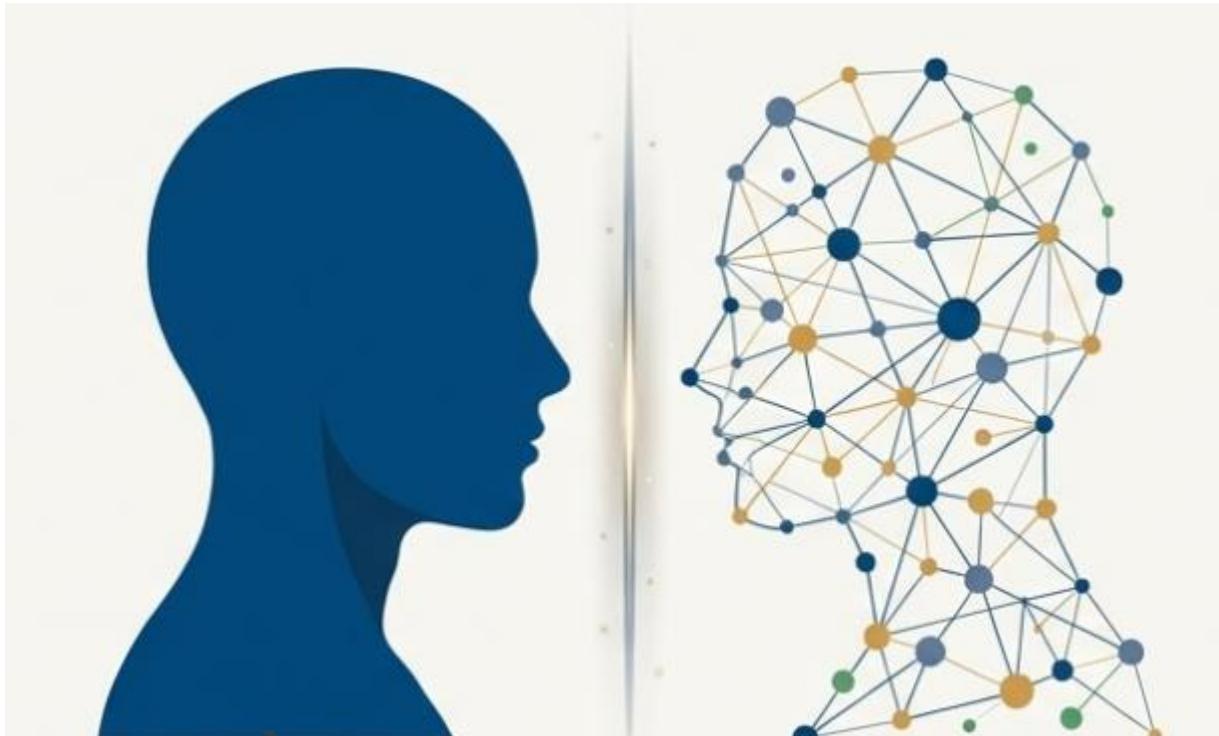




Treasury Board of Canada
Secretariat

Secrétariat du Conseil du Trésor
du Canada

Canada



INTRODUCTION À LA CONCEPTION DE RECHERCHE QUASI EXPÉIMENTALE

Mesurer l'impact lorsque la randomisation n'est pas possible

Syeda Batool : Secteur de la gestion des dépenses (SCT), Ryan Kelly- Secteur de la politique stratégique (ISDE)

AU-DELÀ DES BONNES INTENTIONS : DE « AVONS-NOUS DÉPENSÉ ? » À « A-T-IL FONCTIONNÉ ? »

Une évaluation efficace de l'impact est essentielle. Elle aide les décideurs stratégiques à déterminer si les programmes produisent les effets escomptés, favorise la reddition de comptes et comble les lacunes dans notre compréhension de ce qui fonctionne. Les principaux objectifs en question sont ci-dessous.

Responsabilité

Prouver aux bailleurs de fonds et au public que les ressources génèrent un véritable changement, et pas seulement de l'activité.



Efficacité

Comprendre ce qui fonctionne et ce qui ne fonctionne pas afin de concevoir de meilleures politiques et de meilleurs programmes à l'avenir.



Affectation

Prendre des décisions fondées sur des éléments probants concernant l'élargissement, la modification ou l'abandon de programmes.



ÉVALUATION D'IMPACT

- Mesure des *effets causaux*
- Accent mis sur **l'avant et après et sur l'avec et le sans**
- Application d'une gamme de données et de méthodes statistiques

Question clé

👉 Qu'est-ce qui a changé suite au programme?

EXEMPLE DE L'ÉVALUATION D'IMPACT

Programme : Formation professionnelle au gouvernement

Résultat : Gains après un an

Groupe

Participants

Non-participants

Gains moyens

Supérieurs

Inférieurs

Impact = Différence dans les gains

Principaux points à retenir

Impact \neq corrélation

Impact = différence en raison du programme

EXEMPLE DE L'ÉVALUATION D'IMPACT

Programme : Programme de distribution de manuels scolaires

Objectif : Améliorer les résultats d'apprentissage des élèves

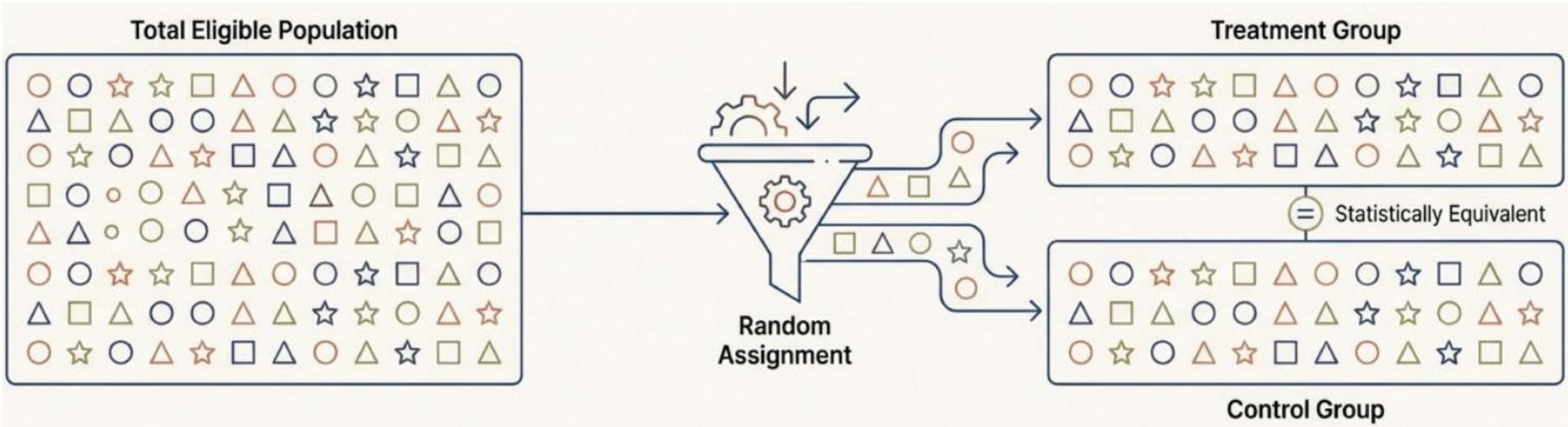
Résultat d'intérêt : Résultats aux examens à la fin de l'année scolaire

Groupe	Ayant reçu des manuels?	Note moyenne à l'examen
Groupe de traitement	Oui	Supérieur
Groupe de contrôle	Non	Inférieur

Impact = Différence dans les résultats moyens aux examens

Question clé de l'évaluation  Les élèves ont-ils appris davantage grâce aux manuels gratuits, ou les résultats auraient-ils été améliorés de toute façon?

POINT DE RÉFÉRENCE : CRÉER UN JUMEAU PARFAIT GRÂCE À LA RANDOMISATION



La randomisation permet de veiller à ce que l'attribution au traitement ne soit pas liée aux caractéristiques observées et non observées; en termes économétriques, l'attribution du traitement est **exogène**.

TOUT CE QUI BRILLE... LIMITES PRATIQUES ET ÉTHIQUES DE LA RÉPARTITION ALÉATOIRE

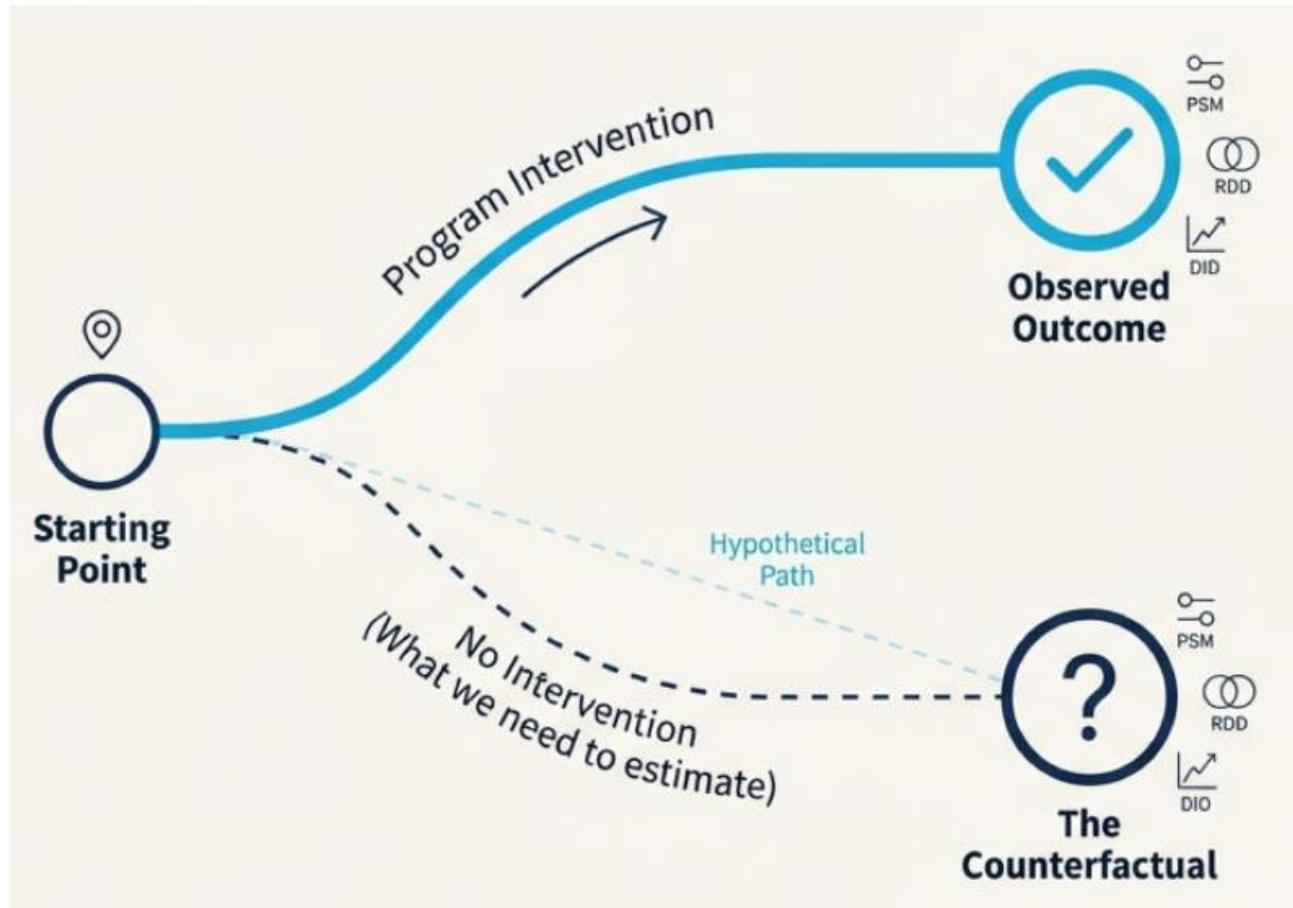
- Pas toujours réalisable ou abordable
- Contraintes éthiques et politiques
- Échelle ou durée limitée
- Problèmes de mise en œuvre et de conformité

VALIDITÉ ANALYTIQUE ET EXTERNE : LIMITES

- Validité externe limitée
- Les effets moyens peuvent masquer l'hétérogénéité
- Attrition et données manquantes
- Pas toujours informatif en ce qui concerne les mécanismes

LE DÉFI PRINCIPAL : LE MONDE INVISIBLE DE L'HYPOTHÉTIQUE

Alors, comment mesurer en toute confiance l'impact lorsque nous ne pouvons pas randomiser?

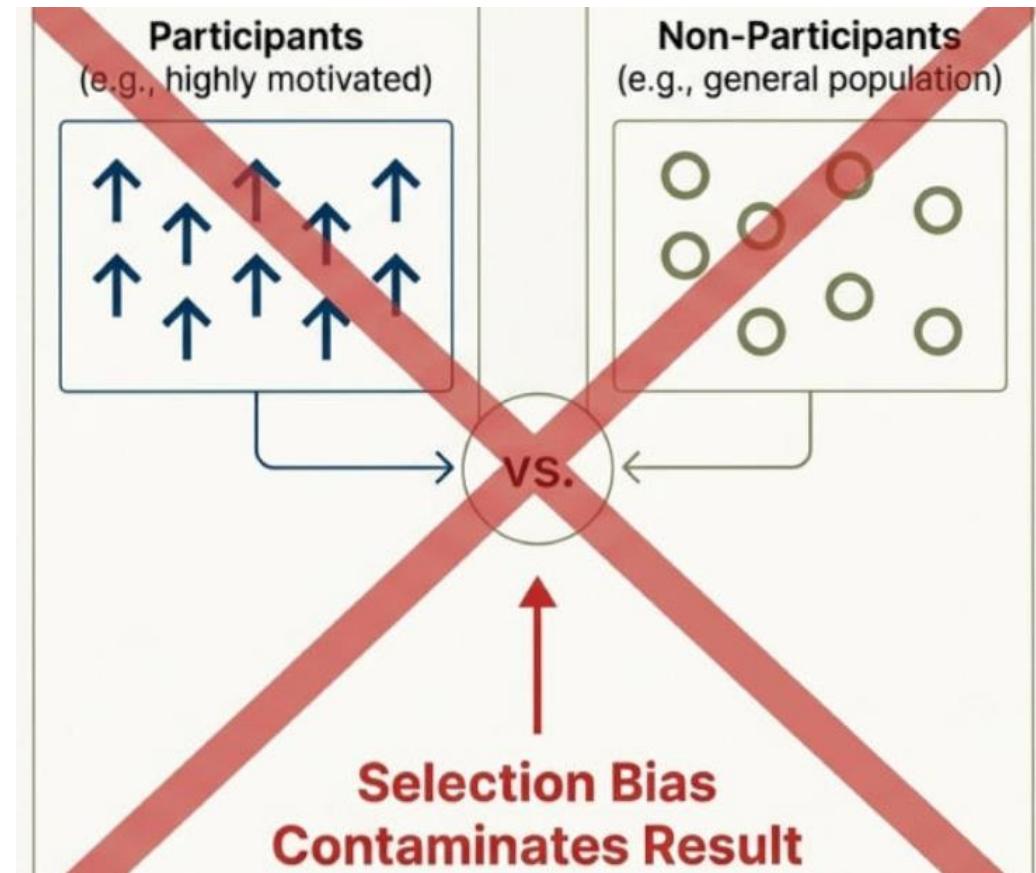


- Notre objectif en matière d'évaluation de l'impact est de répondre à une question : *Notre programme a-t-il causé le changement observé?*
- Pour ce faire, nous devons estimer la situation **hypothétique** – ce qui se serait passé sans notre intervention.

POURQUOI LES COMPARAISONS SIMPLES ÉCHOUENT-ELLES : LE PROBLÈME DU BIAIS DE SÉLECTION

Les participants au programme sont souvent fondamentalement différents des non-participants *avant même le début du programme*. Cette différence systématique est un **biais de sélection**.

Idée clé : Les différences dans les résultats peuvent être attribuées aux personnes qui se sont jointes au programme, et non à ce que le programme a fait.



POURQUOI LES COMPARAISONS SIMPLES ÉCHOUENT-ELLES : LE PROBLÈME DU BIAIS DE SÉLECTION

Autosélection : Les personnes les plus motivées pourraient s'inscrire à un programme de formation.

Placement du programme : Un programme de microcrédit peut être délibérément mis en place dans les villages les plus pauvres.

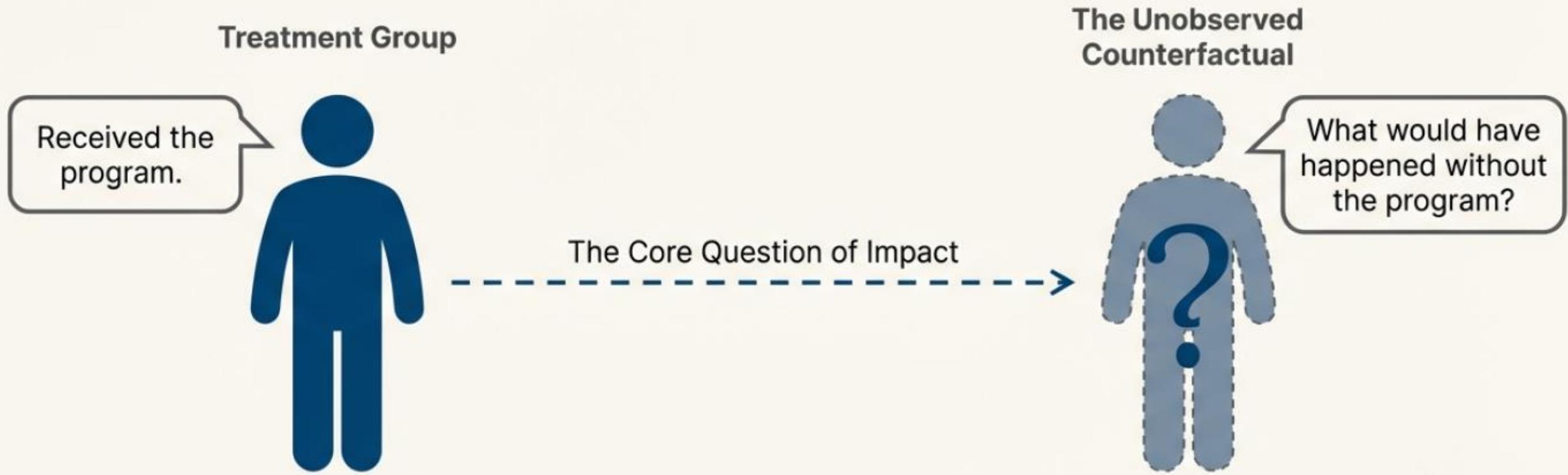
POURQUOI LES COMPARAISONS SIMPLES ÉCHOUENT-ELLES : LE PROBLÈME DU BIAIS DE SÉLECTION

Point clé : Le biais contamine les comparaisons simples, ce qui conduit à des conclusions erronées.

Avec et sans : Compare les participants à un groupe non équivalent, ce qui peut sous-estimer ou surestimer l'impact.

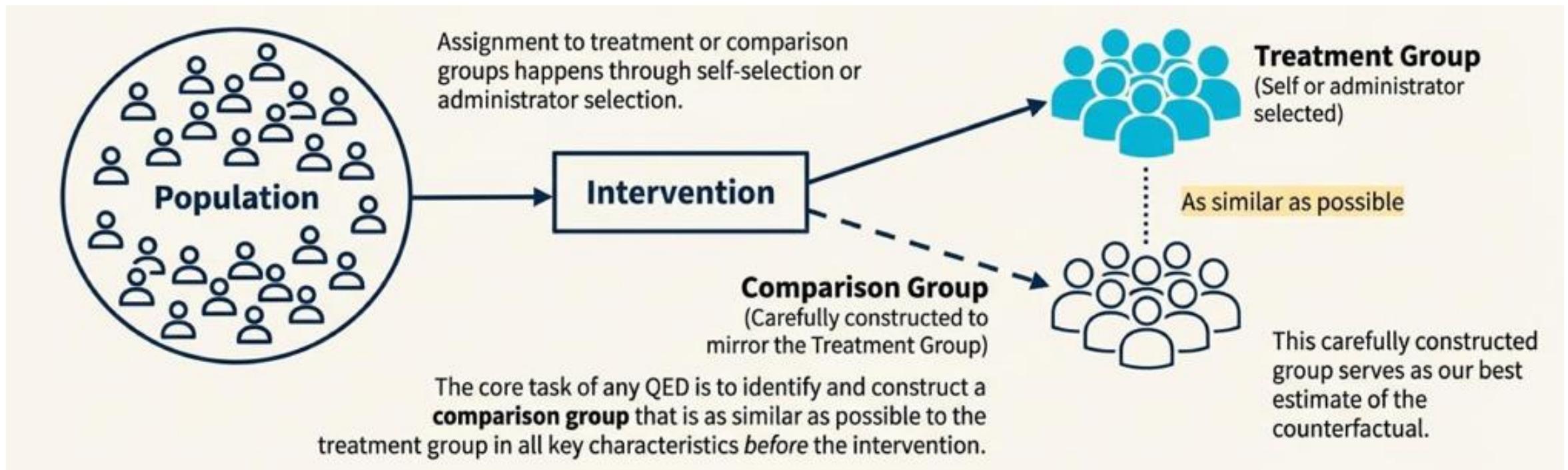
Avant et après : Ne distingue pas les effets du programme des autres facteurs externes qui changent au fil du temps (p. ex. reprise économique, autres tendances macroéconomiques).

LE DÉFI DE L'HYPOTHÉTIQUE



L'évaluation d'impact est la science de la création d'un groupe de comparaison convaincant pour estimer la situation hypothétique!

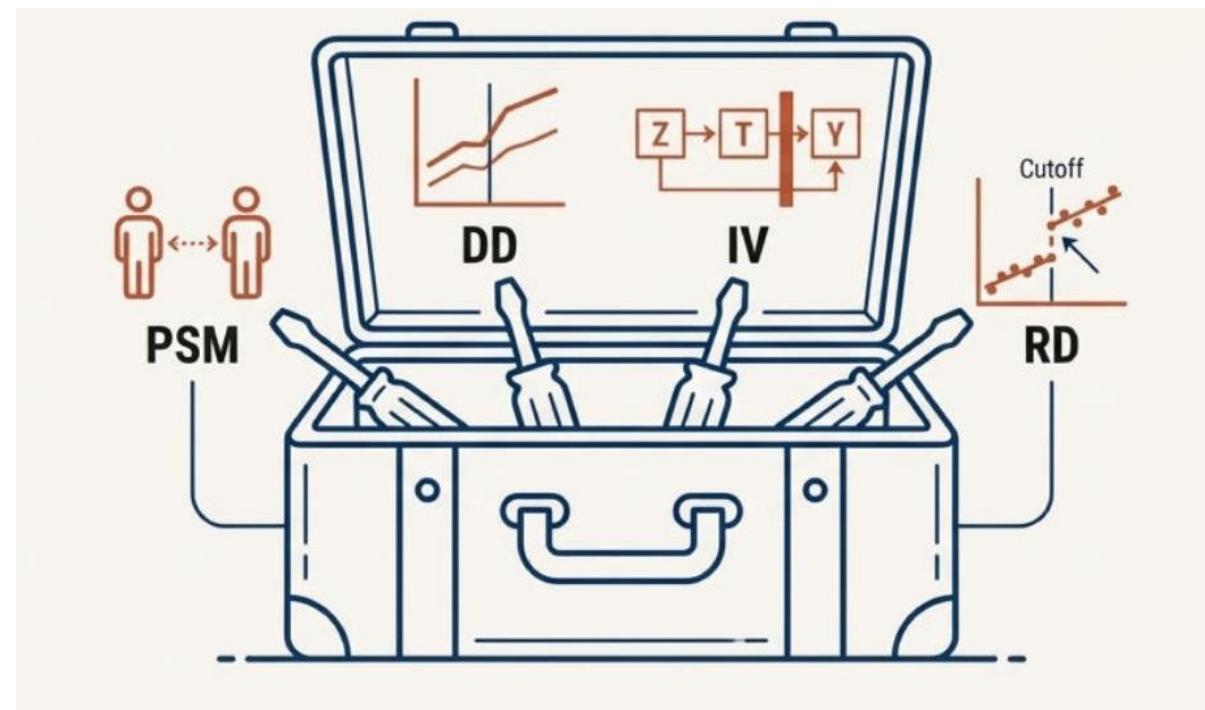
LA SOLUTION : CONCEPTION QUASI EXPÉRIMENTALE (CQE)



LORSQUE LA RANDOMISATION N'EST PAS POSSIBLE : LA TROUSSE D'OUTILS DE L'ÉVALUATEUR

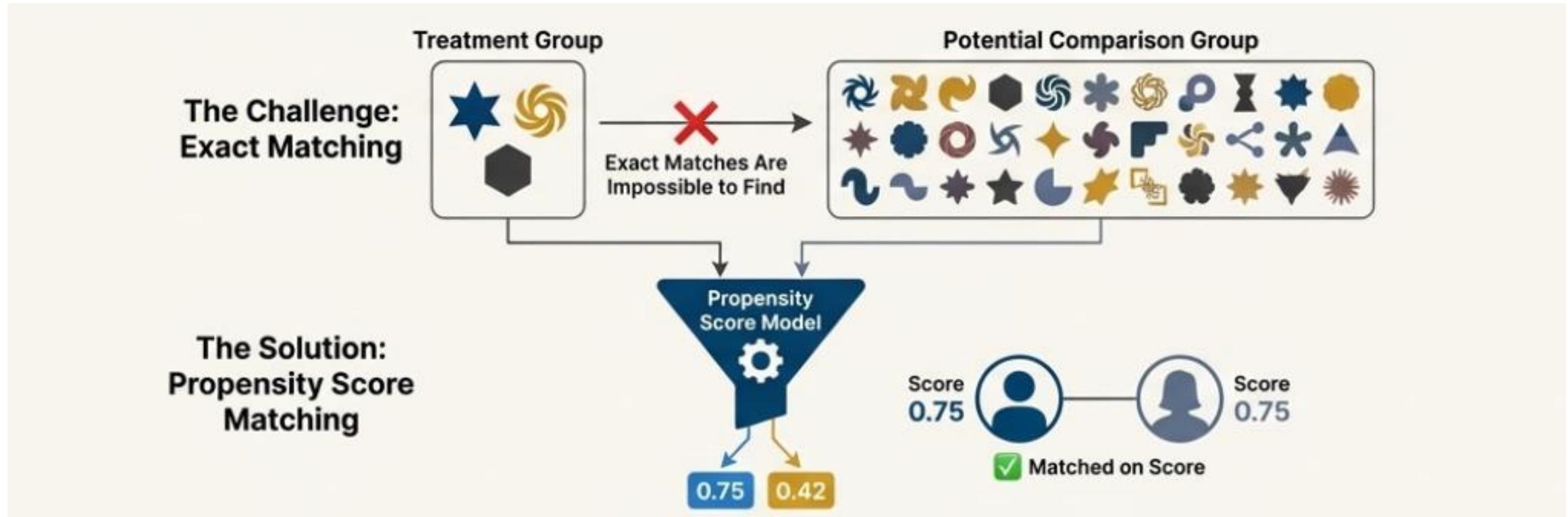
Les méthodes suivantes sont des stratégies puissantes pour « imiter » statistiquement un essai randomisé en contrôlant le biais de sélection.

Point crucial : Chaque méthode repose sur des données différentes et, surtout, **sur des hypothèses différentes** quant à la nature du biais de sélection. Pour choisir le bon outil, il faut comprendre ces hypothèses.



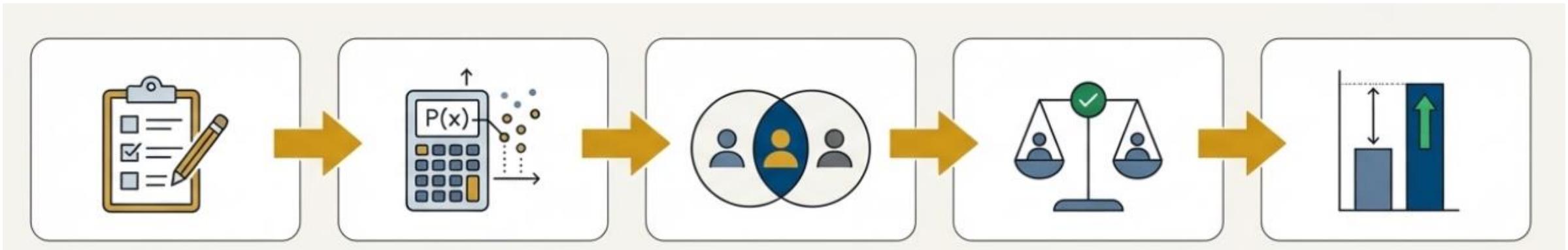
MÉTHODE 1 : APPARIEMENT DES SCORES DE PROPENSION

MÉTHODE 1 : APPARIEMENT DES SCORES DE PROPENSION (ASP)



C'est comme créer un « jumeau statistique » non pas basé sur une correspondance exacte de l'ADN, mais sur un vaste profil de traits observables.

COMMENT FONCTIONNE L'ASP : LE PROCESSUS EN 5 ÉTAPES



Faire l'enquête et l'échantillon

Recueillir des données sur un grand groupe de participants et de non-participants.

Estimer des scores

Utiliser un modèle statistique (logit/probit) pour calculer un seul « score de propension ». Dans le cas de personne, sa probabilité de participation compte tenu de leurs traits observables.

Apparier et affiner

Associez chaque personne traitée à une ou à plusieurs personnes non traitées qui ont un score très similaire. Écartez toute personne qui n'a pas de bonne correspondance (c'est ce qu'on appelle trouver la « région de soutien commun »).

Vérifier l'équilibre

Après l'appariement, vérifier l'équilibre du nouveau groupe de traitement et du groupe de comparaison (statistiquement similaires) sur le plan de leurs caractéristiques observables.

Estimer l'impact

Maintenant que les groupes sont équilibrés, la différence restante dans les résultats peut être attribuée avec plus de confiance au programme.

ÉTUDE DE CAS : LE CASSE-TÊTE DU STAGE À L'ÉTRANGER

Programme : Une intervention propose aux personnes sans emploi un stage volontaire ou un contrat de travail de courte durée à l'étranger.

Objectif : Déterminer si le programme a eu un effet réel sur l'employabilité des participants après leur retour.

Défi : La participation est volontaire. Ainsi, les participants peuvent s'inscrire eux-mêmes au programme. Nous avons des données sur 128 participants (le groupe « traité ») et 272 non-participants (le groupe « témoin »).

- Le résultat qui nous intéresse est de savoir si une personne est embauchée plus tard.

UN PREMIER COUP D'ŒIL IMPARFAIT RÉVÈLE UN PROBLÈME MAJEUR

Tableau : Âge moyen des participants par rapport aux non-participants

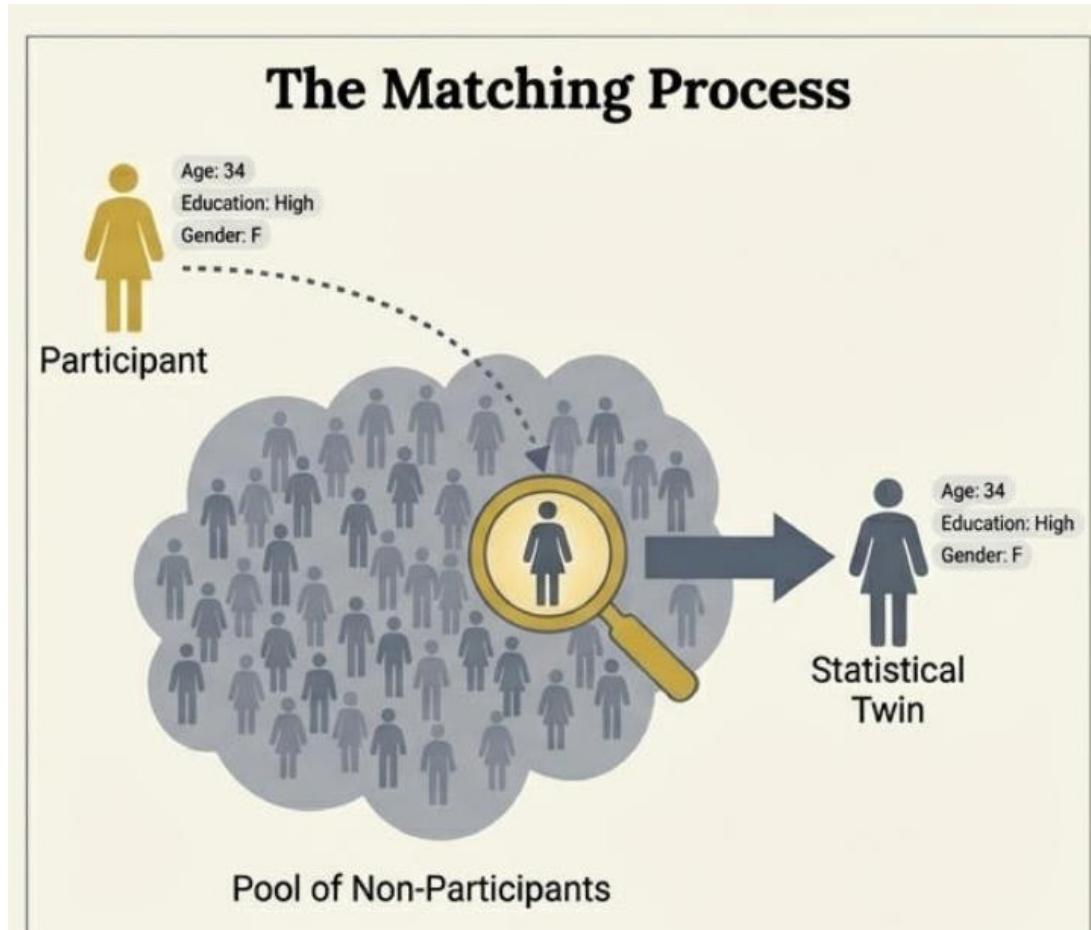
Groupe	Âge moyen	N
Nombre de participants (D=1)	34,4 ans	128
Non-Participants (D=0)	43,4 ans	272
Différence	-9,04*	

(** indique une signification statistique)*



Indication : Nous ne pouvons pas nous fier à une simple comparaison des résultats en matière d'emploi. Les participants sont-ils plus employables grâce au programme, ou simplement parce qu'ils sont plus jeunes? C'est le biais de sélection en action.

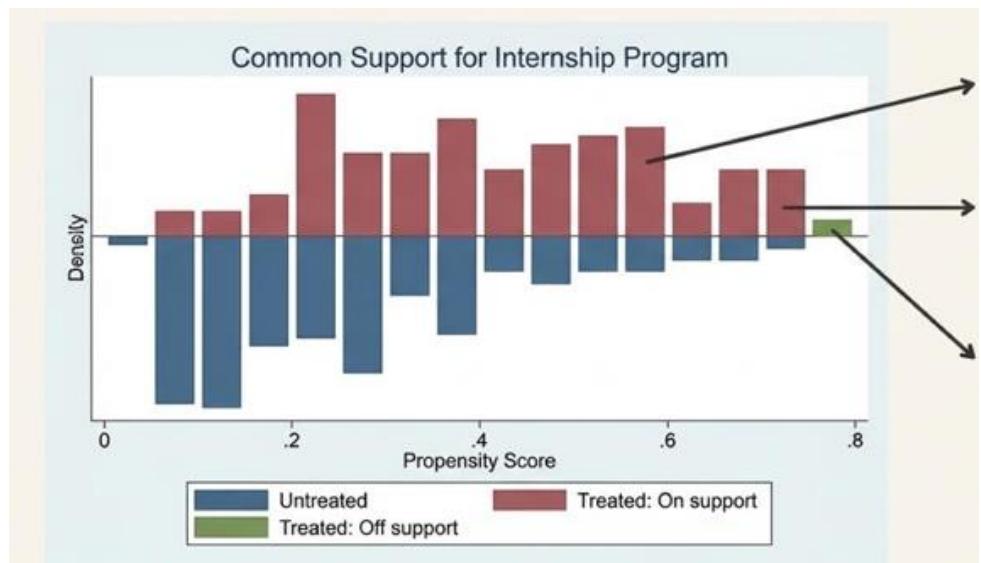
SOLUTION : TROUVER UN « JUMEAU STATISTIQUE » AVEC APPARIEMENT DES SCORES DE PROPELSION



ASP EN ACTION : LES ÉTAPES D'UNE COMPARAISON ÉQUITABLE

Étape 1 : Estimer le score de propension de toutes les personnes de l'échantillon.

Étape 2 : Vérifier la condition du « soutien commun »



Barres bleues : Distribution des scores de propension dans le cas du groupe non traité.

Barres rouges : Personnes traitées « soutien » (bonnes correspondances).

Barre verte : Personnes traitées « pas de soutien » (trop différentes, écartées de l'analyse).

Étape 3 : Choisissez un algorithme d'appariement (p. ex., les voisins les plus proches, l'appariement avec étrier) pour appairer les participants avec leurs « jumeaux ».

LE MOMENT DE VÉRITÉ : LE JUMELAGE A-T-IL CORRIGÉ LE BIAIS?

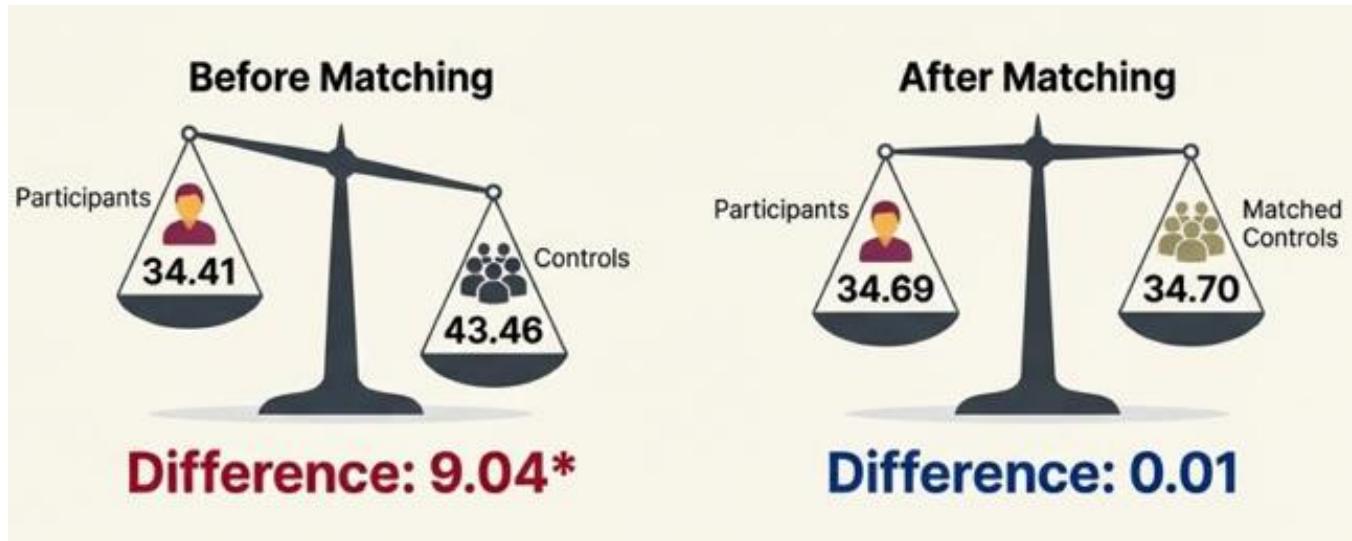


Tableau : Équilibre des covariables d'âge

	Nombre de participants (D=1)	Contrôles (D=0)	Différence
Avant l'appariement	34,41	43,46	9,04*
Après l'appariement	34,69	34,70	0,01

Résultat : Succès. L'écart d'âge de 9 ans a été éliminé. Le score de propension a réussi à équilibrer les deux groupes sur cette caractéristique clé. Nous avons maintenant un groupe de comparaison crédible.

CONCLUSION : LE PROGRAMME DE STAGES A EU UN IMPACT POSITIF SIGNIFICATIF

Tableau : Effet moyen estimé du traitement sur les traités (ETT)

Algorithme d'appariement	OPTIONS	Incidence (ETT) sur l'employabilité
Voisins les plus proches	2 Contrôles	+16,7 pp **
Voisins les plus proches	4 Contrôles	+21,8 pp ***
Appariement avec étrier	Rayon de 0,02	+24,3 pp ***
Correspondance du noyau	Fonction 2	+25,2 pp ***



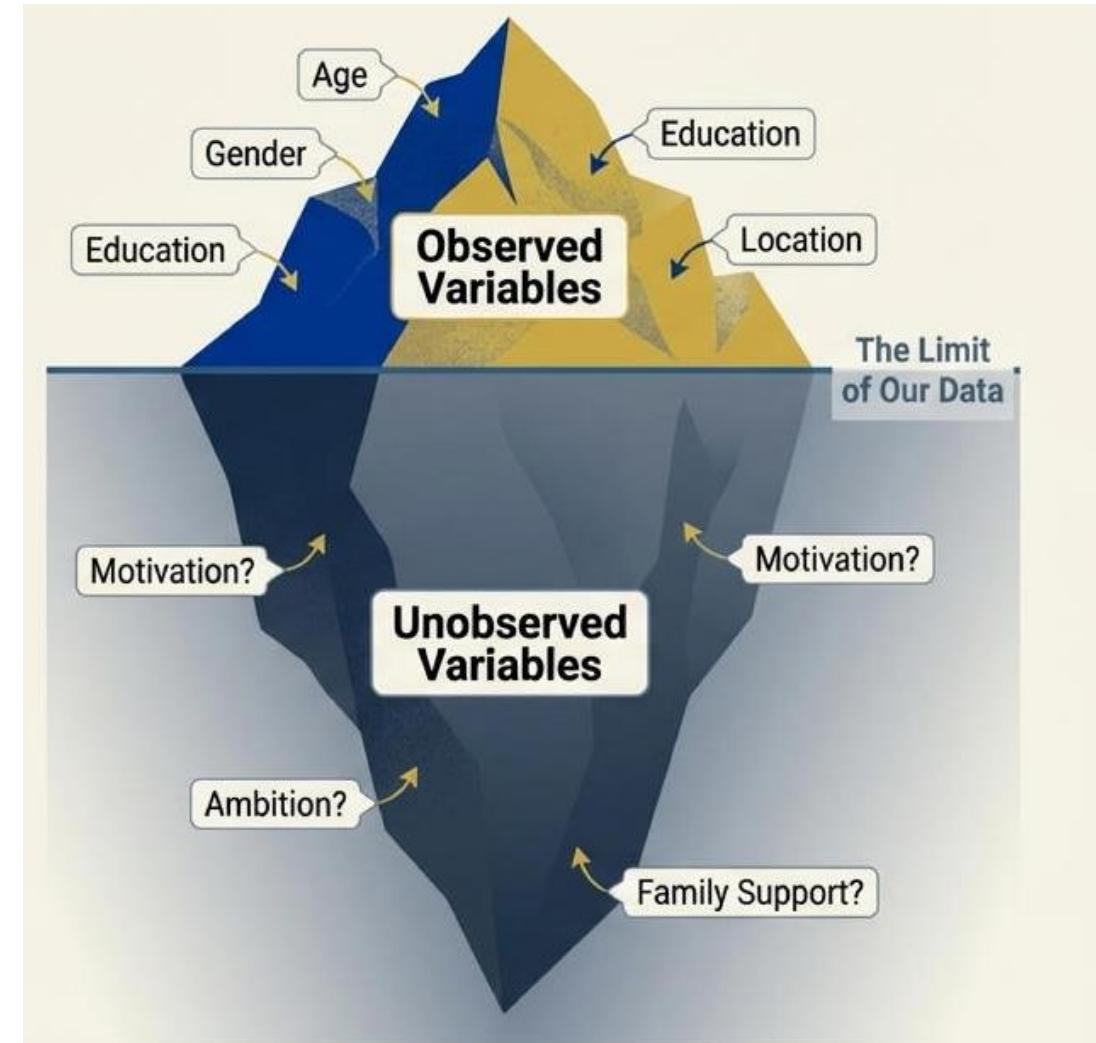
Remarque : pp = points de pourcentage. ** et *** indiquent l'importance.

Conclusion : Le programme de stage et d'emploi à l'étranger a augmenté la probabilité d'être employé de 17 à 25 points de pourcentage au profit des participants

LE GRAND « SI » : L'HYPOTHÈSE LA PLUS IMPORTANTE DE L'ASP

L'ASP est une méthode puissante, mais sa validité repose sur une hypothèse forte et invérifiable :

L'hypothèse d'indépendance conditionnelle ou « sélection à partir des variables observables »



MÉTHODE 2 : DIFFÉRENCE DANS LA DIFFÉRENCE (DD)

LOGIQUE DE LA MÉTHODE DE DD : UTILISATION D'UN GROUPE TÉMOIN POUR DÉGAGER LES TENDANCES TEMPORELLES

La méthode de DD élimine les biais des comparaisons entre un groupe de traitement et un groupe témoin en tenant compte des tendances qui ont un effet sur les deux groupes au fil du temps.

Comment ça fonctionne

1. Calculez l'évolution des résultats du groupe de traitement avant et après l'intervention.
2. Calculez la variation du résultat du groupe témoin au cours de la même période.
3. L'effet du traitement est la différence entre ces deux différences.

Le calcul de la « double différence » :

Effet = (traitement après – traitement avant) – (contrôle après – contrôle avant)

LOGIQUE DE LA MÉTHODE DE DD : UTILISATION D'UN GROUPE TÉMOIN POUR DÉGAGER LES TENDANCES TEMPORELLES

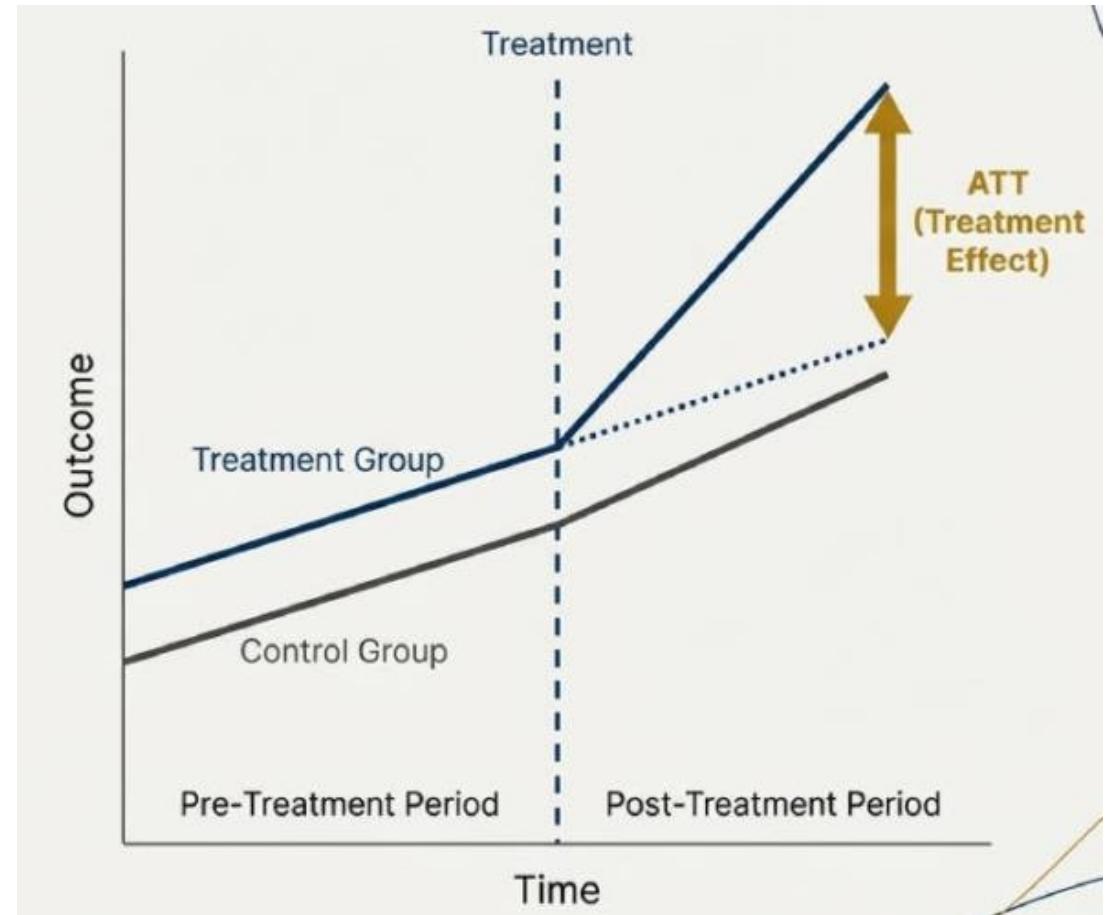
Exigences en matière de données

- Les données doivent couvrir :
 - la période avant et après l'intervention;
 - les éléments liés aux groupes de traitement et de contrôle.
- Options relatives à la structure des données
 - Données de panel : mêmes personnes observées au fil du temps
 - Coupes transversales répétées : Différentes personnes d'un même groupe au fil du temps

L'HYPOTHÈSE CRUCIALE : DES TENDANCES PARALLÈLES

Hypothèse clé

En l'absence du traitement, le résultat moyen du groupe de traitement et du groupe témoin aurait suivi la même tendance au fil du temps.



ÉTUDE DE CAS : LE PARADOXE DE L'ÉDUCATION À L'ENTREPRENEURIAT

• Programme

Un programme d'entrepreneuriat de premier plan destiné aux étudiants universitaires aux Pays-Bas, conçu pour accroître leurs compétences et leur motivation à démarrer une entreprise.

Question de recherche

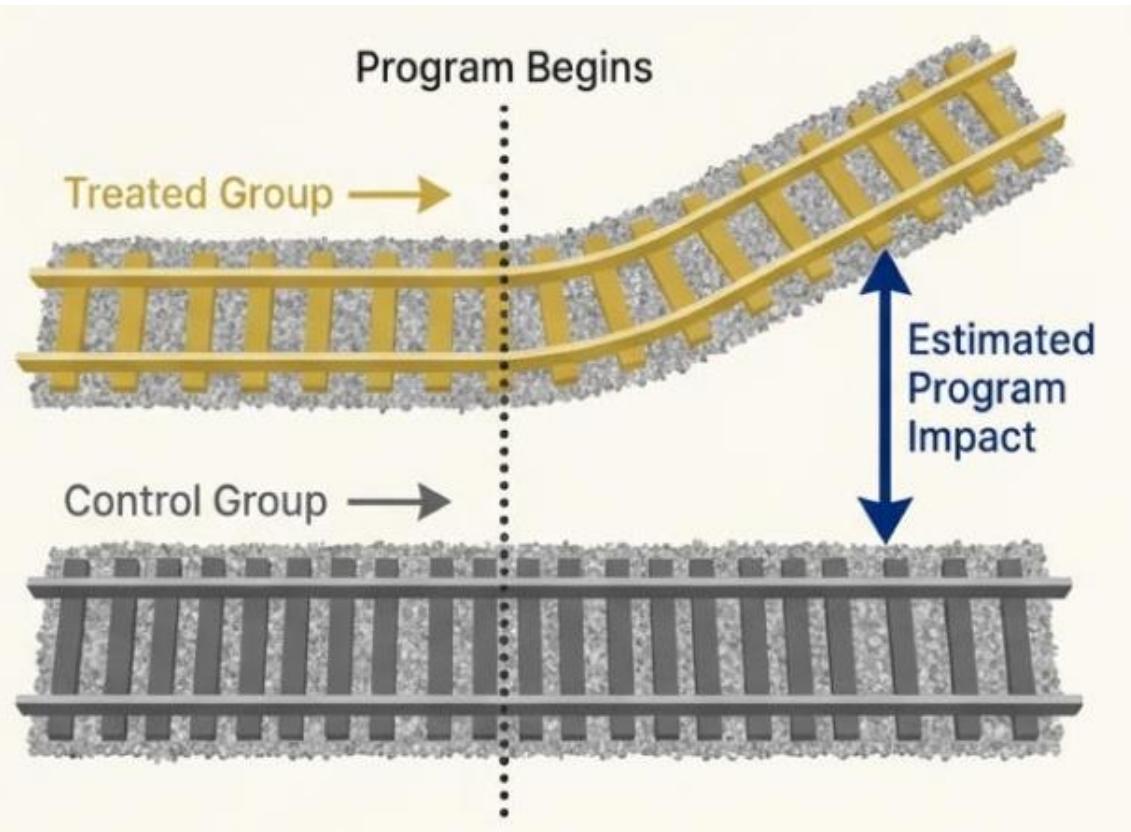
Est-ce qu'un programme de mini-entreprises destiné aux étudiants augmente les compétences entrepreneuriales et la motivation des étudiants?

Mise en place

- Groupe de traitement : Étudiants d'un campus d'un collège professionnel néerlandais à Breda, où le programme en question était obligatoire.
- Groupe témoin : Étudiants d'un campus différent (Den Bosch) du même collège, qui n'offrait pas encore le programme.
- Données : Des enquêtes auprès des étudiants des deux campus ont été menées avant et après l'année universitaire.

SOLUTION : COMPARER LES TRAJECTOIRES EN UTILISANT LA MÉTHODE DE LA DIFFÉRENCE DANS LA DIFFÉRENCE

Au lieu d'apparier les personnes, la méthode de la DD compare le changement dans le résultat au fil du temps du groupe traité au changement dans le résultat du groupe témoin.



RÉVÉLATION SURPRENANTE : UN IMPACT NÉGATIF SUR LES INTENTIONS

Les chercheurs ont comparé le changement dans les compétences et les intentions entrepreneuriales entre les étudiants de Breda (traités) et de Den Bosch (contrôle). Les résultats n'étaient pas ceux auxquels les décideurs s'attendaient.

Bien qu'il y ait eu des résultats mixtes liés à des compétences précises, la constatation la plus frappante concernait l'objectif principal.

Mesure des résultats	Estimation de l'effet de DD
Compétences de l'entrepreneur	-0.188**
Créativité	-0.360**
Intentions entrepreneuriales	**-0.553***

Constatation : Le programme a entraîné une diminution statistiquement significative des intentions déclarées des étudiants de devenir entrepreneurs.

*(Source : Oosterbeek et al., 2010. Colonne 7 du tableau 5)

LE PARADOXE EXPLIQUÉ : UNE DOSE DE RÉALITÉ



- Le programme offrait une vision réaliste de ce qu'il faut pour gérer une entreprise.
- Il a aidé certains étudiants à voir que l'entrepreneuriat n'était pas la bonne voie à prendre dans leur cas.
- Nuance importante : effet davantage prononcé dans le cas des étudiantes; la question de l'équilibre entre les affaires et d'autres priorités a entraîné une baisse des intentions entrepreneuriales.

Ce qu'il faut retenir : Une évaluation rigoureuse ne nous dit pas seulement si un programme a fonctionné, elle peut nous aider à comprendre comment et pour qui il a fonctionné.

ÉTUDE DE CAS : L'EFFET DU SALAIRE MINIMUM SUR L'EMPLOI

La question classique (Card et Krueger, 1994)

L'augmentation du salaire minimum entraîne-t-elle une baisse de l'emploi?

L'expérience naturelle

Traitemet : En avril 1992, le New Jersey (NJ) a augmenté son salaire minimum de 4,25 \$ à 5,05 \$ de l'heure.



ÉTUDE DE CAS : L'EFFET DU SALAIRE MINIMUM SUR L'EMPLOI

Mise en place : Les chercheurs ont utilisé ce changement de politique pour créer une étude de DD.

- **Groupe de traitement :** Restauration rapide dans le New Jersey.
- **Groupe de contrôle :** Restauration rapide dans l'est de la Pennsylvanie, où le salaire minimum est resté à 4,25 \$.

Données : Les chercheurs ont enquêté sur -400 établissements de restauration rapide dans les deux États avant (février 1992) et après (novembre 1992) l'augmentation du salaire minimum.

DD EN ACTION : DÉBALLAGE DE LA DOUBLE DIFFÉRENCE

Tableau : Emploi moyen d'équivalents temps plein (ETP) par magasin

	Pennsylvanie (contrôle)	New Jersey (traité)	Différence (NJ - PA)
Avant (févr. 92)	23,33	20,44	-2,89
Après (nov. 92)	21,17	21,03	-0,14
Changement (après-avant)	-2,16	+0,59	+2,76

Première différence
(contrôle) : L'emploi en PA
(contrôle) a diminué de
2,16 ETP.

La première différence (traité)
: L'emploi dans le New Jersey
(traité ★) a augmenté* de
0,59 ETP.

La deuxième différence
(l'estimateur DD) : Effet DD =
(variation de NJ) - (variation de
PA) = (+0,59) -(-2,16) = +2,76

Conclusion : Contrairement à la simple théorie économique, l'augmentation du salaire minimum dans le New Jersey n'a pas diminué l'emploi dans le secteur de la restauration rapide. Les éléments de preuve indiquent que le changement pourrait avoir légèrement l'emploi par rapport au groupe témoin.

CHOIX DU STRATÈGE : L'APPARIEMENT DU SCORE DE PROPENSION PAR RAPPORT À LA DIFFÉRENCE DANS LA DIFFÉRENCE

Appariement des scores de propension (ASP)

Logique de base : Sélectionne un groupe témoin qui a la même apparence en ce qui concerne les caractéristiques observées avant le traitement.

Hypothèse clé : indépendance conditionnelle (sélection sur les éléments observables). Suppose qu'aucune variable non observée n'affecte simultanément la participation et les résultats.

Besoins en données : Riches données transversales avec de nombreuses covariables de prétraitemet.

Poignées : Biais de sélection fondé sur des facteurs mesurables (âge, niveau d'études, entre autres).

Vulnérabilité : Caractéristiques non observées comme la motivation, le talent, le soutien familial.

Différence dans les différences (DiD)

Logique de base : Utilise un groupe témoin pour modéliser ce qui se serait passé au fil du temps sans traitement.

Hypothèse clé : Tendances parallèles. Suppose que les groupes de traitement et de contrôle auraient suivi la même tendance au fil du temps.

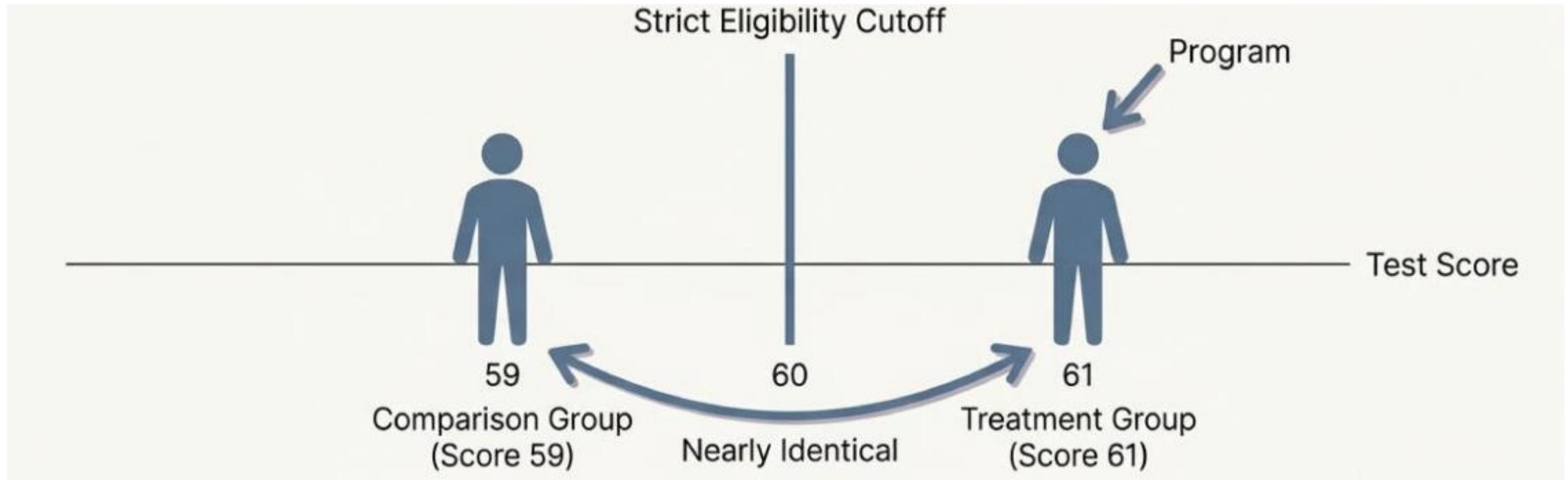
Besoins en données : Données de panel ou coupes transversales répétées (au moins deux périodes).

Poignées : biais de sélection à partir de caractéristiques non observées invariantes dans le temps (par exemple, la motivation innée).

Vulnérabilité : Événements ou chocs autres que le traitement qui ont un effet seulement sur le groupe de traitement au fil du temps.

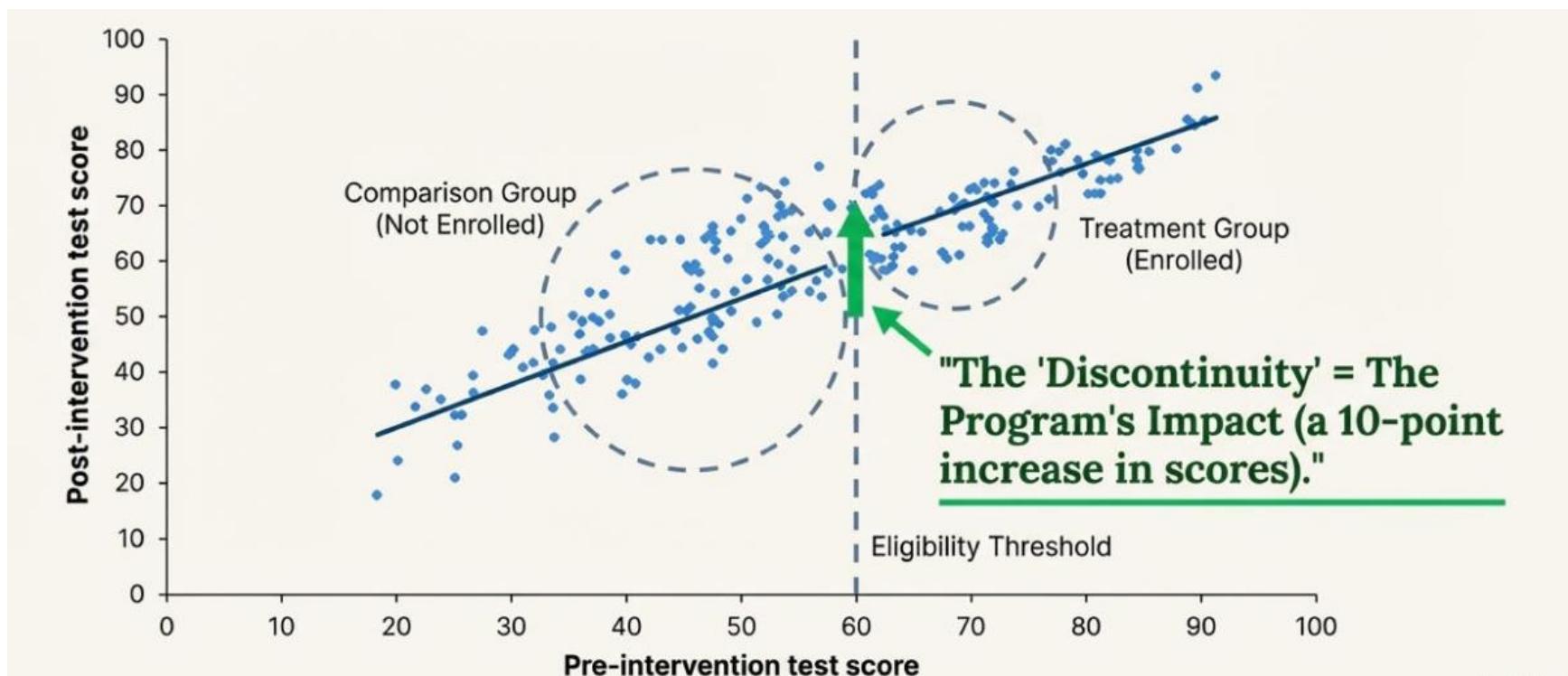
MÉTHODE 3 : MODÈLE DE DISCONTINUITÉ DE LA RÉGRESSION

MÉTHODE 3 : MODÈLE DE DISCONTINUITÉ DE LA RÉGRESSION (MDR)



MDR EN ACTION : LA PUISSANCE DE LA COUPURE

Un programme de rattrapage est offert aux élèves qui obtiennent une note inférieure à 60 à un test préalable. Nous voulons mesurer son impact sur leurs résultats aux tests effectués par la suite.



AVANTAGES ET INCONVÉNIENTS DU MDR

Avantages

- **Gère les éléments inobservables**

Parce que les personnes juste autour de la limite sont si similaires, le MDR est beaucoup plus efficace pour traiter les caractéristiques non observées que l'ASP. Ses affirmations causales sont souvent considérées comme plus crédibles que celles obtenues au moyen de l'ASP.

- **Transparence**

La règle d'admissibilité est claire et l'analyse est visuellement intuitive.

Inconvénients

- **Effet local**

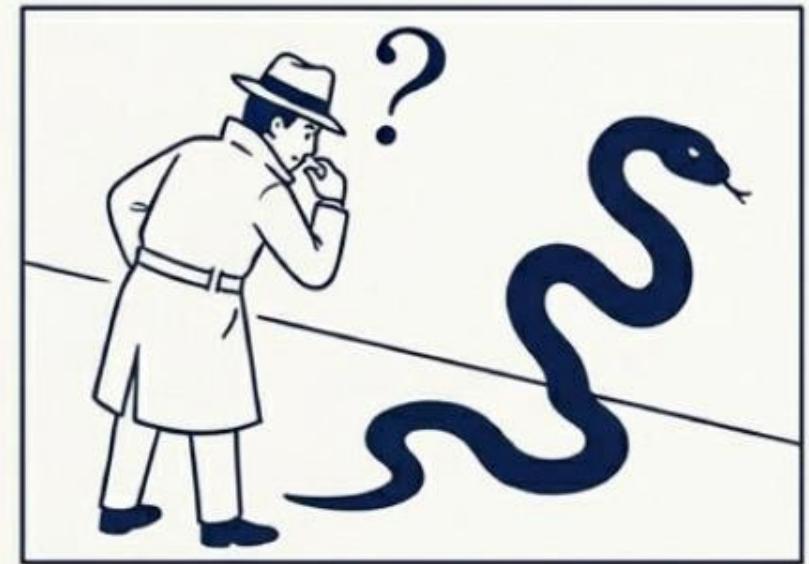
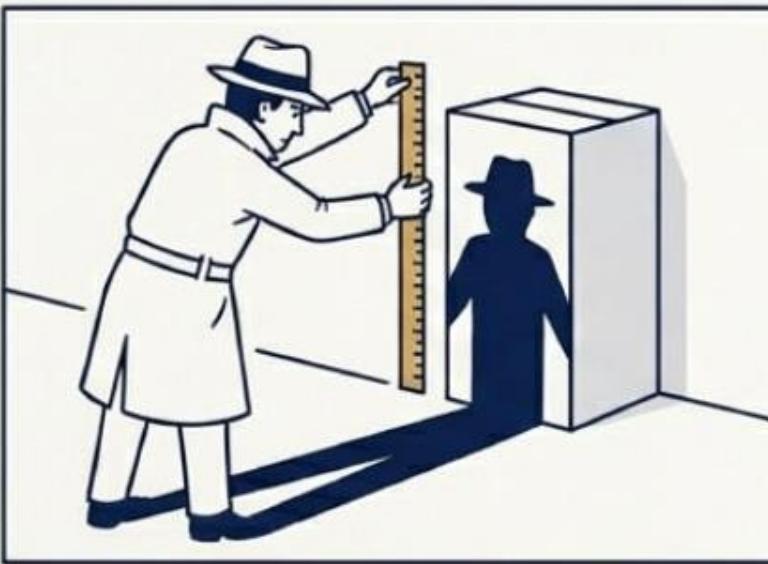
L'estimation de l'impact n'est valable que dans le cas de la population située juste autour de la limite. L'effet pourrait être différent dans le cas des personnes qui sont loin du seuil.

- **Champ d'application limitée**

Il faut un programme ayant une règle d'admissibilité précise et clairement définie ainsi que suffisamment de données regroupées autour de cette limite.

LE DÉFI LE PLUS DIFFICILE : QUE SE PASSE-T-IL SI LE BIAIS EST CACHÉ ET CHANGE AU FIL DU TEMPS?

- Le **ASP** échoue si des facteurs non observés sont importants.
- La **méthode de DD** échoue si le biais non observé change au fil du temps (p. ex., la motivation des participants augmente *en raison* du programme, ou le placement du programme cible des domaines ayant une possibilité de croissance changeante).



ASP : Trouve ce qui se trouve sous la lumière
(variables observables).

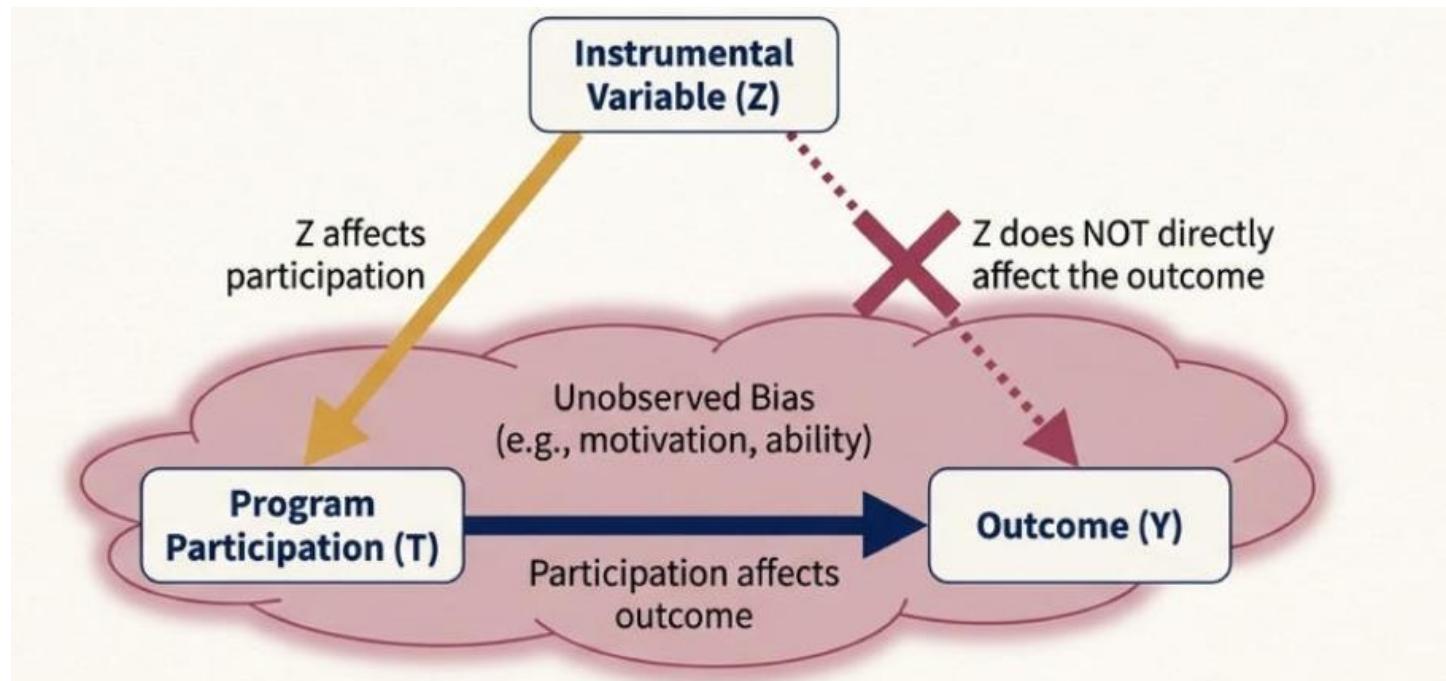
DD : Mesure une ombre constante (biais invariant dans le temps).

Le problème : une ombre changeante et cachée
(biais variant dans le temps).

MÉTHODE 4 : VARIABLE INSTRUMENTALE (VI)

SOLUTION : LES VARIABLES INSTRUMENTALES (VI)

L'idée : trouver une source de variation – un « instrument » – qui soit aussi bonne qu'aléatoire. Cet instrument agit comme un « coup de pouce aléatoire » encourageant certaines personnes à participer, mais pas d'autres, sans avoir une incidence directe sur leurs résultats.



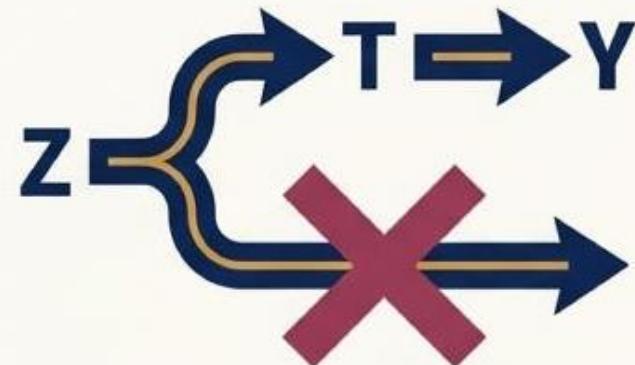
LES DEUX RÈGLES D'OR À RESPECTER POUR CHOISIR UN BON INSTRUMENT

Pour être un instrument valide, une variable valide doit satisfaire à deux conditions strictes;



1. The Relevance Condition

$$\text{cov}(Z, T) \neq 0$$



2. The Exclusion Restriction

$$\text{cov}(Z, \varepsilon) = 0$$

OÙ TROUVE-T-ON CES « INSTRUMENTS » DANS LE MONDE RÉEL?

Pour trouver un instrument valide, il faut une connaissance approfondie de la conception et du contexte du programme. Les sources courantes sont présentées ci-dessous.

Géographie du placement du programme

Parfois, les programmes sont mis en œuvre dans certaines régions, mais pas dans d'autres, et les raisons à la base de ce choix n'ont rien à voir avec les résultats (p. ex., des raisons de commodité administrative).

La distance par rapport à un centre de programmes peut être un instrument.

OÙ TROUVE-T-ON CES « INSTRUMENTS » DANS LE MONDE RÉEL?

Règles d'admissibilité

Un programme peut avoir une limite nette et arbitraire concernant l'admissibilité (p. ex., âge, revenu ou propriété foncière).

Le fait d'être juste au-dessus ou en dessous de la limite peut servir d'instrument.

OÙ TROUVE-T-ON CES « INSTRUMENTS » DANS LE MONDE RÉEL?

Encouragement aléatoire

Au lieu de randomiser le programme lui-même, nous pouvons donner au hasard à certaines personnes une incitation ou des informations supplémentaires pour les encourager à s'inscrire.

L'encouragement lui-même est l'instrument.

OÙ TROUVE-T-ON CES « INSTRUMENTS » DANS LE MONDE RÉEL?

Conception des politiques

Les caractéristiques de la mise en œuvre d'une politique, par exemple si les hommes et les femmes doivent rejoindre des groupes séparés dans un programme de microfinance, peuvent créer des variations exogènes dans la participation.

UN CAS PARTICULIER : LA DISCONTINUITÉ DE LA RÉGRESSION (DR)

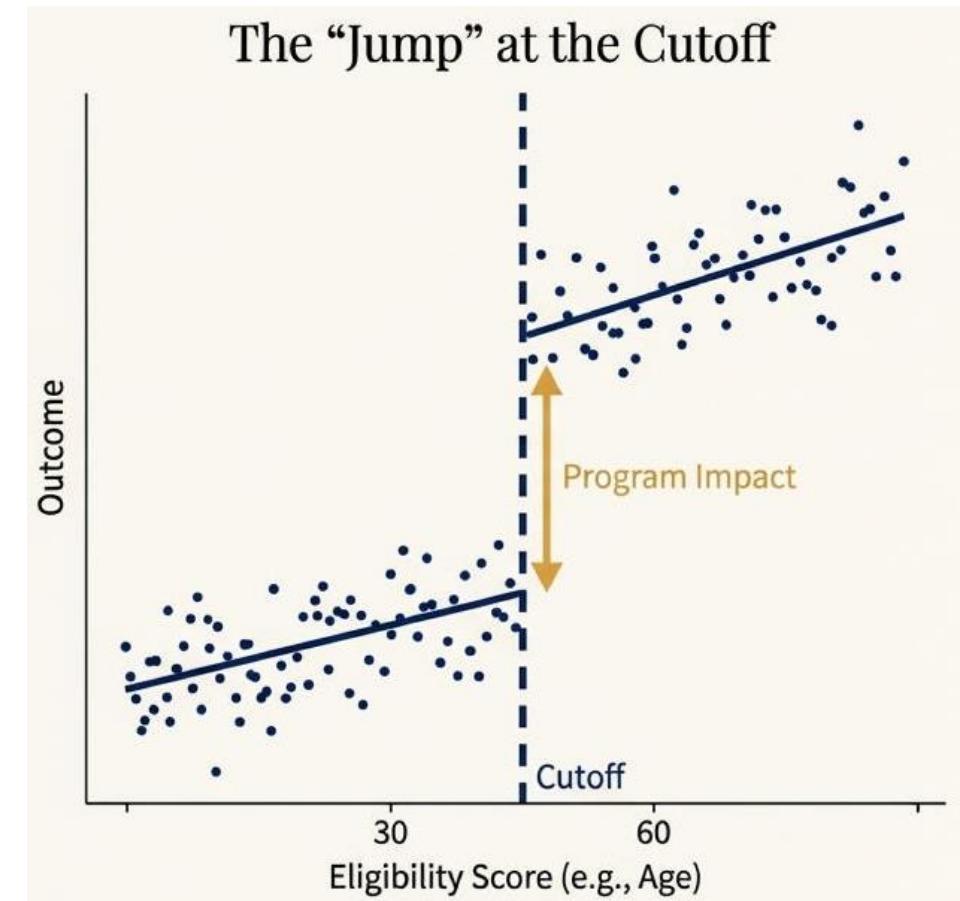
La DR est un modèle puissant qui utilise une règle d'admissibilité comme instrument.

L'idée : Nous pouvons comparer les personnes qui sont à peine admissibles à un programme avec celles qui sont à peine inadmissibles.

Exemple : Exploitation des règles d'admissibilité en Afrique du Sud

Un programme de pension sociale a une limite d'âge stricte en ce qui concerne l'admissibilité. Il est raisonnable de supposer que les personnes âgées de 60 ans (tout juste admissibles) sont très semblables aux personnes âgées de 59 ans et 11 mois (tout juste inadmissibles) à tous les autres égards. La différence dans leurs résultats peut être attribuée au programme.

Dans ce cas, l'instrument est le seuil d'admissibilité lui-même. Il prédit puissamment la participation, mais il est peu probable qu'il soit directement corrélé avec d'autres facteurs ayant une incidence sur les résultats juste autour de ce point limite.



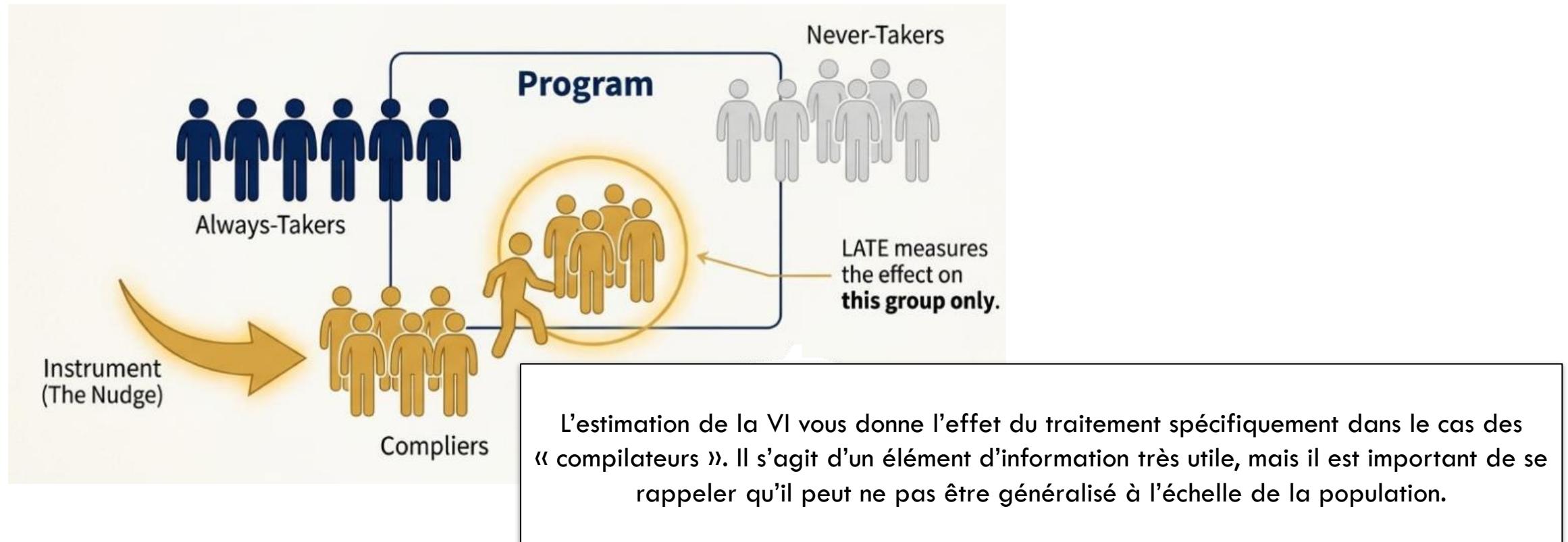
MISE EN GARDE : LE DANGER DES « INSTRUMENTS FAIBLES »



Il vaut mieux ne pas avoir d'instrument qu'un instrument faible. Les chercheurs doivent tester rigoureusement la résistance des instruments qu'ils proposent avant de tirer des conclusions.

UNE QUESTION ESSENTIELLE : QUE MESURE RÉELLEMENT LA VI?

Une estimation de la VI ne mesure pas l'effet moyen dans le cas de chaque participant. Elle mesure l'effet moyen du traitement local : l'effet moyen du programme sur le sous-groupe spécifique de personnes qui ont été incitées à participer par l'instrument.



LA BOÎTE À OUTILS DU DÉTECTIVE : UNE COMPARAISON DES MÉTHODES

Méthode	Comment elle gère le biais de sélection	Hypothèse clé	Besoins habituels en matière de données
Appariement des scores de propension (MSP)	Contrôle des différences observables par la création d'un « jumeau statistique »	Absence de confusion : Pas de sélection des caractéristiques non observées.	Riches données transversales ayant de nombreuses covariables préalables au programme.
Double différence (DD)	Contrôle des différences non observées, mais invariables dans le temps, en comparant les changements au fil du temps.	Tendances parallèles : Les groupes de traitement et de contrôle auraient suivi des tendances similaires sans le programme.	Données de panel ou coupes transversales répétées (données avant et après).
Variables instrumentales (VI)	Contrôle des différences non observées et variant dans le temps à l'aide d'une source externe de variation.	Pertinence et exclusion : L'instrument doit avoir une incidence sur la participation, mais pas directement sur le résultat.	Données transversales ou données de panel en plus d'un instrument valide.
Discontinuité de régression (DR)	Un cas particulier de VI qui contrôle les biais en comparant les unités juste au-dessus et au-dessous d'un seuil d'admissibilité.	Continuité : Les unités situées de part et d'autre de la limite sont comparables à tous les autres égards.	Données sur la variable d'affectation qui détermine l'admissibilité.

DE QUELLES DONNÉES AVEZ-VOUS BESOIN?

La qualité de toute évaluation d'impact dépend de la qualité et du type de données recueillies.

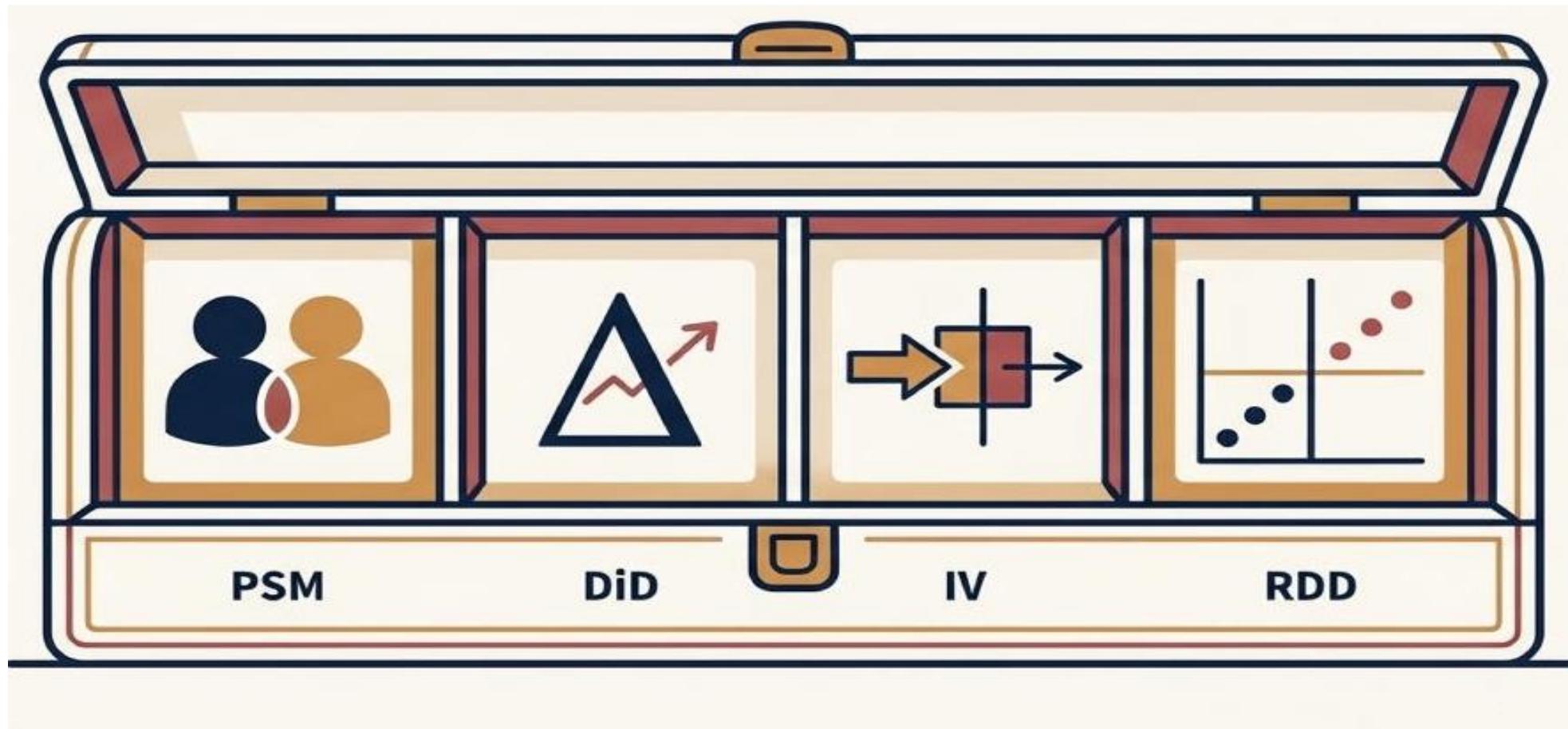
Dans le cas de toutes les méthodes non expérimentales

- **Covariables riches** : Recueillir des données détaillées sur les caractéristiques des ménages, des personnes et des communautés avant le début du programme. Cette étape est essentielle pour faire l'ASP et pour vérifier l'équilibre dans d'autres méthodes.
- **Instrument de sondage commun** : Utiliser le même questionnaire et la même méthodologie de sondage auprès des groupes de participants et de non-participants afin d'assurer la comparabilité.
- **Grand échantillon de non-participants** : Un grand échantillon représentatif de non-participants admissibles est essentiel pour trouver de bonnes correspondances (ASP) et assurer la puissance statistique.

Dans les cas de DD et de la VI de panel

- **Une enquête de référence** est essentielle : il est essentiel de recueillir des données avant l'intervention. Il vous permet de tester l'hypothèse de tendances parallèles (DD) et de contrôler les conditions initiales.
- **Données de panel** : Suivre les mêmes personnes ou ménages au fil du temps est la norme à suivre pour contrôler l'hétérogénéité non observée.

IL N'Y A PAS DE SOLUTION MIRACLE, SEULEMENT LE BON OUTIL POUR FAIRE LE TRAVAIL



MERCI!

Syeda.Batool@tbs-sct.gc.ca
ryan.kelly@ised-isde.gc.ca