

Qu'est-ce qui pourrait aussi l'expliquer

Un facteur de confusion cause le résultat et diffère entre les groupes. Quatre candidats séparent les districts de cette épidémie — l'un a été retiré, l'un n'est pas mesuré, et l'un n'est pas un facteur de confusion du tout : c'est le mécanisme.

Pierre Robentz Cassion

Leçon 7 sur 8

Santé publique et épidémiologie appliquée — Ce que la comparaison peut affirmer

Ce que couvre cette leçon

- La définition, et pourquoi il vaut la peine d'être strict
- Les quatre candidats de cette épidémie
- Stratifiez pour le voir
- Quand la stratification s'épuise
- Les trois questions à poser à toute comparaison
- La suite

La définition, et pourquoi il vaut la peine d'être strict

- elle est **associée à l'exposition** — elle diffère entre les groupes comparés ;
- elle est une **cause du résultat**, indépendamment de l'exposition ;
- elle n'est **pas sur le chemin causal** entre l'exposition et le résultat.

Les quatre candidats de cette épidémie

Candidat	Diffère entre districts ?	Cause le résultat ?	Sur le chemin ?
Structure par âge	Oui, nettement	Oui	Non — facteur de confusion, déjà retiré
Source d'eau et densité	Presque certainement	Oui	Non — facteur de confusion, non mesuré
Distance au traitement	Oui	Oui, pour la létalité	Oui — c'est le mécanisme
Qualité du registre	Oui	Non	Non — pas un facteur de confusion, un biais

Les quatre candidats de cette épidémie

- La structure par âge — a été traitée
- La source d'eau, l'assainissement et la promiscuité — sont les causes réelles de la transmission du choléra, elles. . .

Les quatre candidats de cette épidémie — En Python

```
print("Available for adjustment: age, sex, district")
print("Known to matter and unavailable: water source, sanitation, crowding, "
      "displacement status")
```

Les quatre candidats de cette épidémie — En R

The list of what you could not adjust for belongs in the limitations section.

Les quatre candidats de cette épidémie

- La distance au traitement — est l'intéressante et n'est pas du tout un facteur de confusion
- La qualité du registre — n'est pas non plus un facteur de confusion

Stratifiez pour le voir — En Python

```
import pandas as pd

cases = pd.read_csv("cholera-line-list-2024.v1.csv")
with_outcome = cases[cases["outcome"].notna()]

stratified = (
    with_outcome.assign(died=with_outcome["outcome"] == "died")
    .groupby(["district", "age_band"])
    .agg(cases=("died", "size"), deaths=("died", "sum"))
)
stratified["cfr"] = stratified["deaths"] / stratified["cases"]
print((stratified["cfr"].unstack() * 100).round(1))
```

Stratifiez pour le voir — En R

```
library(dplyr)

cases |>
  filter(!is.na(outcome)) |>
  summarise(cases = n(), cfr = mean(outcome == "died"), .by = c(district, age_band)) |>
  tidyr::pivot_wider(id_cols = age_band, names_from = district, values_from = cfr)
```

- L'écart persiste-t-il à l'intérieur des strates? — Si la létalité de Nord est plus élevée dans chaque tranche d'âge,...
- Les strates sont-elles cohérentes? — Un écart grand dans une tranche et absent des autres est une modification d'effet...
- La modification d'effet est un constat; la confusion est une nuisance — Réduire la première à un nombre unique détruit...

Quand la stratification s'épuise — En Python

```
cells = with_outcome.groupby(["district", "age_band", "sex"]).size()
print(f"{len(cells)} cells, smallest {cells.min()}, {(cells < 30).sum()} below 30")
```

Quand la stratification s'épuise — En R

```
cases |> filter(!is.na(outcome)) |> count(district, age_band, sex) |>  
  summarise(cells = n(), smallest = min(n), under_30 = sum(n < 30))
```

Les trois questions à poser à toute comparaison

- 1. Qu'est-ce qui diffère entre ces groupes en dehors de ce que j'étudie ? — Listez-le
- 2. Lesquels causent le résultat ? — Ceux-là seuls sont des facteurs de confusion
- 3. Lesquels sont sur le chemin causal ? — N'ajustez pas dessus

Les trois questions à poser à toute comparaison — En Python

```
adjustment_plan = pd.DataFrame({
    "variable": ["age", "sex", "water source", "delay to treatment", "register quality"],
    "differs": [True, False, True, True, True],
    "causes_outcome": [True, True, True, True, False],
    "on_causal_path": [False, False, False, True, False],
    "action": ["adjust", "no need", "unmeasured; state it", "do not adjust",
               "fix the register"],
})
print(adjustment_plan)
```

Les trois questions à poser à toute comparaison — En R

```
tibble::tribble(  
  ~variable,      ~action,  
  "age",          "adjust",  
  "water source", "unmeasured; state it",  
  "delay",        "do not adjust; it is the mechanism",  
  "register",      "fix it; this is bias, not confounding"  
)
```

Les trois questions à poser à toute comparaison

- Écrivez ce tableau avant l'analyse, non après — Décider sur quoi ajuster une fois qu'on a vu quel ajustement donne la...

- Vous avez nommé ce qui pourrait aussi expliquer un écart et retiré ce que vous pouviez.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/public-health-epidemiology/
epi-07-confounding](https://data-analysis.cassion.dev/fr/cours/public-health-epidemiology/epi-07-confounding)