

# Causalité

## Expérimentation

Visseho Adjivanou, PhD.

Département de Sociologie - UQAM

16 October 2022

# Plan de présentation

- 1 Questions causales en sciences sociales et terminologie
- 2 Effets causaux et contrefactuel
- 3 Essais contrôlés randomisés (*Randomized controlled trials*) et causalité
- 4 Applications:
  - Discrimination raciale sur le marché du travail

# Introduction

# Introduction

- Dans ce chapitre, nous considérons la causalité, l'un des concepts les plus centraux des sciences sociales quantitatives.
- Une grande partie de la recherche en sciences sociales s'intéresse aux effets causaux de diverses politiques et autres facteurs sociétaux.

# Introduction

- Dans ce chapitre, nous considérons la causalité, l'un des concepts les plus centraux des sciences sociales quantitatives.
- Une grande partie de la recherche en sciences sociales s'intéresse aux effets causaux de diverses politiques et autres facteurs sociétaux.
- Par exemple:

# Introduction

- Dans ce chapitre, nous considérons la causalité, l'un des concepts les plus centraux des sciences sociales quantitatives.
- Une grande partie de la recherche en sciences sociales s'intéresse aux effets causaux de diverses politiques et autres facteurs sociétaux.
- Par exemple:
- Est-ce que le vaccin A protège contre la maladie X?

# Introduction

- Dans ce chapitre, nous considérons la causalité, l'un des concepts les plus centraux des sciences sociales quantitatives.
- Une grande partie de la recherche en sciences sociales s'intéresse aux effets causaux de diverses politiques et autres facteurs sociétaux.
- Par exemple:
  - Est-ce que le vaccin A protège contre la maladie X?
  - Les classes de petite taille augmentent-elles les résultats des tests standardisés des élèves?

# Introduction

- Dans ce chapitre, nous considérons la causalité, l'un des concepts les plus centraux des sciences sociales quantitatives.
- Une grande partie de la recherche en sciences sociales s'intéresse aux effets causaux de diverses politiques et autres facteurs sociétaux.
- Par exemple:
  - Est-ce que le vaccin A protège contre la maladie X?
  - Les classes de petite taille augmentent-elles les résultats des tests standardisés des élèves?
  - Les soins de santé universels amélioreraient-ils la santé et les finances des pauvres?



# Introduction

- Par exemple:
  - L'éducation réduit-elle le nombre d'enfants?

# Introduction

- Par exemple:
  - L'éducation réduit-elle le nombre d'enfants?
  - La rémunération des gens sur Wikipedia augmente-t-elle leur productivité?

# Introduction

- Par exemple:
  - L'éducation réduit-elle le nombre d'enfants?
  - La rémunération des gens sur Wikipedia augmente-t-elle leur productivité?
  - Est-ce que l'augmentation du salaire minimum réduit l'activité économique?

## Questions de recherche

## Questions de recherche

- Une question de recherche est au cœur d'un projet de recherche, d'une étude ou d'une revue de littérature.
- Il concentre l'étude, détermine la méthodologie et guide toutes les étapes de la recherche, de l'analyse et de la production de rapports.
- Peut être **associatif** ou **causal**

## Exemple 1

- 1 Le salaire minimum augmente-t-il le taux de chômage?
- Le taux de chômage a augmenté après l'augmentation du salaire minimum.

## Exemple 1

- 1 Le salaire minimum augmente-t-il le taux de chômage?
  - Le taux de chômage a augmenté après l'augmentation du salaire minimum.
  - Le taux de chômage aurait-il augmenté si l'augmentation du salaire minimum n'avait pas eu lieu?

## Exemple 2

- 2 La race/l'ethnie a-t-elle une incidence sur les perspectives d'emploi?
- Mohamed a postulé pour un emploi mais ne l'a pas obtenu.
  - Mohamed aurait-il trouvé un travail s'il était blanc (avait un nom européen)?



## Exemple 3

- 3** Est-ce que fumer cause une maladie coronarienne?
- Jean, fumeur, a eu une maladie coronarienne.
  - Est-ce que Jean aurait eu la même maladie s'il n'était pas fumeur?

## Exemple 4

- 4 Quelle est l'importance des questions souverainistes dans la victoire de François Legault?
  - Au cours des élections de 2018, la question souverainiste a été laissée de côté et François Legault a gagné.
  - François Legault aurait-il gagné les élections si ces questions étaient présentes?

## Exemple 5

- 5 L'augmentation des salaires des enseignants augmente-t-elle l'apprentissage des élèves?

# Terminologie

## 1 Réponse ou variable dépendante, *outcome*

- C'est ce que nous voulons expliquer.

- *Exemples:*

- Taux de chômage
- Perspective d'emploi
- Maladie coronarienne
- Victoire de François Legault

# Terminologie

## 2 Variable indépendante, facteur de risque

- Tout facteur pouvant influencer la variable de réponse
- Peut être de différents niveaux
- Leur choix dépend de la théorie
- *Exemples:*
  - Salaire minimum
  - Ethnie / Race
  - Fumer
  - Questions souverainistes

## 3 Variables de contrôle

# Comment répondre à ces questions

- Regarder les différences dans les données
  - Par exemple: On peut constater que les élèves réussissent mieux dans les écoles qui offrent des salaires élevés aux enseignants

## Comment répondre à ces questions

- Problèmes: Les écoles où les enseignants gagnent plus peuvent être différentes à bien des égards.

## Comment répondre à ces questions

- Problèmes: Les écoles où les enseignants gagnent plus peuvent être différentes à bien des égards.
- Par exemple, les élèves des écoles où les salaires des enseignants sont élevés peuvent provenir de familles plus riches.



## Comment répondre à ces questions

- Problèmes: Les écoles où les enseignants gagnent plus peuvent être différentes à bien des égards.
- Par exemple, les élèves des écoles où les salaires des enseignants sont élevés peuvent provenir de familles plus riches.
- Ainsi, ce qui ressemble à un effet des enseignants pourrait simplement provenir de la comparaison de différents types d'élèves.

## Comment répondre à ces questions

- Problèmes: Les écoles où les enseignants gagnent plus peuvent être différentes à bien des égards.
- Par exemple, les élèves des écoles où les salaires des enseignants sont élevés peuvent provenir de familles plus riches.
- Ainsi, ce qui ressemble à un effet des enseignants pourrait simplement provenir de la comparaison de différents types d'élèves.
- Ces différences non mesurées entre les étudiants sont appelées des **facteurs de confusion** et, en général, la possibilité de facteurs de confusion nuit à la capacité des chercheurs à répondre aux questions de cause à effet en recherchant des modèles dans les données existantes.

## Comment répondre à ces questions

- Une solution au problème des facteurs de confusion consiste à essayer de faire des comparaisons équitables en ajustant les différences observables entre les groupes.
- Par exemple,
- vous pourrez peut-être télécharger des données sur la taxe foncière à partir d'un certain nombre de sites Web gouvernementaux.

## Comment répondre à ces questions

- Une solution au problème des facteurs de confusion consiste à essayer de faire des comparaisons équitables en ajustant les différences observables entre les groupes.
- Par exemple,
- vous pourrez peut-être télécharger des données sur la taxe foncière à partir d'un certain nombre de sites Web gouvernementaux.
- Ensuite, vous pouvez comparer les performances des élèves dans les écoles où les prix des logements sont similaires mais les salaires des enseignants sont différents

## Comment répondre à ces questions

- Une solution au problème des facteurs de confusion consiste à essayer de faire des comparaisons équitables en ajustant les différences observables entre les groupes.
- Par exemple,
- vous pourrez peut-être télécharger des données sur la taxe foncière à partir d'un certain nombre de sites Web gouvernementaux.
- Ensuite, vous pouvez comparer les performances des élèves dans les écoles où les prix des logements sont similaires mais les salaires des enseignants sont différents
- Vous pourriez toujours constater que les élèves apprennent davantage dans les écoles où les salaires des enseignants sont plus élevés.

# Comment répondre à ces questions

- Qu'en est-il des autres facteurs de confusion:

## Comment répondre à ces questions

- Qu'en est-il des autres facteurs de confusion:
- Niveau d'éducation des parents

## Comment répondre à ces questions

- Qu'en est-il des autres facteurs de confusion:
- Niveau d'éducation des parents
- proximité des bibliothèques publiques



# Comment répondre à ces questions

- Qu'en est-il des autres facteurs de confusion:
- Niveau d'éducation des parents
- proximité des bibliothèques publiques
- Salaire élevé des directeurs dans les écoles qui paient de gros salaires aux enseignants

## Comment répondre à ces questions

- Vous pouvez également essayer de mesurer et d'ajuster ces facteurs, mais la liste des facteurs de confusion possibles est essentiellement infinie.

## Comment répondre à ces questions

- Vous pouvez également essayer de mesurer et d'ajuster ces facteurs, mais la liste des facteurs de confusion possibles est essentiellement infinie.
- Dans de nombreuses situations, vous ne pouvez tout simplement pas mesurer et ajuster tous les facteurs de confusion possibles.

## Comment répondre à ces questions

- Vous pouvez également essayer de mesurer et d'ajuster ces facteurs, mais la liste des facteurs de confusion possibles est essentiellement infinie.
- Dans de nombreuses situations, vous ne pouvez tout simplement pas mesurer et ajuster tous les facteurs de confusion possibles.
- En réponse à ce défi, les chercheurs ont développé un certain nombre de techniques pour faire des estimations causales à partir de données non expérimentales (deuxième partie)

# Présentation de Joshua Angrist

- Ceteris Paribus: Public vs. Private University

<https://youtu.be/iPBV3BIV7jk>

**Type de relation**

# Association

- On dit que deux variables A et B sont **associées** quand l'une se trouve plus communément en présence de l'autre.
- Se détecte souvent à partir d'un tableau dit de **contingence** ou **tableau croisé** ou d'un graphique

# Association

- Exemple
- Existe-il une association entre le degré d'ouverture d'un pays et l'attitude face à la violence contre les femmes?



# Association

- Exemple
- Existe-il une association entre le degré d'ouverture d'un pays et l'attitude face à la violence contre les femmes?
- Pierotti, Rachel. (2013). "Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion." *American Sociological Review*, 78: 240-265.

# Association

- Exemple
- Existe-il une association entre le degré d'ouverture d'un pays et l'attitude face à la violence contre les femmes?
- Pierotti, Rachel. (2013). "Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion." *American Sociological Review*, 78: 240-265.
- Nous utilisons les données des enquêtes démographiques et de santé (EDS), qui représentent un ensemble de plus de 300 enquêtes représentatives à l'échelle nationale, régionale et résidentielle menées dans des pays en développement du monde entier depuis 1992.

## Association

Name	Description
beat_burnfood	Pourcentage de femmes dans chaque pays qui pensent qu'un mari a le droit de battre sa femme si elle brûle la nourriture (quantitative)
beat_burnfood_cat	Variable beat_bunfood en 4 catégories (qualitative)
no_media	Pourcentage de femmes dans chaque pays qui ont rarement accès un journal, une radio ou une télévision (quantitative)
no_media_cat	Variable no_media en 3 catégories (qualitative)
country	pays

# Association

```
## [1] "/Users/visseho/Library/CloudStorage/OneDrive-UQAM/"
## # A tibble: 6 x 11
##       X1 beat_burnfood beat_goesout sec_school no_media c
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl> <
## 1     1           4.4           18.6           25.2           1.5 A
## 2     4           4.9           19.9           67.7           8.7 A
## 3     5           2.1           10.3           67.6           2.2 A
## 4     6           0.3           3.1            46            6.4 A
## 5     7          12.1          42.5          74.6           7.4 A
## 6     8           NA           NA             24          41.9 P
## # ... with 3 more variables: beat_burnfood_cat <fct>, be
## #   no_media_cat <fct>
```

# Association

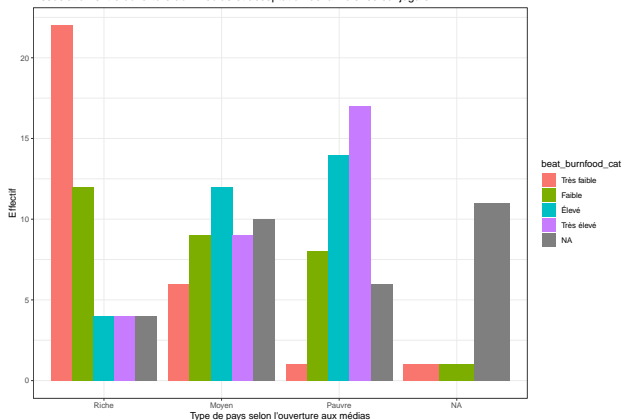
## ■ Résumé de l'information contenue dans la base de donnée

```
##          X1          beat_burnfood      beat_goesout      sec_
##  Min.      : 1.00      Min.      : 0.10      Min.      : 0.30      Min.
## 1st Qu.: 40.50      1st Qu.: 4.50      1st Qu.:11.85      1st Qu.
## Median : 79.00      Median :11.85      Median :28.10      Median
## Mean   : 80.53      Mean   :15.04      Mean   :28.60      Mean
## 3rd Qu.:119.50      3rd Qu.:22.25      3rd Qu.:42.08      3rd Qu.
## Max.    :160.00      Max.    :64.50      Max.    :82.70      Max.
##          NA's      :31      NA's      :27      NA's
##      no_media      country      year      re
##  Min.      : 0.80      Length:151      Min.      :1999      Length
## 1st Qu.:11.25      Class :character      1st Qu.:2004      Class
## Median :29.15      Mode  :character      Median :2007      Mode
## Mean   :28.40      Mean   :2007
## 3rd Qu.:43.23      3rd Qu.:2011
```

# Association

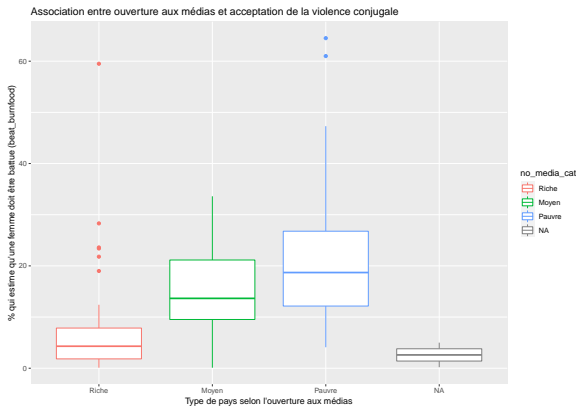
## ■ Association entre deux variables qualitatives

Association entre ouverture aux médias et acceptation de la violence conjugale



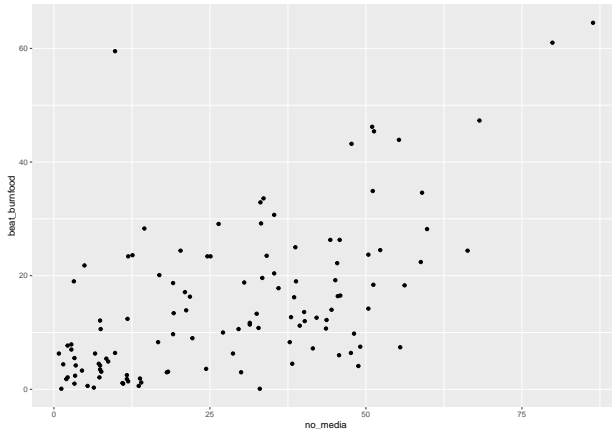
# Association

- Représentation graphique (boxplot) dans le cas d'une variable qualitative et d'une variable quantitative



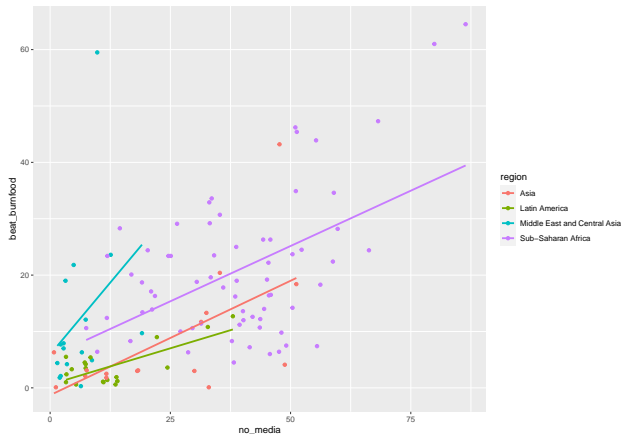
# Association

- Représentation linéaire (scatterplot) et de calcul d'indicateurs (corrélation de Pearson): dans le cas de deux variables quantitatives.





# Association



## Relation associative

- Une association (linéaire) peut être
  - **1** **positive** si les deux variables vont dans le même sens (une augmentation de l'une est associée à une augmentation de l'autre);
  - *Exemple* : éducation et revenu, durée de résidence et emploi

## Relation associative

- Une association (linéaire) peut être
  - **2 négative** si les deux variables vont dans des sens opposés (une augmentation de l'un est associée à une diminution de l'autre);
  - *Exemple* : scolarisation et racisme, revenu et obésité, niveau de développement d'un pays et niveau de mortalité infantile

# Relation associative

- Une association (linéaire) peut être
  - **3** **nulle** (Absence d'association).
  - *Exemple:*

# Causalité

## Relation causale

- L'association est une **condition nécessaire** à la causalité (Mais elle **n'est pas suffisante**).
- Toutes les associations ne sont pas causales. L'association peut arriver par hasard.

## Relation causale

- L'association est une **condition nécessaire** à la causalité (Mais elle **n'est pas suffisante**).
- Toutes les associations ne sont pas causales. L'association peut arriver par hasard.
- **L'analyse statistique à elle seule ne peut constituer une preuve d'un lien de causalité**

## Relation causale

- L'association est une **condition nécessaire** à la causalité (Mais elle **n'est pas suffisante**).
- Toutes les associations ne sont pas causales. L'association peut arriver par hasard.
- **L'analyse statistique à elle seule ne peut constituer une preuve d'un lien de causalité**
- Comparaison entre *factuel* et *contrefactuel*



## Relation causale

- L'association est une **condition nécessaire** à la causalité (Mais elle **n'est pas suffisante**).
- Toutes les associations ne sont pas causales. L'association peut arriver par hasard.
- **L'analyse statistique à elle seule ne peut constituer une preuve d'un lien de causalité**
- Comparaison entre *factuel* et *contrefactuel*
- Problème fondamental de l'inférence causale:

## Relation causale

- L'association est une **condition nécessaire** à la causalité (Mais elle **n'est pas suffisante**).
- Toutes les associations ne sont pas causales. L'association peut arriver par hasard.
- **L'analyse statistique à elle seule ne peut constituer une preuve d'un lien de causalité**
- Comparaison entre *factuel* et *contrefactuel*
- Problème fondamental de l'inférence causale:
- Il faut déduire des résultats contrefactuels

## Relation causale

- L'association est une **condition nécessaire** à la causalité (Mais elle **n'est pas suffisante**).
- Toutes les associations ne sont pas causales. L'association peut arriver par hasard.
- **L'analyse statistique à elle seule ne peut constituer une preuve d'un lien de causalité**
- Comparaison entre *factuel* et *contrefactuel*
- Problème fondamental de l'inférence causale:
- Il faut déduire des résultats contrefactuels
- Il n'y a pas de causalité sans manipulation: caractéristiques immuables

## Relation causale

- La clé pour comprendre la causalité est de penser au **contrefactuel**.
- L'inférence causale est une comparaison entre le factuel (ce qui s'est réellement passé) et le contrefactuel (ce qui se serait passé si une condition était différente).

## Relation causale

- La clé pour comprendre la causalité est de penser au **contrefactuel**.
- L'inférence causale est une comparaison entre le factuel (ce qui s'est réellement passé) et le contrefactuel (ce qui se serait passé si une condition était différente).
- Contrefactuels ne sont pas observés, sauf dans les films.

## Relation causale

- La clé pour comprendre la causalité est de penser au **contrefactuel**.
- L'inférence causale est une comparaison entre le factuel (ce qui s'est réellement passé) et le contrefactuel (ce qui se serait passé si une condition était différente).
- Contrefactuels ne sont pas observés, sauf dans les films.
- <https://www.youtube.com/watch?v=BvUbv4iwbDs&rel=0&modestbranding=1&autohide=1&showinfo=0>

## **Essais contrôlés randomisés (Expérimentation)**

# Expérimentation

- Les expériences permettent aux chercheurs d'aller au-delà des corrélations dans les données naturelles afin de répondre de manière fiable à certaines questions de cause à effet.
- À l'ère de l'analogique, les expériences étaient souvent difficiles et coûteuses d'un point de vue logistique.



# Expérimentation

- Les expériences permettent aux chercheurs d'aller au-delà des corrélations dans les données naturelles afin de répondre de manière fiable à certaines questions de cause à effet.
- À l'ère de l'analogique, les expériences étaient souvent difficiles et coûteuses d'un point de vue logistique.
- Aujourd'hui, à l'ère du numérique, les contraintes logistiques disparaissent progressivement. Non seulement il est plus facile de faire des expériences comme celles faites dans le passé, mais il est maintenant possible d'exécuter de nouveaux types d'expériences.

# Expérimentation

- Deux manières différentes de voir la chose
- 1 Expérience: Dans une expérience, un chercheur intervient dans le monde et mesure ensuite un résultat.

# Expérimentation

- Deux manières différentes de voir la chose
- 1 Expérience: Dans une expérience, un chercheur intervient dans le monde et mesure ensuite un résultat.
- Cette approche est décrite comme **«perturber et observer»**.

# Expérimentation

- Deux manières différentes de voir la chose
- 1 Expérience: Dans une expérience, un chercheur intervient dans le monde et mesure ensuite un résultat.
- Cette approche est décrite comme **«perturber et observer»**.
- Les expériences de perturbation et d'observation ne concernent qu'un seul groupe ayant reçu l'intervention, et les résultats peuvent donc conduire les chercheurs à la mauvaise conclusion.

## Expérimentation

- Dans une **expérience contrôlée randomisée**, un chercheur intervient pour certaines personnes et pas pour d'autres, et le chercheur décide quelles personnes reçoivent l'intervention par randomisation (par exemple, lancer une pièce).
- Les expériences contrôlées randomisées créent des comparaisons équitables entre deux groupes: l'un qui a reçu l'intervention et l'autre qui n'en a pas.

# Expérimentation

- 2 Dans une **expérience contrôlée randomisée**, un chercheur intervient pour certaines personnes et pas pour d'autres, et le chercheur décide quelles personnes reçoivent l'intervention par randomisation (par exemple, lancer une pièce).
- Les expériences contrôlées randomisées créent des comparaisons équitables entre deux groupes: l'un qui a reçu l'intervention et l'autre qui n'en a pas.
- En d'autres termes, les expériences contrôlées randomisées sont une solution aux problèmes de facteurs de confusion.

# Expérimentation

- 2 Dans une **expérience contrôlée randomisée**, un chercheur intervient pour certaines personnes et pas pour d'autres, et le chercheur décide quelles personnes reçoivent l'intervention par randomisation (par exemple, lancer une pièce).
- Les expériences contrôlées randomisées créent des comparaisons équitables entre deux groupes: l'un qui a reçu l'intervention et l'autre qui n'en a pas.
- En d'autres termes, les expériences contrôlées randomisées sont une solution aux problèmes de facteurs de confusion.
- Malgré les différences importantes entre les expériences et les expériences contrôlées randomisées, les chercheurs en sciences sociales utilisent souvent ces termes de manière interchangeable.

## Essais contrôlés randomisés

- Idée clé: la **randomisation** du traitement rend les groupes de **traitement** et de **contrôle** en moyenne “identiques”
- Les deux groupes sont similaires en termes de toutes les caractéristiques (**observées et non observées**)



## Essais contrôlés randomisés

- Idée clé: la **randomisation** du traitement rend les groupes de **traitement** et de **contrôle** en moyenne “identiques”
- Les deux groupes sont similaires en termes de toutes les caractéristiques (**observées et non observées**)
- Peut attribuer les différences moyennes de résultats à la différence de traitement

## Essais contrôlés randomisés

- Idée clé: la **randomisation** du traitement rend les groupes de **traitement** et de **contrôle** en moyenne “identiques”
- Les deux groupes sont similaires en termes de toutes les caractéristiques (**observées et non observées**)
- Peut attribuer les différences moyennes de résultats à la différence de traitement
- Effet du Traitement Moyen (Sample Average Treatment Effect, SATE)

# Essais contrôlés randomisés

$$SATE = \frac{1}{n} \sum_{i=1}^n Y_i(1) - Y_i(0)$$

- Essais contrôlés randomisés comme **norme d'excellence** (*Gold standard*)

## Essais contrôlés randomisés

- La SATE n'est pas directement observable.
- Pour le groupe de traitement qui a reçu le traitement, nous avons observé le résultat moyen sous le traitement, mais nous ne savons pas quel aurait été leur résultat moyen sans le traitement.

## Essais contrôlés randomisés

- La SATE n'est pas directement observable.
- Pour le groupe de traitement qui a reçu le traitement, nous avons observé le résultat moyen sous le traitement, mais nous ne savons pas quel aurait été leur résultat moyen sans le traitement.
- Le même problème existe pour le groupe témoin car ce groupe ne reçoit pas le traitement et, par conséquent, nous n'observons pas le résultat moyen qui se produirait dans les conditions de traitement.

## Essais contrôlés randomisés

- La SATE n'est pas directement observable.
- Pour le groupe de traitement qui a reçu le traitement, nous avons observé le résultat moyen sous le traitement, mais nous ne savons pas quel aurait été leur résultat moyen sans le traitement.
- Le même problème existe pour le groupe témoin car ce groupe ne reçoit pas le traitement et, par conséquent, nous n'observons pas le résultat moyen qui se produirait dans les conditions de traitement.
- Pour estimer le résultat contrefactuel moyen du traitement, nous pouvons utiliser le résultat moyen observé du groupe témoin.

## Essais contrôlés randomisés

- De même, nous pouvons utiliser le résultat moyen observé du groupe de traitement comme une estimation du résultat contrefactuel moyen pour le groupe de contrôle.
- La SATE peut être estimée en calculant la différence entre le résultat moyen entre les groupes de traitement et témoin

## Essais contrôlés randomisés

- De même, nous pouvons utiliser le résultat moyen observé du groupe de traitement comme une estimation du résultat contrefactuel moyen pour le groupe de contrôle.
- La SATE peut être estimée en calculant la différence entre le résultat moyen entre les groupes de traitement et témoin
- En clair, la grande question de la causalité n'est qu'une question de soustraction :)



## Essais contrôlés randomisés

- Dans un essai contrôlé randomisé (ECR), chaque unité est assignée de manière aléatoire au groupe de traitement ou au groupe de contrôle.

## Essais contrôlés randomisés

- Dans un essai contrôlé randomisé (ECR), chaque unité est assignée de manière aléatoire au groupe de traitement ou au groupe de contrôle.
- La **randomisation** de l'assignation de traitement garantit que la différence moyenne de résultats entre les groupes de traitement et de contrôle peut être attribuée uniquement au traitement, car les deux groupes sont en moyenne identiques pour toutes les caractéristiques de prétraitement (observées et non observées).

# Essais contrôlés randomisés

## 1 Forces

- **Validité interne** - mesure dans laquelle les hypothèses de causalité sont satisfaites dans l'étude

## 2 Limites

- **Validité externe** - mesure dans laquelle les conclusions peuvent être généralisées au-delà d'une étude particulière

## 3 Autres problèmes

- Explication causale faible

# Essais contrôlés randomisés

## 1 Forces

- **Validité interne** - mesure dans laquelle les hypothèses de causalité sont satisfaites dans l'étude

## 2 Limites

- **Validité externe** - mesure dans laquelle les conclusions peuvent être généralisées au-delà d'une étude particulière

## 3 Autres problèmes

- Explication causale faible
- Considérations éthiques

# Essais contrôlés randomisés

## 1 Forces

- **Validité interne** - mesure dans laquelle les hypothèses de causalité sont satisfaites dans l'étude

## 2 Limites

- **Validité externe** - mesure dans laquelle les conclusions peuvent être généralisées au-delà d'une étude particulière

## 3 Autres problèmes

- Explication causale faible
- Considérations éthiques
- Possibilité de contamination

# Applications

# Exemple 1: Est-ce que la mammographie sauve des vies?

## Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial



OPEN ACCESS

Anthony B Miller *professor emeritus*<sup>1</sup>, Claus Wall *data manager*<sup>1</sup>, Cornelia J Baines *professor emerita*<sup>1</sup>, Ping Sun *statistician*<sup>2</sup>, Teresa To *senior scientist*<sup>3</sup>, Steven A Narod *professor*<sup>1,2</sup>

<sup>1</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario M5T 3M7, Canada; <sup>2</sup>Women's College Research Institute, Women's College Hospital, Toronto, Ontario M5G 1N8, Canada; <sup>3</sup>Child Health Evaluative Services, The Hospital for Sick Children, Toronto, Ontario, Canada

### Abstract

**Objective** To compare breast cancer incidence and mortality up to 25 years in women aged 40-59 who did or did not undergo mammography screening.

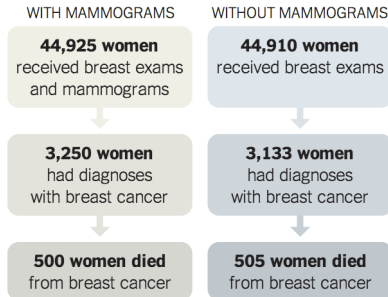
**Design** Follow-up of randomised screening trial by centre coordinators,

**Conclusion** Annual mammography in women aged 40-59 does not reduce mortality from breast cancer beyond that of physical examination or usual care when adjuvant therapy for breast cancer is freely available. Overall, 22% (106/484) of screen detected invasive breast cancers were over-diagnosed, representing one over-diagnosed breast cancer for every 424 women who received mammography screening in the trial.

# Exemple 1: Est-ce que la mammographie sauve des vies?

## Study Results

A large, 25-year study of Canadian women aged 40 to 59 found no benefit for women who were randomly assigned to have mammograms.



The death rate from breast cancer was the same in both groups, but **1 in 424 women** who had mammograms received unnecessary cancer treatment, including surgery, chemotherapy and radiation.



## Exemple 2 discrimination raciale sur le marché du travail

### 1 Question de recherche

- La discrimination raciale existe-t-elle sur le marché du travail?
- Ou bien les disparités raciales dans le taux de chômage devraient-elles être attribuées à d'autres facteurs tels que les écarts raciaux dans le niveau d'instruction?

## Exemple 2 discrimination raciale sur le marché du travail

### 2 Expérimentation

- En réponse aux annonces dans les journaux, les chercheurs ont envoyé les CV de candidats fictifs à des employeurs potentiels.
- Changé seulement le nom du demandeur d'emploi
  - Noms afro-américains
  - Noms à consonance caucasienne
- Les autres informations sont inchangées

## Exemple 2 discrimination raciale sur le marché du travail

### 3 Variable dépendante

- Taux de rappel

## Exemple 2 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain

## Exemple 2 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain
- **Groupe de traitement** (unités traitées): Afro-américains

## Exemple 2 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain
- **Groupe de traitement** (unités traitées): Afro-américains
- **Groupe de contrôle** (unités non traitées): Caucasiens

## Exemple 2 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain
- **Groupe de traitement** (unités traitées): Afro-américains
- **Groupe de contrôle** (unités non traitées): Caucasiens
- **Réponse** (variable de réponse) **Y**: si un rappel a été effectué

## Exemple 2 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain
- **Groupe de traitement** (unités traitées): Afro-américains
- **Groupe de contrôle** (unités non traitées): Caucasiens
- **Réponse** (variable de réponse) **Y**: si un rappel a été effectué
- Que signifie "**T cause Y**"?



## Exemple 2 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain
- **Groupe de traitement** (unités traitées): Afro-américains
- **Groupe de contrôle** (unités non traitées): Caucasiens
- **Réponse** (variable de réponse) **Y**: si un rappel a été effectué
- Que signifie “**T cause Y**”?
- Contrefactuels, “**Quoi si**” : Les Afro-Américains auraient-ils été rappelés s'ils n'avaient pas de noms afro-américains?

## Exemple 2 discrimination raciale sur le marché du travail

- **Deux résultats possibles:**  $Y(1)$  et  $Y(0)$
- **Effet causal:**  $Y(1) - Y(0)$
- **Problème fondamental d'inférence causale:** un seul des deux résultats potentiels est observable

## Exemple 2 discrimination raciale sur le marché du travail

- Comment pouvons-nous comprendre les contrefactuels?
  - L'association n'est pas un lien de causalité
  - Trouvez une unité similaire!  $\implies$  **Matching**
  - Est-ce-que Jamal n'a été rappelé à cause de sa race?
  - Trouver une personne blanche qui ressemble à Jamal
- Le problème: on ne peut pas correspondre sur tout
- Facteurs de **confusion non observés**: variables associées au traitement et au résultat  $\implies$  **biais de sélection**

## Exemple 2 discrimination raciale sur le marché du travail

- La clé pour comprendre la causalité est de penser au contrefactuel. L'inférence causale est une comparaison entre le factuel (ce qui s'est réellement passé) et le contrefactuel (ce qui se serait passé si une condition était différente).

CV (i)	Noms à consonnance afro-américain ( $T_i$ )	Appellé pour interview?		Age	Niveau d'éducation
		$Y_i(1)$	$Y_i(0)$		
1	1	1	?	25	Collège
2	0	?	0	55	Secondaire
3	0	?	1	40	Collège
n	1	0	?	62	Secondaire

Figure 1: A caption

## Labo 4.1: Données expérimentales

## Exemple 2 discrimination raciale sur le marché du travail

## Exemple 2 discrimination raciale sur le marché du travail

##	firstname	sex	race	call
## 1	Allison	female	white	0
## 2	Kristen	female	white	0
## 3	Lakisha	female	black	0
## 4	Latonya	female	black	0
## 5	Carrie	female	white	0
## 6	Jay	male	white	0

## Exemple 2 discrimination raciale sur le marché du travail

```
freq(resume$sex)
```

```
## Frequencies
```

```
## resume$sex
```

```
## Type: Character
```

```
##
```

##		Freq	% Valid	% Valid Cum.	% Total
##	-----	-----	-----	-----	-----
##	female	3746	76.92	76.92	76.92
##	male	1124	23.08	100.00	23.08
##	<NA>	0			0.00
##	Total	4870	100.00	100.00	100.00



## Exemple 2 discrimination raciale sur le marché du travail

```
freq(resume$race)
```

```
## Frequencies
```

```
## resume$race
```

```
## Type: Character
```

```
##
```

```
##           Freq  % Valid  % Valid Cum.  % Total
```

```
## -----
```

```
##      black  2435    50.00      50.00    50.00
```

```
##      white  2435    50.00     100.00    50.00
```

```
##      <NA>     0      0.00      100.00    0.00
```

```
##      Total  4870   100.00     100.00   100.00
```

## Exemple 2 discrimination raciale sur le marché du travail

```
freq(resume$call)
```

```
## Frequencies
```

```
## resume$call
```

```
## Type: Integer
```

```
##
```

```
##           Freq  % Valid  % Valid Cum.  % Total
```

```
## -----
```

```
##           0   4478    91.95      91.95    91.95
```

```
##           1    392     8.05     100.00     8.05
```

```
##          <NA>     0      0.00      0.00
```

```
##          Total  4870   100.00    100.00   100.00
```

## Y'a-t-il discrimination ou pas?

```
ctable(resume$race, resume$call)
```

```
## Cross-Tabulation, Row Proportions
```

```
## race * call
```

```
## Data Frame: resume
```

```
##
```

```
## -----
```

```
##           call                0                1                To
```

```
##    race
```

```
##  black      2278 (93.6%)    157 ( 6.4%)    2435 (100.0%)
```

```
##  white      2200 (90.3%)    235 ( 9.7%)    2435 (100.0%)
```

```
##  Total      4478 (92.0%)    392 ( 8.0%)    4870 (100.0%)
```

```
## -----
```

■  $\text{SATE} = 9,65 - 6,45 = 3,2\%$

## Est-ce que les deux groupes étaient similaires au début?

```
ctable(resume$race, resume$sex)
```

```
## Cross-Tabulation, Row Proportions
```

```
## race * sex
```

```
## Data Frame: resume
```

```
##
```

```
## -----
```

```
##           sex           female           male           Total
```

```
##    race
```

```
##  black      1886 (77.5%)      549 (22.5%)      2435 (100%)
```

```
##  white      1860 (76.4%)      575 (23.6%)      2435 (100%)
```

```
##  Total      3746 (76.9%)      1124 (23.1%)      4870 (100%)
```

```
## -----
```

## Exemple 3: Expérimentation à l'ère digitale

Lire chapitre 4 de Matthews Salganik

# Présentation par Joshua Angrist

- Randomized Trials: The Ideal Weapon

<https://www.youtube.com/watch?v=eGRd8jBdNYg>

## Références

## Références

- Salganik(<https://www.bitbybitbook.com/en/1st-ed/running-experiments/>)
- Gerber, Alan S. and Donald P. Green, Field Experiments: Design, Analysis, and Interpretation. Chapters 1 and 2.
- Stephen, Morgan & Christopher Winship. 2018. Counterfactuals and Causal Inference: methods and principles for Social research, 2nd edition. Chapter 2.
- Shelley, Clark et al. 2019. "The Impact of Childcare on Poor Urban Women's Economic Empowerment in Africa". Demography, 56, 4.
- Mirna Safi et Patrick Simon. 2013. "Les discriminations ethniques et raciales dans l'enquête Trajectoires et Origines: représentations, expériences subjectives et situations vécues." Économie et Statistiques, No 464-465-466. (À discuter en classe)