atlantis

- Est-ce que la modernisation des routes rurales contribue au développement économique local?
- **Contexte:**
 - **Problème:** Le Département des Transports de La République de Atlantis veut moderniser les routes rurales du Nord et de l'Est. L'objectif étant de réduire les coûts de transport de/et vers les zones agricoles afin de favoriser le développement économique local
 - **Objectifs principal:**
 - Promouvoir le Développement économique local



Etude de cas: Routes Rurales de l'Atlantis

- **Données complémentaires:**
 - **Janvier 2014:** les routes rurales sont classées en fonction de leur statut.
 - **2,000** villages avec des routes prioritaires sont sélectionnés et invités à candidater pour faire partir du programme. **1,021** répondent.
 - **Mars 2014 - Mars 2015 :** Le programme “**Revitalisation rurale - phase 1**” est mis en œuvre. Les **1,021** premiers villages voient leurs routes rurales être mises à niveau.
 - **Juin 2016:** Les ménages des villages avec routes prioritaires sont enquêtés et leur consommation mesurée.
- **La modernisation des routes rurales contribue-t-elle au développement économique local?**



Méthodes pour construire un contrefactuel valide

Faux Contrefactuels

Avant – Après

Participants –
Non-participants

**Impact causal sous certaines
hypotheses & fortes limitations**

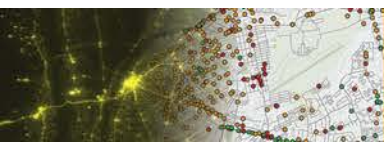
Double différence

Appariement

Discontinuité de la
régression

Impact causal

Méthodes
expérimentales
(randomization)

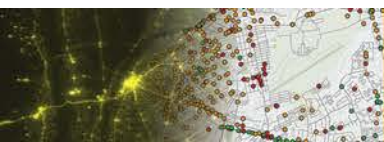


Synthèse: méthodes non-expérimentales

- Différentes méthodes produisent des résultats différents
- Les méthodes faibles peuvent conduire aux résultats biaisés
- Certaines méthodes non-expérimentales nécessitent des hypothèses fortes et sont exigeantes en matière de données de bonne qualité
- Les résultats de l'évaluation d'impact sont valides seulement si nous utilisons des méthodes rigoureuses.

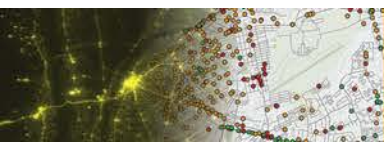
| Méthodes | Résultats (Impact estimé) |
|-------------------|---------------------------|
| Simple différence | 82.5* |
| Appariement | 55.8* |
| Double différence | 27.1* |

Note: * statistiquement significatif au seuil de 5%



Essais Contrôlés Randomisés (RCTs)

- Le département de transport choisit aléatoirement 4,000 villages
- Toutes les routes doivent être réhabilitées, cependant
 - Le Département n'a aucune information sur l'ordre de priorité: **quel village devrait prioritairement bénéficier du programme?**
- Dans un pareil cas, l'assignation aléatoire de l'ordre de traitement est possible
- Des chercheurs réussissent à convaincre le Département Randomiser l'ordre dans lequel les villages reçoivent le programme
 - il sera ainsi possible d'isoler l'impact du programme de revitalisation des routes
- Il accepte et décide d'assigner aléatoirement la moitié des villages au programme (**Traitement**) et l'autre moitié est reportée pour plus tard (**contrôle**)



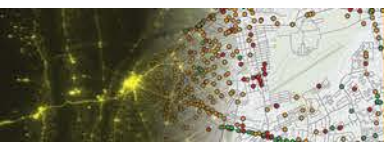
Assignation Aléatoire (Randomization)

- Le choix aléatoire des villages devant prioritairement bénéficier du programme
- En moyenne les villages de contrôle et de traitement auront des caractéristiques similaires (observables et non-observables)
 - la participation ou non au programme est purement aléatoire. (Pas besoin d'habiletés pour appliquer)
- Seule différence entre les villages: la participation ou non au programme
 - la comparaison entre les villages de contrôle et de traitement après l'intervention va nous donner l'impact causal

*4,000 ménages
(Echantillon éligible)*

*2,000 ménages
(Groupe cible)*

*2,000 ménages
(Groupe de contrôle)*



Données : Caractéristiques des Villages

Avant l'intervention, il n'y a pas de différence significative entre les deux groupes

| | Traitement | Contrôle | Différence |
|------------------------------|------------|----------|------------|
| Nombre d'utilisateurs | 37.85 | 38.27 | 0.42 |
| Densité de population | 10.20 | 10.89 | -0.69 |
| Marché | 0.34 | 0.37 | -0.03 |
| Nombre d'enfants par ménages | 5.05 | 5.03 | 0.02 |
| Diversification (%) | 25.51 | 25.67 | -0.16 |
| Taille de l'échantillon | 2,000 | 2,000 | |

Consommation moyenne par habitants après l'intervention

| | Traitement | Contrôle | |
|---------------------------|------------|----------|--------------|
| Consommation moyenne/hbts | 146.2 | 121.3 | 24.9* |

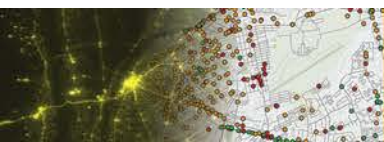
Note: * statistiquement significatif au seuil de 5%

Expérience Aléatoire Contrôlées (RCTs)

- L'impact causal peut-être attribué au programme d'entretien routier
- Comparaison avec les autres méthodes: **DIFFÉRENCES ÉNORMES**

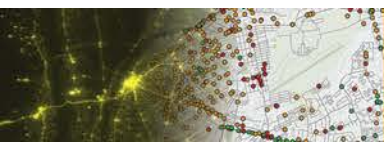
| Méthodes | Résultats (Impact estimé) |
|-------------------|---------------------------|
| Simple différence | 82.5* |
| Appariement | 55.8* |
| Double différence | 27.1* |
| Randomisation | 24.9* |

- **Problèmes potentiels:**
 - Généralisation
 - Menaces a la validité interne
 - Validité externe



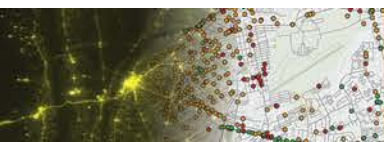
Etude de cas 2: Politique Complémentaire

- Randomiser les programmes routiers (ou l'ordre) n'est pas toujours évident
- Comment randomiser une politique complémentaire (qui a lieu en même temps que l'intervention)?
- En collaboration avec le ministère des affaires économiques et sociales,
- **Objectif:** Maximiser l'impact de l'investissement en transport en favorisant la diversification des revenus des femmes rurales



Etude de cas 2: Politique Complémentaire

- **Intervention:** Tester si une subvention (outil) de transport peut aider les femmes à écouler plus facilement les produits de leur récolte
- 900 femmes sont choisies aléatoirement
 - 450 femmes reçoivent une moto
 - 450 autres ne reçoivent pas
- Comme précédemment, le choix des femmes devant bénéficier du programme est aléatoire
- Puisque la seule différence entre ces *femmes est la participation ou non au programme de subvention*, alors la comparaison entre ces femmes après l'intervention va nous donner **l'impact causal**

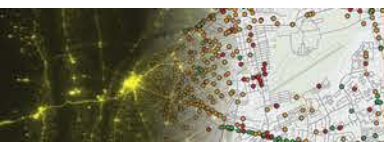


Données : Caractéristiques des femmes

Avant l'intervention, il n'y a pas de différence significative entre les deux groupes

| | Traitement | Comparaison | Différence |
|---|------------|-------------|------------|
| Nombre de semaines travaillées au champ | 38.14 | 38.08 | 0.06 |
| Nombre de kilomètre moyen vers le marché le plus proche | 10.52 | 10.70 | -0.18 |
| Marché Local [1= Yes] | 0.34 | 0.37 | -0.03 |
| Nombre d'enfants par ménage | 5.02 | 4.90 | 0.12 |
| Cout moyen de la moto | 24.64 | 25.20 | -0.56 |
| Taille de l'échantillon | 450 | 450 | |

Les groupes de traitement et de comparaison dans ce tableau semblent similaires. **Est-ce à cela que vous vous attendiez** ? Pourquoi? Quels sont les différences entre ce tableau et le tableau précédent? (Appariement et simple différence?)



Données : Caractéristiques des femmes

Avant l'intervention, il n'y a pas de différence significative entre les deux groupes

| | Traitement | Comparaison | Différence |
|---|------------|-------------|------------|
| Nombre de semaines travaillées au champ | 38.14 | 38.08 | 0.06 |
| Nombre de kilomètre moyen vers le marché le plus proche | 10.52 | 10.70 | -0.18 |
| Marché Local [1= Yes] | 0.34 | 0.37 | -0.03 |
| Nombre d'enfants par ménage | 5.02 | 4.90 | 0.12 |
| Cout moyen de la moto | 24.64 | 25.20 | -0.56 |
| Taille de l'échantillon | 450 | 450 | |

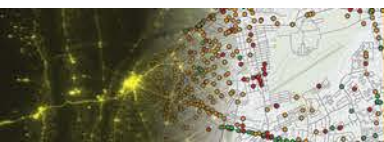
Consommation moyenne par habitants après l'intervention

| | Traitement | Comparaison | |
|---------------------------|------------|-------------|---------------|
| Consommation moyenne/hbts | 86.17 | 65.30 | 20.87* |

Menaces à la validité interne?

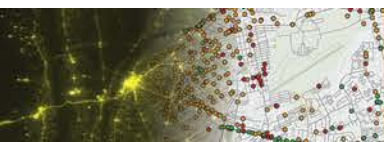
• Menaces à la validité interne?

- Faible taux d'adhésion
- Problèmes de conformité
- Perte de participants
- Débordement ou contamination



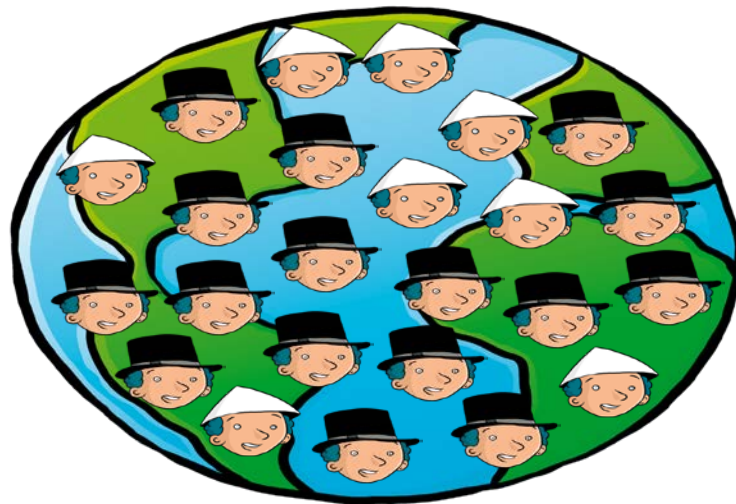
Validité externe

- **Les résultats sont-ils généralisables?**
 - L'impact estimé est causal, mais seulement pour la population cible ou de même caractéristiques
 - Généraliser les résultats au delà de la population cible nécessite la réplication de l'évaluation
 - Ou alors...faire l'évaluation sur une population représentative



Sélection aléatoire \neq assignation aléatoire

1. Population



Validité Externe

2. Echantillon d'Evaluation



Validité Interne

3. Assignation Aléatoire



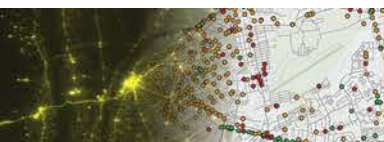
Contrôle



Traitement

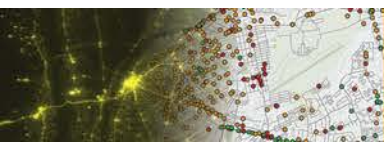
Défis Potentiels

- Contraintes liées à la randomisation dans le secteur de transport?
- Echantillonnage?
- Contraintes d'éthiques et de politiques
 - Mais il y a des mesures d'atténuation
- Défi de puissance
 - Besoin d'avoir un échantillon suffisant



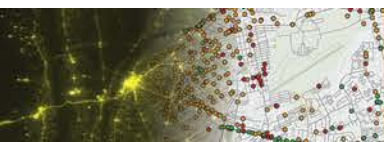
Randomisation dans le transport?

- **Contraintes liées à la randomisation dans le secteur de transport: Bon contrefactuel?**
 - Que ce serait-il passé si les investissements avaient été distribués différemment?
- Ne doit pas être perçu comme inéquitable
- Doit être éthique
 - Si on est sûr qu'un programme marche; est-ce éthique d'avoir un groupe de contrôle?
- Doit être politiquement faisable
 - Est-ce que les autorités acceptent l'idée de l'assignation aléatoire?



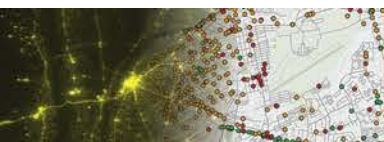
Biais de sélection

- Qu'est ce que le biais de sélection (auto-sélection)?
 - C'est l'erreur introduite par le fait que les participants à une intervention ne sont pas représentatifs de la population en général



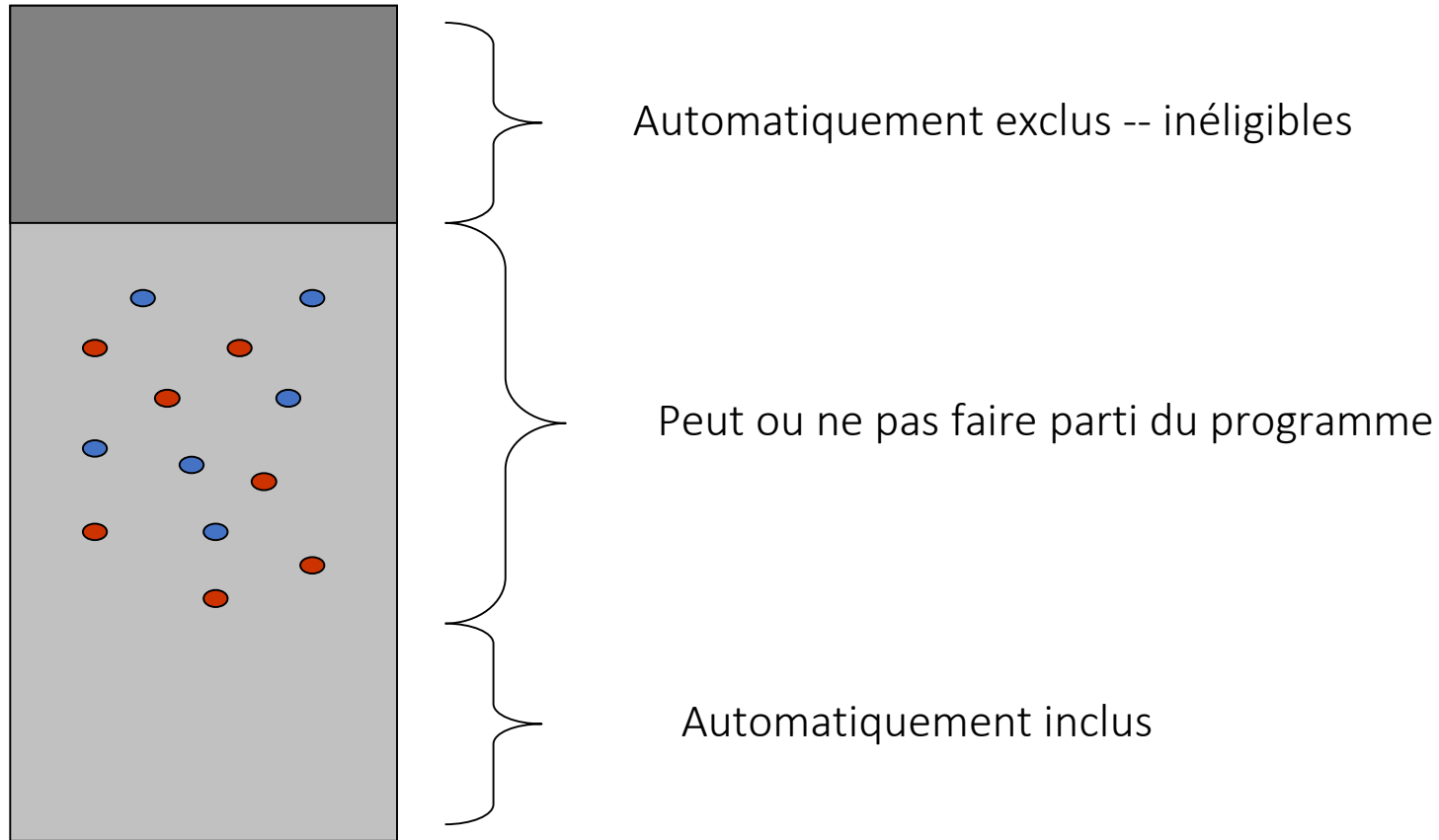
Expériences Contrôlées et Echantillon

- Combien d'unités (taille de l'échantillon) est-il nécessaire pour chaque groupe?
 - Exemple: Combien de villages et femmes doit-on choisir pour les groupes de contrôle et de traitement?
- En pratique: Ce sont les experts statisticiens qui aident au calcul de la taille de l'échantillon
- Formule de la taille de l'échantillon N:
$$N = \left[\frac{4\sigma^2(z_{\alpha/2} + z_{\beta})^2}{D^2} \right] [1 + \rho(H - 1)]$$
- **Ingrédients fondamentaux**
 - Quel est l'**impact minimum** qui justifierait d'investir dans le programme ?
 - Quelle est la **variance** de l'indicateur de résultat qui vous intéresse ?
 - Le programme génère-t-il des **grappes**?



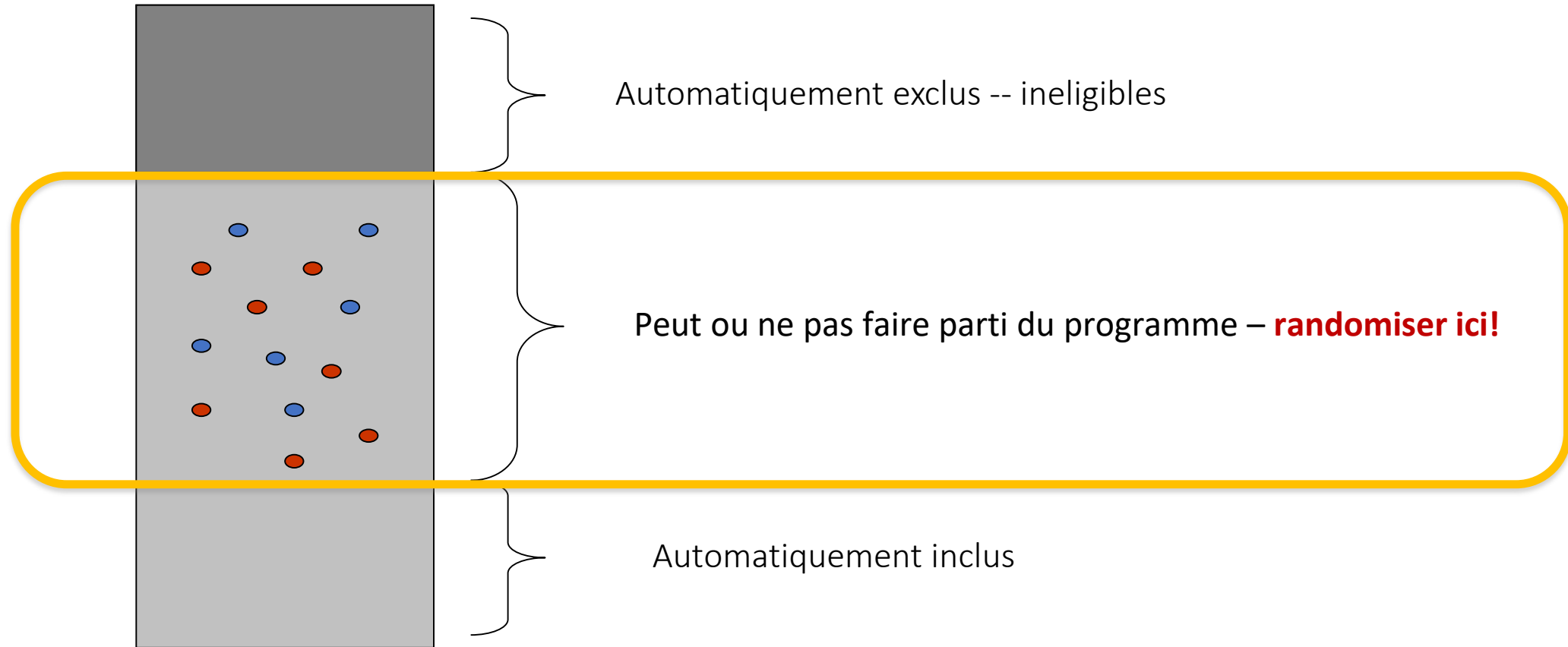
Ethiques de la selection aléatoire

Le programme cible les communautés/villages démunis afin qu'ils puissent rattraper



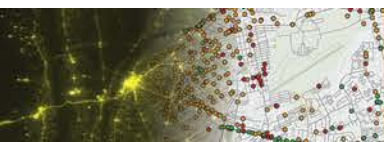
Ethiques de la selection aléatoire

Le programme cible les communautés/villages démunis afin qu'ils puissent rattraper



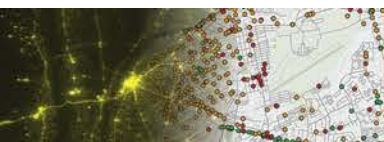
Défi de puissance

- Evaluation de l'impact: **comparer** deux échantillons
- Grands échantillons: plus de précision, mais les ressources sont limitées
 - Les grands échantillons permettent d'estimer plus précisément les caractéristiques de la population
- **Petits échantillons**: risque de fausses conclusions
 - Des petits échantillons créent des risques de tirer des conclusions politiques erronées
- Solution: **Calculs de puissance**
- Voir présentation de Serge Demain!



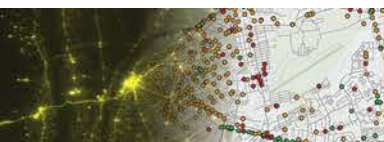
Défi de puissance (1)

- Les calculs de puissance nous informent sur la taille requise pour les échantillons.
- Des échantillons plus grands sont nécessaires pour estimer précisément un impact si nous nous attendons à un impact faible, ou si le programme génère des grappes,...
- Les calculs peuvent être réalisés dans de nombreux logiciels statistiques: STATA, R, Optimal Design
- Si vous avez besoin de faire des calculs de puissance statistique: contactez un spécialiste d'échantillonnage...



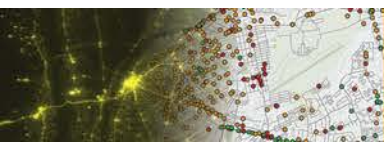
Avant de pouvoir vous lancer, qui devez-vous convaincre?

- Vos supérieurs hiérarchiques
- Le Ministre
- Les autorités locales
- Vos collègues et partenaires (ceux qui vont mettre en place le programme)
- Le comité d'éthique
- Les participants du groupe témoin



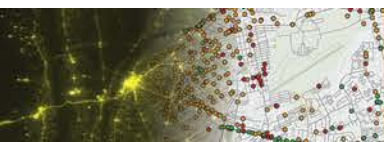
Comment Convaincre les Praticiens?

- Trois éléments essentiels :
 - Connaître l'identité des principaux acteurs et leurs points de vue
 - Ecouter. Trouver des manières de répondre aux questions *opérationnelles* tout en effectuant la recherche
 - Importance de la culture
- Il est parfois difficile d'être en désaccord en personne. Y a-t-il une véritable acceptation?



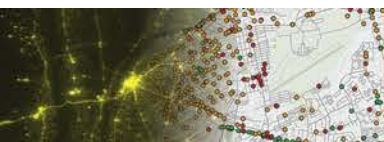
Une Bonne Evaluation d'Impact

- Répond non seulement à la question « **Est-ce que le programme marche ?** »
- Mais aussi aux questions « **Pourquoi ca marche (ou pas) et comment pourrait-il marcher mieux ?** »
- Pour que votre évaluation soit acceptable et utile, il est important:
 - D'identifier les questions que se posent les acteurs/partenaires
 - De vous assurer que l'évaluation y réponde



Conclusion

- L'inférence causale sur base du raisonnement contrefactuel permet l'évaluation d'impact
- La randomisation est une meilleure méthode de construction d'un bon contrefactuel
 - Elle est simple, directe et donne une estimation non biaisée de l'impact
- Beaucoup de possibilités pour la randomisation
 - Déploiement Échelonné
 - Traitements Multiples
- D'autres méthodes existent, mais elles reposent sur plus d'hypothèses et des techniques plus sophistiquées



Calendrier de l'Evaluation d'impact

Conception De
l' EI

- Identification des problèmess et questions importantes
- Identification des innovations à tester

Enquête de
base

- Administration des questionnaires aux groupes de contrôle et traitement
- Analyse des données de référence

Enquête
mi-parcours

- Administration des questionnaires aux groupes de contrôle et traitement
- Résultats préliminaires

Enquête finale

- Administration des questionnaires aux groupes de contrôle et traitement
- Résultats finaux