

LES RISQUES SUR LA VALIDITÉ

Bruno Crépon
CREST / J-PAL

22 -26 juin 2015

Abidjan, Côte d'Ivoire

PLAN DE LA SEMAINE

1. Théorie du changement et mesure
2. Pourquoi tirer au sort ?
3. Comment tirer au sort ?
4. Méthodes non expérimentales
5. Risques sur la validité
6. Intégrer une évaluation dans un projet
7. Une évaluation de A à Z



INTRODUCTION

- L'évaluation d'impact s'appuie sur l'idée de comparer un groupe recevant un programme à un groupe de contrôle **identique** ne le recevant pas
- L'assignation aléatoire est la méthode la plus fiable pour construire un groupe de contrôle identique
- Mais un certain nombre de menaces peuvent biaiser la validité de l'évaluation
- Il faut distinguer les risques liés à la **validité interne** et ceux liés à la **validité externe**

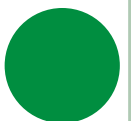
PLAN DU COURS

○ **Menaces sur la validité interne**

- Data mining
- Attrition
- Externalités
- Conformité partielle (ITT / ToT)

○ **Menaces sur la validité externe**

- Réponses comportementales aux évaluations
- Généralisation des résultats



DATA MINING

- Sans même mentionner les rares cas de fraude...
- ...le data mining (= trouver « quelque chose » à tout prix) est un problème

=> Plus vous examinez de variables, plus vous avez de chances d'en trouver au moins une affectée de façon significative par le programme

=> Enregistrement et Plans de Pré-Analyse



LE AEA RCT REGISTRY

AEA RCT Registry

<https://www.socialscienceregistry.org/site/about>



AEA RCT Registry

The American Economic Association's registry for randomized controlled trials

[Create Account](#) [Sign in](#)

[About RCTs](#) [Registration Guidelines](#) [FAQ](#)

Advanced Search

SEARCH

[REGISTER A TRIAL >](#)

ABOUT THE REGISTRY

Welcome.

This is the American Economic Association's registry for randomized controlled trials.

Randomized Controlled Trials (RCTs) are widely used in various fields of economics and other social sciences. As they become more numerous, a central registry on which trials are on-going or complete (or abandoned) becomes important for various reasons: as a source of results for meta-analysis; as a one-stop resource to find out about available survey instruments and data.

Because existing registries are not well suited to the need for social sciences, in April 2012, the AEA executive committee decided to establish such a registry for economics and other social sciences.

If you are running or have run a trial: Registration is free and you do not need to be a member of the AEA to register. We encourage you to register any new study at its outset. However, given the backlog of existing trials, we invite you to also register past studies.

If you are searching for results: Please browse the data base. More results are forthcoming!

© Copyright 2012-2015, MIT.

[About](#) [FAQ](#) [Contact](#)

LES PLANS DE PRÉ-ANALYSE

- Surtout utiles quand:
 - Plusieurs façons de mesurer un résultat
 - Beaucoup de sous-groupes
- Mais des inconvénients:
 - Quid des résultats inattendus?
 - Comment s'adapter?
- ⇒ PPA conditionnels... et leur coût
- ⇒ Un choix laissé à chaque équipe de recherche



PLAN DU COURS

○ **Menaces sur la validité interne**

- Data mining
- **Attrition**
- Externalités
- Conformité partielle (ITT / ToT)

○ **Menaces sur la validité externe**

- Réponses comportementales aux évaluations
- Généralisation des résultats



ATTRITION

- Est-ce un problème si certaines personnes de l'expérimentation disparaissent avant que vous ayez recueilli les données ?
- Des exemples ?
- Pourquoi est-ce un problème si le profil des personnes qui disparaissent est lié au programme ?



BIAIS D'ATTRITION : UN EXEMPLE

- Certains enfants ne vont pas à l'école parce qu'ils sont trop faibles (sous-alimentés)
- Vous lancez un programme de distribution de repas à domicile
 - Vous avez un groupe test et un groupe témoin
- Les enfants faibles et chétifs commencent à aller plus à l'école s'ils vivent près d'une école test
- Premier effet de votre programme : l'assiduité des élèves augmente
- Vous voulez également mesurer son impact sur la croissance des enfants :
 - Deuxième résultat intéressant : le poids des enfants
- Vous vous rendez dans toutes les écoles (tests et témoins) et vous pesez tous les présents un jour donné
- Est-ce que la différence de poids entre test et témoin sera sur ou sous estimée ?



BIAIS D'ATTRITION: UN EXEMPLE

	Avant le programme			Après le programme	
	Test	Témoin		Test	Témoin
	20	20		22	20
	25	25		27	25
	30	30		32	30
Moyenne	25	25		27	25
	Différence	0		Différence	2



BIAIS D'ATTRITION: UN EXEMPLE

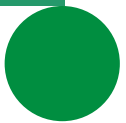
Et qu'en est-il si seuls les enfants de plus de 20 kilos viennent à l'école ?



BIAIS D'ATTRITION : UN EXEMPLE

Et qu'en est-il si seuls les enfants de plus de 20 kilos viennent à l'école ?

	Avant le programme			Après le programme		
	Test	Témoin		Test	Témoin	
	[absent]	[absent]		22	[absent]	
	25	25		27	25	
	30	30		32	30	
Moyenne	27,5	27,5		27	27,5	
	Difference	0		Difference	-0,5	



BIAIS D'ATTRITION

- Prévoir des fonds pour être en mesure de garder la trace des participants quand ils quittent le programme
- S'il reste encore de l'attrition, vérifiez qu'il n'y a pas de différence entre les groupes tests et témoins. Est-ce suffisant ?
- Vérifiez également que ce n'est pas lié aux observables: balancing test sur les répondants
- Essayez de délimiter l'étendue du biais :
 - Supposez que chaque personne qui a abandonné le groupe test a eu le score le plus faible et supposez que chaque personne qui a abandonné le groupe témoin a eu le plus élevé...



COMMENT TROUVER TOUT LE MONDE ?

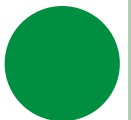
- Informations initiales :
 - Recueillir autant d'informations que possible sur les différents contacts
- Tirez profit des réseaux
- Mais les adresses sont souvent folkloriques...

- Les téléphones portables sont une invention formidable !
- Essayez de trouver les migrants où qu'ils soient
 - engagez un Sherlock Holmes (et même plusieurs)
 - soyez prêt à dépenser beaucoup d'argent (essence, salaire des recenseurs, ...) : cela peut revenir très cher !
- Quelques pistes :
 - Recherches intensives seulement sur des sous-échantillons tirés au sort
- Et même si vous les trouvez, ils ne seront peut-être pas disposés à fournir des informations !



D'AUTRES CAUSES D'ATTRITION

1. Les non réponses : refus de participer à l'enquête
2. Les ménages temporairement absents
3. Les mauvaises informations sur les adresses initiales
 - 3 devrait être équilibré entre témoin et test
 - 1 et 2 peuvent en revanche être déséquilibrés



PLAN DU COURS

○ **Menaces sur la validité interne**

- Data mining
- Attrition
- **Externalités**
- Conformité partielle

○ **Menaces sur la validité externe**

- Réponses comportementales aux évaluations
- Généralisation des résultats



EXTERNALITÉS

Les programmes ont des effets directs et indirects, aussi appelés “externalités” (“spillovers”). Il y en a différents types:

- Physiques : deparasitage, malaria
 - D'information : on donne de l'information aux membres du groupe traitement.... qui la diffuse
 - Comportementaux : effets de pairs
 - D'équilibre général : Cash transfert dans des villages.
Effet d'entraînement économique
- => Peuvent être positifs ou négatifs
- => Rendent difficile ou impossible de mesurer l'impact du programme



LE PROBLEME DES EXTERNALITÉS?

Les externalités peuvent biaiser la mesure de l'impact pour les bénéficiaires : si le groupe contrôle est affecté par exemple

Les externalités sont souvent faibles et le groupe de contrôle n'est que peu affecté, l'effet sur les bénéficiaires est bien mesuré

Mais les externalités s'appliquent à un très grand groupe: l'impact des externalités sur chaque individu est faible mais sur la population est grand

L'impact du programme dans son ensemble, pas seulement sur les bénéficiaires directs, peut être fortement biaisé même si les externalités sont faibles



COMMENT GÉRER LES EXTERNALITÉS?

1/ Les éviter

⇒ En jouant sur le niveau de randomization. Par exemple information à des élèves: l'unité de randomisation peut être l'individu, la classe, l'école...

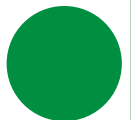
2/ Les mesurer

⇒ cf l'exemple suivant



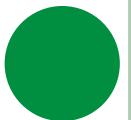
EXEMPLE : INFORMATION SUR LES PRIX

- Fournir aux fermiers des informations sur les prix futurs par téléphone portable
- Est-ce qu'on doit s'attendre à des externalités?
- Tirage au sort : à l'échelle des individus ou des villages ?
 - Tirage au sort au niveau des villages
 - Puissance statistique faible
 - « des groupes témoins plus purs »
 - Tirage au sort au niveau des individus
 - Plus de puissance statistique (si les externalités sont petites)
 - Mais un paramètre biaisé si il y a des externalités



EXEMPLE : INFORMATION SUR LES PRIX

- Est-ce qu'on peut faire les deux et mesurer les externalités ?
- Tirer au sort pour répartir les villages en 4 groupes(A, B, C, et D)
- Villages du groupe A
 - SMS avec des informations sur les prix à 50% des individus possédant un téléphone portable
- Villages du groupe B
 - Pas de SMS avec des informations sur les prix
- Deux « groupes contrôle »
 - G1 : Village B
 - G2 village A ne recevant pas l'information
- Un groupe traitement T village A recevant l'information
- Comparer T et G1 : mesure du vrai effet
- Comparer G2 et G1 mesure de l'externalité



PLAN DU COURS

○ **Menaces sur la validité interne**

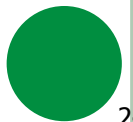
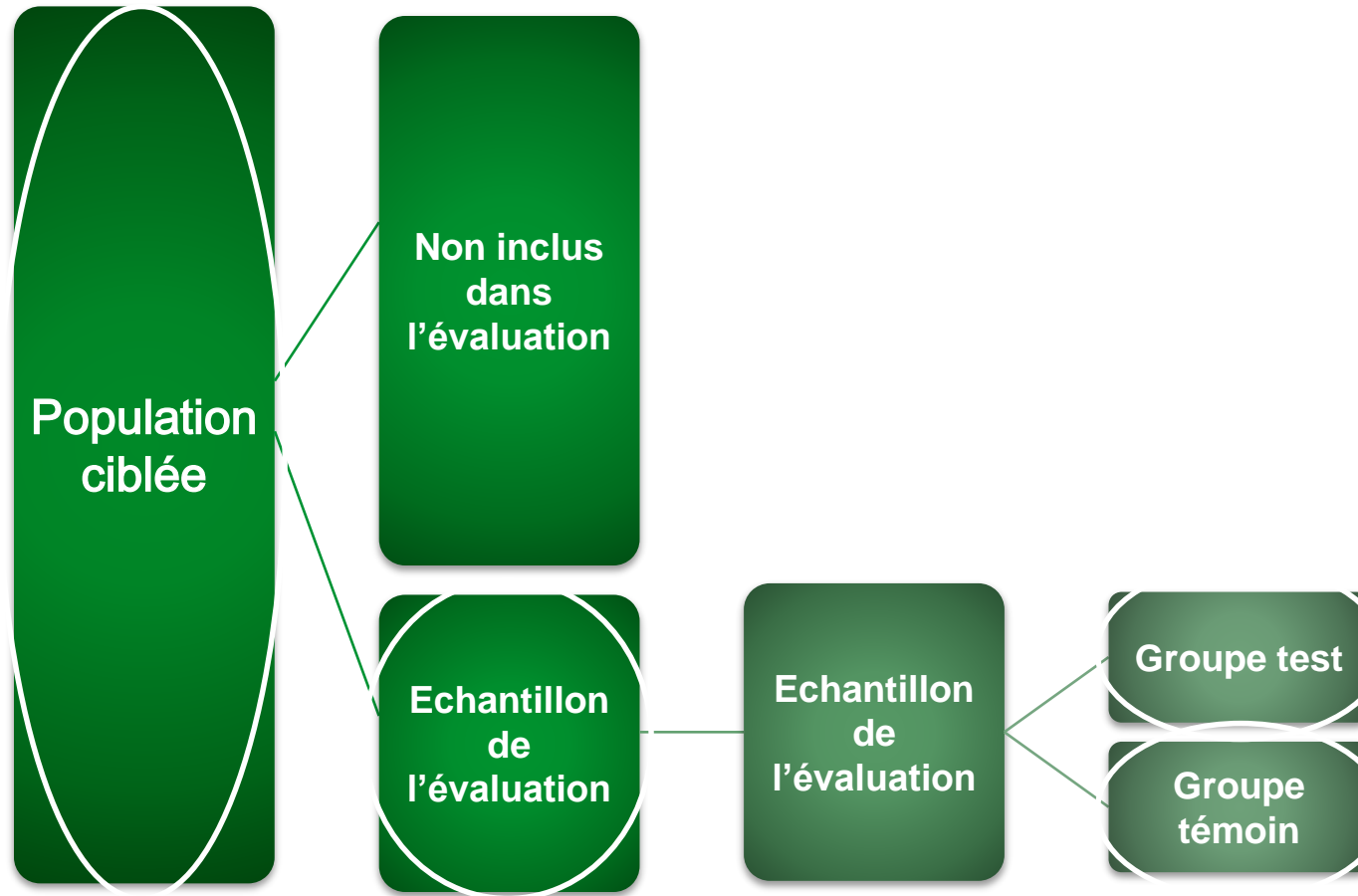
- Data mining
- Attrition
- Externalités
- **Conformité partielle (ITT / ToT)**

○ **Menaces sur la validité externe**

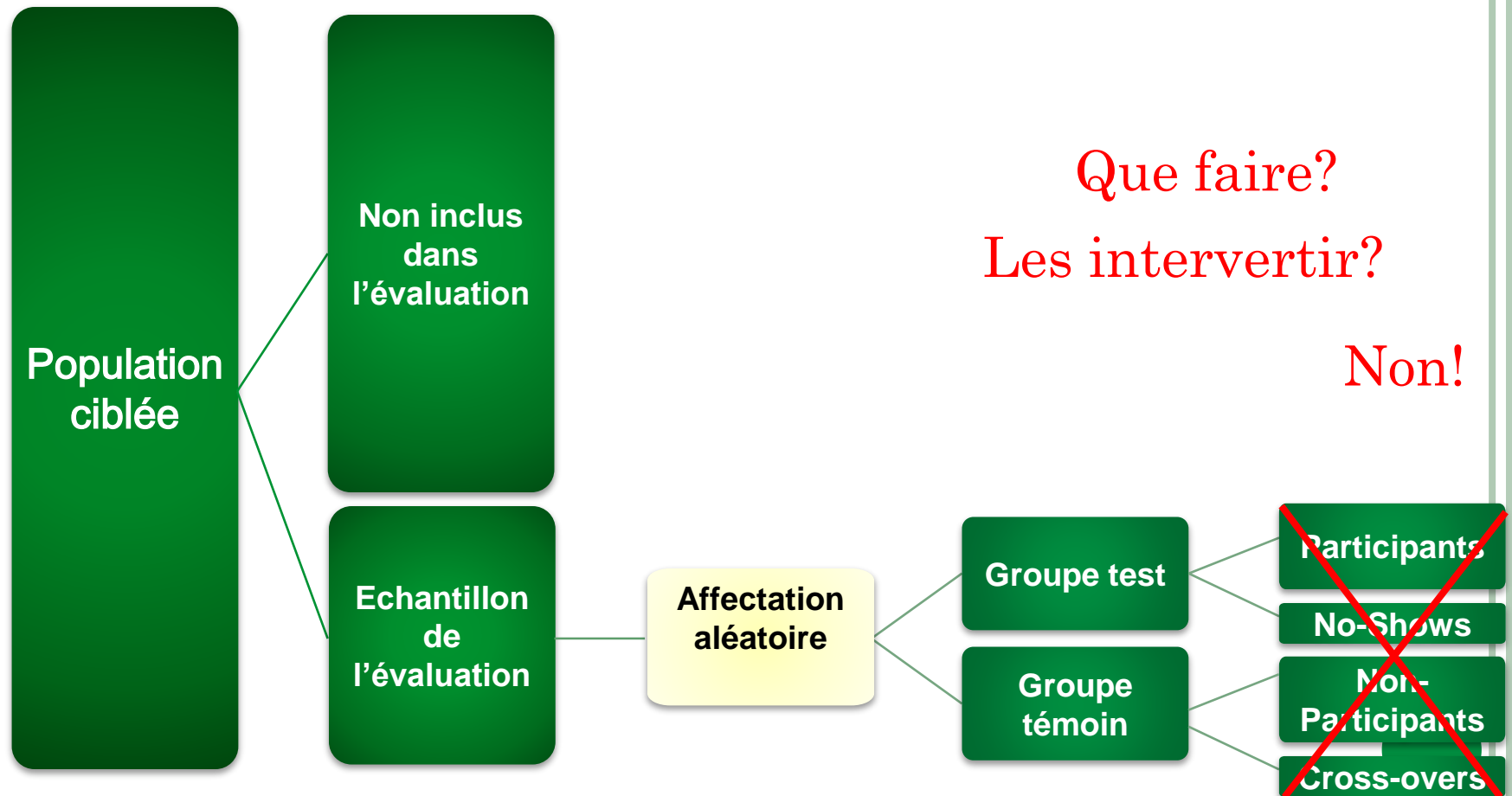
- Réponses comportementales aux évaluations
- Généralisation des résultats



PROTOCOLE DE BASE DE L'ÉVALUATION ALÉATOIRE



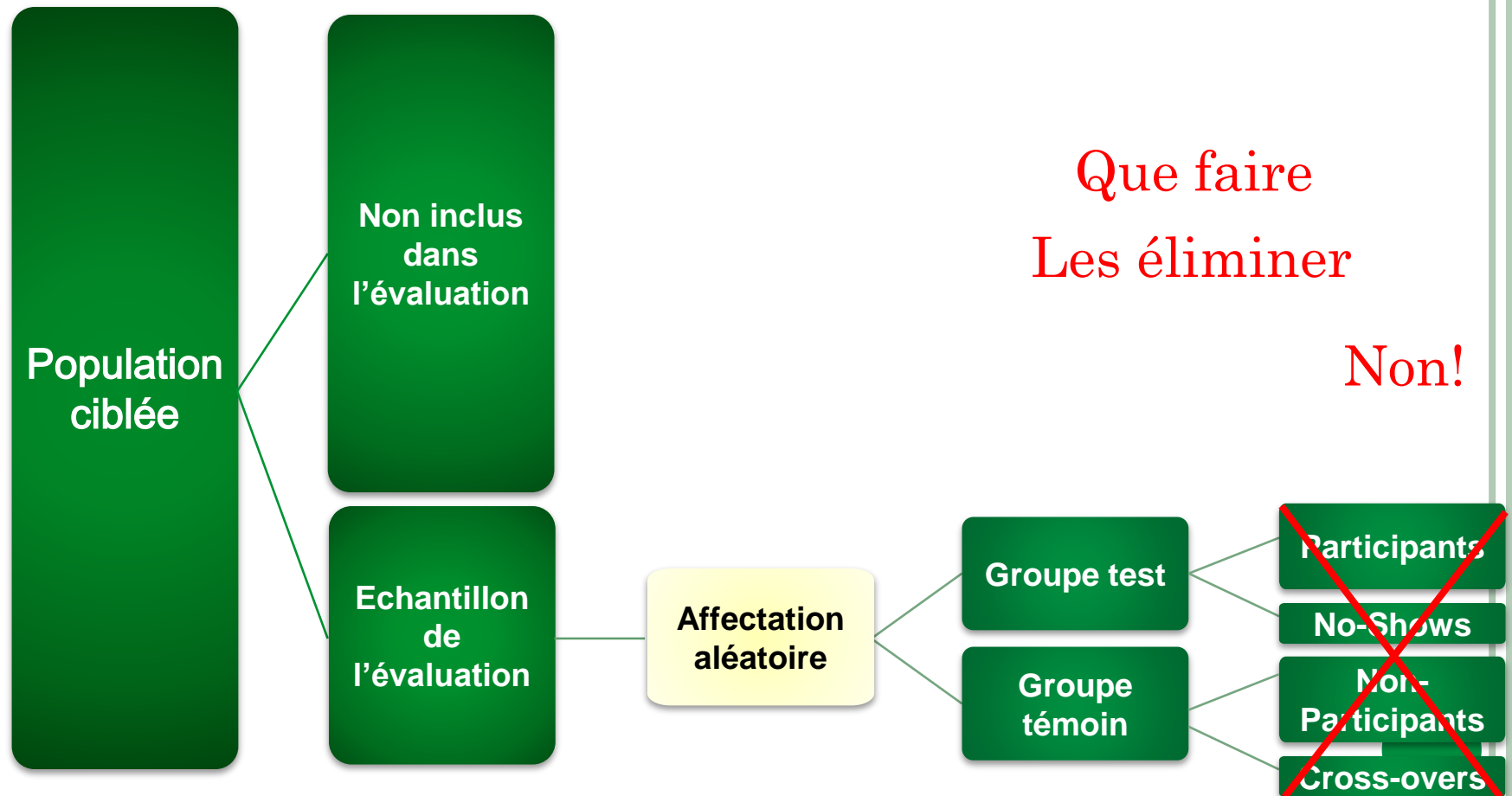
CONFORMITÉ IMPARFAITE (NON COMPLIANCE)



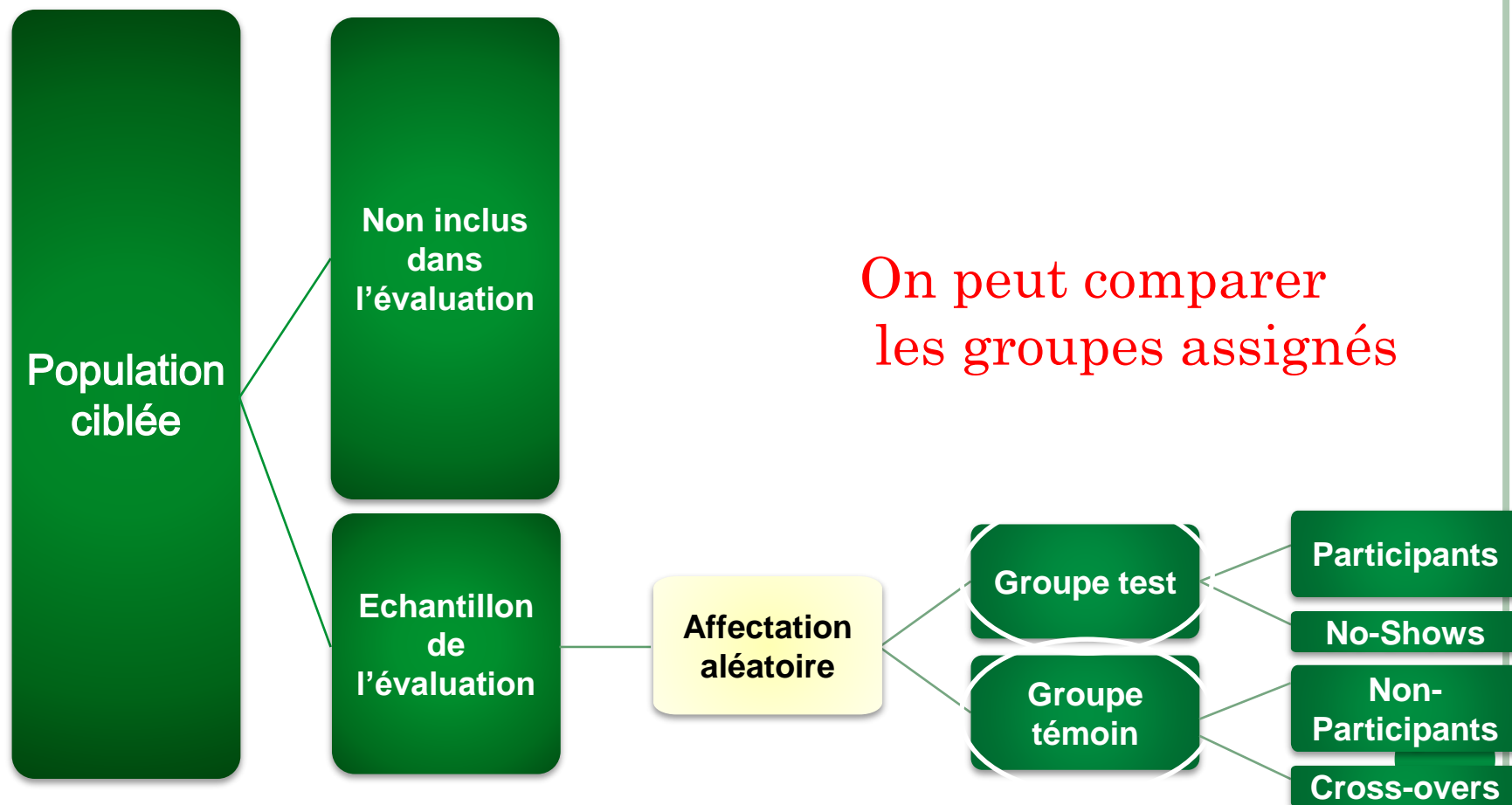
Que faire?
Les intervertir?

Non!

CONFORMITÉ IMPARFAITE (NON COMPLIANCE)



CONFORMITÉ IMPARFAITE (NON COMPLIANCE)

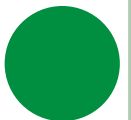


CONFORMITÉ PARTIELLE

- Certains individus assignés au groupe témoin peuvent aller dans le groupe test
- Et parallèlement, des individus affectés au groupe test peuvent ne pas recevoir le programme

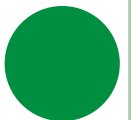
On considère deux types de paramètres:

- ITT: Intention to Treat
- ToT: Treatment on the Treated



ITT / ToT

- L'Intention to Treat ne se préoccupe pas de la conformité partielle: on mesure **l'impact du fait d'avoir été assigné** au programme
- Le Treatment on the Treated permet de palier le problème de la conformité partielle et de mesurer **un impact du programme mais... sur une sous population** particulière: ceux qui l'ont effectivement reçu à cause de l'assignation
- Idée de la population des « compliers »



ESTIMER LE TOT ET L'ITT

- De quoi avons nous besoin
- $Y(AT)$ la moyenne sur le groupe Assigné au groupe Test (AT)
- $Y(AC)$ la moyenne sur le groupe Assigné au groupe Controle (AC)

- $\text{Prob}[T | AT]$ = Proportion de traités dans AT
- $\text{Prob}[T | AC]$ = Proportion of traités dans AC

- Ces proportions sont appelées **take-up** du programme



TREATMENT ON THE TREATED (TOT)

- L'estimateur ITT est simplement la différence entre la moyenned es assignés

$$ITT = [Y(AT) - Y(AC)]$$

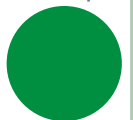
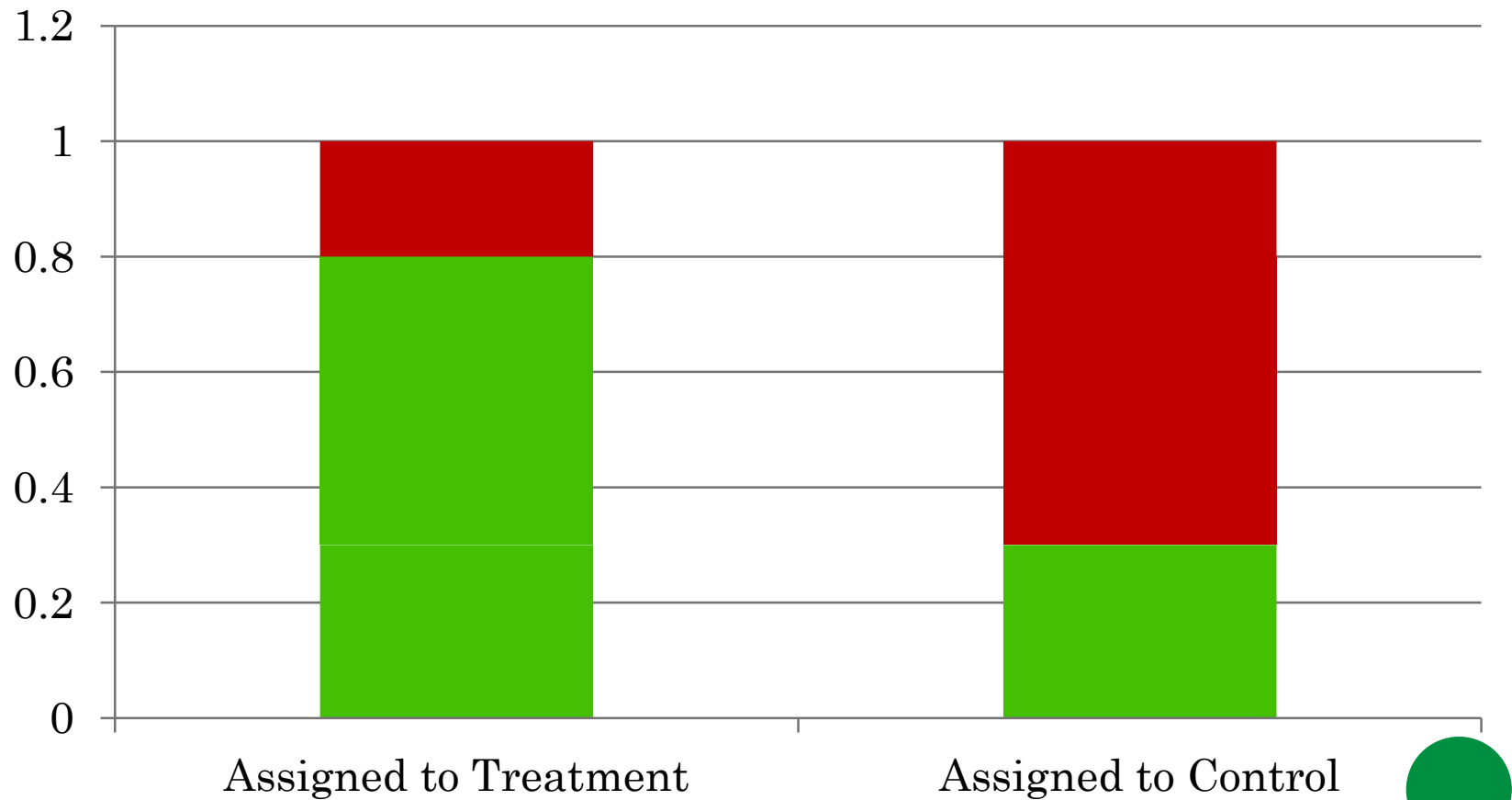
- L'estimateur TOT est simplement

$$TOT = [Y(AT) - Y(AC)] / [\text{Prob}[T | AT] - \text{Prob}[T | AC]]$$

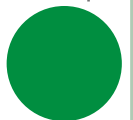
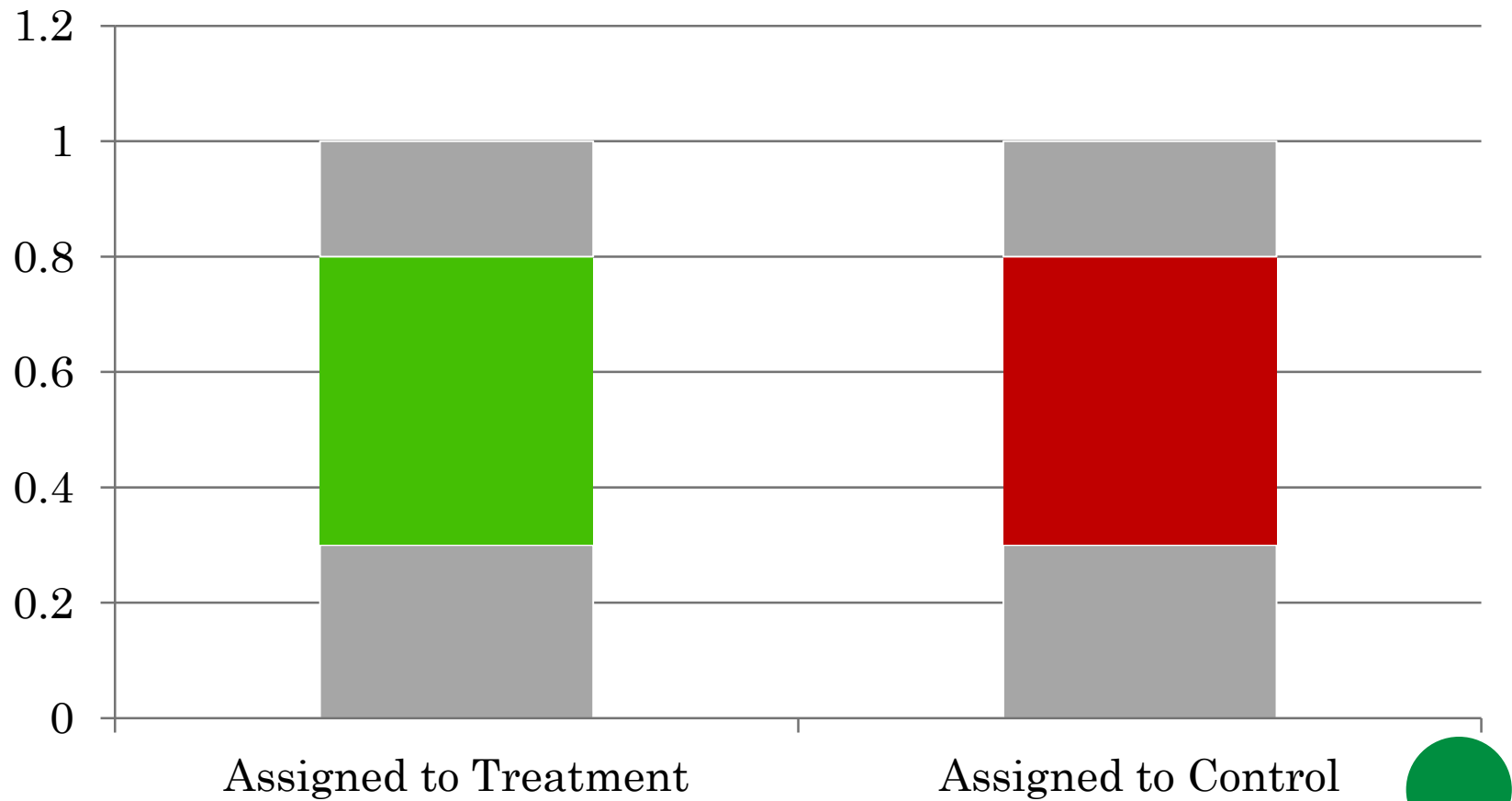
- Le **ratio** de l'**ITT** par la **différence des take-up**



ESTIMER LE TOT A APRTIR DE L'ITT



ESTIMER LE TOT A APRTIR DE L'ITT



EXEMPLE DES BILLETS DE 10.000 CFA

- Prenons un exemple très simple: imaginons que mon programme consiste à distribuer des billets de 10.000 CFA
- J'ai un groupe test de 400 personnes et un groupe contrôle de 400 personnes
- Je veux mesurer l'impact sur la richesse de mes bénéficiaires



CAS DE CONFORMITÉ PARFAITE

	Groupe test	Groupe témoin
Nombre de personnes recevant les 10.000 CFA	400	0
Nombre de personnes ne recevant pas les 10.000 CFA	0	400

Dans ce cas-ci, l'impact moyen du programme sur les bénéficiaires est identique qu'on l'exprime en ITT ou en ToT: chaque bénéficiaire a vu sa richesse augmenter de 10.000 CFA par rapport aux non-bénéficiaires.



CAS DE CONFORMITÉ PARTIELLE

	Groupe test	Groupe témoin
Nombre de personnes recevant les 10.000 CFA	300	0
Nombre de personnes ne recevant pas les 10.000 CFA	100	400

Ici, les résultats peuvent être analysés de deux manières:

- En ITT:

L'accroissement moyen de la richesse est le même dans les deux groupes: 5.000 CFA en moyenne. Mon programme a un impact sur les assignés de $300/400 * 10000 = 7500$

- En ToT:

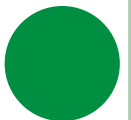
Mais si je m'intéresse à ceux qui ont reçu les 10.000 CFA, grace à l'assignation l'impact sur eux reste de 10.000 CFA



QUAND L'ITT EST-ELLE UTILE ?

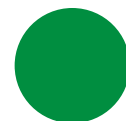
- Elle est utile pour estimer concrètement l'impact des programmes, dans leurs conditions réelles
- Par exemple, on peut ne pas être intéressé par l'effet de recevoir 10.000 CFA, mais par ce qui arriverait concrètement en cas de mise en œuvre d'un tel programme

=> L'ITT peut être plus utile en termes de politiques publiques



RÉSUMÉ SUR LA NON-CONFORMITÉ

- Effet de l'intention de traiter (ITT)
 - Souvent approprié pour des évaluations de programme
 - Simple à calculer et à expliquer
- Effet du traitement sur les traités (ToT)
 - Mesure l'effet du programme sur ceux qui le reçoivent parce qu'ils y ont été assignés
 - Peut être plus intéressant si l'on s'intéresse au concept-même du programme



PLAN DU COURS

- Menaces sur la validité interne
 - Data mining
 - Attrition
 - Externalités
 - Conformité partielle (ITT / ToT)
- **Menaces sur la validité externe**
 - **Réponses comportementales aux évaluations**
 - Généralisation des résultats



VALIDITÉ EXTERNE

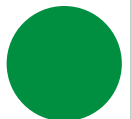
- La problématique de la validité externe est très différente de celle de la validité interne
 - Si l'on surmonte les obstacles méthodologiques vus précédemment, la validité interne d'une évaluation aléatoire est très forte: on est sûr que l'impact mesuré est conforme à la réalité
 - La validité externe, pour tout type d'évaluation, est **toujours** incertaine: on n'est jamais sûr qu'un résultat sera toujours valable à un autre endroit / moment / échelle
- => La validité externe est plus ou moins crédible, mais toujours hypothétique



RÉPONSES COMPORTEMENTALES AUX ÉVALUATIONS

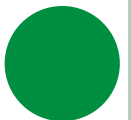
L'évaluation elle-même peut changer les comportements:

- Pour le groupe test:
 - Effet Hawthorne
- Pour le groupe témoin:
 - Effet John Henry
 - Démoralisation
 - Anticipation (en cas de passe-in)
- Pour les deux: effets d'enquête
 - Les traités peuvent répondre aux enquête non pas ce qu'ils pensent vraiment mais ce queils pensent que leur statut de traité devrait les conduire à répondre



COMMENT LES LIMITER?

- Jouer sur le niveau de randomisation
 - Si la randomisation est proche de l'entrée dans le programme les réponses comportementales sont davantage possible
- Assurer l'impartialité du staff
- Essayer de les mesurer sur un sous-échantillon



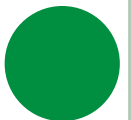
PLAN DU COURS

- Menaces sur la validité interne
 - Data mining
 - Attrition
 - Externalités
 - Conformité partielle (ITT / ToT)
- **Menaces sur la validité externe**
 - Réponses comportementales aux évaluations
 - **Généralisation des résultats**

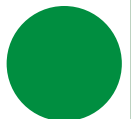
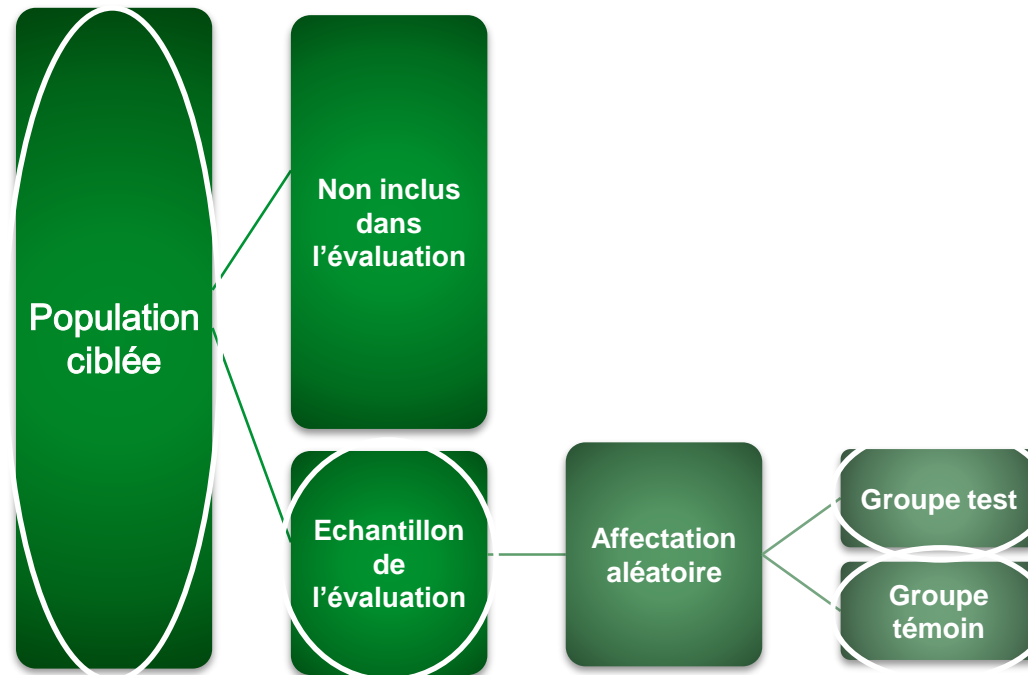


PENSER À LA GÉNÉRALISATION DÈS LE DÉBUT

- Choix du lieu, de la population, du partenaire
=> Éviter les contextes trop particuliers
- Tester des programmes qui ne reposent pas sur des inputs inhabituels: la possibilité de généraliser dépend essentiellement du programme mis en oeuvre!
 - Par exemple la présence d'un concepteur impliqué
- => Indispensable de très bien connaître le programme et son fonctionnement
- Tirer au sort l'échantillon pour qu'il soit représentatif de la population cible (cf slide suivante)



PROTOCOLE DE BASE DE L'ÉVALUATION ALÉATOIRE



REPRODUIRE OU TESTER SIMULTANÉMENT

- Dans différents lieux
- Avec différents partenaires
- A différentes échelles



COMBINER LES APPROCHES

- Qualitatif et quantitatif

=> Ouvrir la « black box »: **indicateurs intermédiaires** et **approches quali**

- Combiner les méthodes quanti entre elles: les grandes évaluations multi-pays informent sur les fortes corrélations, les RCTs sur les causalités
- Combiner test et théorie
 - Pour déterminer si le programme est sensible au contexte
 - Pour déterminer les similarités des contextes



CONCLUSION

- Il y a beaucoup de menaces de la validité interne et externe des évaluations aléatoires...
- ... comme il y en a pour tous les autres types d'études
- Mais les évaluations aléatoires :
 - rendent l'analyse simple et transparente
 - Fournissent peu de « degrés de liberté » dans l'analyse des données (ce qui est une bonne chose!)
 - permettent des tests clairs de la validité de l'expérimentation

