

La réponse était fixée avant la mesure du premier enfant

Vingt-quatre écoles de cinquante enfants chacune peuvent détecter 4,9 points.
Les mêmes 1 200 enfants randomisés individuellement pourraient en détecter 2,3.
Le protocole a décidé de ce que l'évaluation était capable de trouver, des années avant que quiconque lise un résultat.

Pierre Robentz Cassion

Leçon 6 sur 8

Méthodes d'évaluation d'impact — Des règles, et des limites

Ce que couvre cette leçon

- Calculez ce que le protocole peut trouver
- D'où vient le nombre
- Des grappes, pas des enfants
- La corrélation intra-classe est l'entrée que personne n'a
- Ce qu'une étude sous-dimensionnée a le droit de conclure
- La puissance a posteriori n'existe pas
- Rapportez-le en entier
- La suite

Calculez ce que le protocole peut trouver — En Python

```
import numpy as np

SD = 0.1429          # within-school SD of attendance, from the register
ICC = 0.065          # intra-cluster correlation, from the empty mixed model
M = 50               # children per school
Z_ALPHA, Z_BETA = 1.96, 0.84      # 5% two-sided, 80% power

deff = 1 + (M - 1) * ICC
print(f"design effect {deff:.2f}")

def mde(k_treated, k_control):
    n1, n0 = k_treated * M, k_control * M
    return (Z_ALPHA + Z_BETA) * SD * np.sqrt(deff * (1 / n1 + 1 / n0))

print(f"15 vs 9 schools: {mde(15, 9) * 100:.2f} points")
```

Calculez ce que le protocole peut trouver — En R

clusterPower::cpa.normal(), or the formula above. Either way, before the study.

Calculez ce que le protocole peut trouver

- Effet de plan 4,19. Effet minimal détectable 4,88 points
- Cette évaluation n'aurait jamais pu trouver un effet inférieur à environ cinq points de présence — quoi qu'ait fait. . .

D'où vient le nombre

Entrée	Ici	Décision ou fait ?
Écart-type du résultat	0,143	Fait, tiré d'un pilote ou d'un passage antérieur
Corrélation intra-classe	0,065	Fait, et celui que personne n'a d'ordinaire
Grappes par bras	15 et 9	Décision
Enfants par grappe	50	Décision, et le levier le plus faible

D'où vient le nombre — En Python

```
for k in (12, 24, 40, 60, 100):  
    print(f"{k:3} schools: MDE {mde(k // 2, k // 2) * 100:.2f} points")
```

D'où vient le nombre — En R

The whole planning conversation in five rows.

D'où vient le nombre

Écoles	Enfants	EMD
12	600	6,68 pts
24	1 200	4,73 pts
40	2 000	3,66 pts
60	3 000	2,99 pts
100	5 000	2,32 pts

D'où vient le nombre

- Diviser par deux l'effet détectable coûte environ quatre fois plus d'écoles — La précision s'améliore comme la racine. . .

Des grappes, pas des enfants — En Python

```
individual = (Z_ALPHA + Z_BETA) * SD * np.sqrt(4 / 1200)
print(f"1,200 children, individually randomised: {individual * 100:.2f} points")
print(f"1,200 children in 24 schools: {mde(12, 12) * 100:.2f} points")
print(f"ratio: {np.sqrt(deff):.2f}")
```

Des grappes, pas des enfants — En R

Same children, same outcome, same budget. Twice the detectable effect.

Des grappes, pas des enfants

- 2,31 points individuellement, 4,73 points en grappes — un facteur 2,05, qui est la racine carrée de l'effet de plan
- Ajouter des enfants aux écoles existantes n'aide presque pas — Avec une CCI de 0,065, passer de 50 à 100 enfants par. . .
- Ajouter des écoles aide proportionnellement — C'est la règle de planification : quand le programme est attribué par. . .

La corrélation intra-classe est l'entrée que personne n'a — En Python

```
for icc in (0.01, 0.03, 0.065, 0.10, 0.20):  
    d = 1 + (M - 1) * icc  
    m = (Z_ALPHA + Z_BETA) * SD * np.sqrt(d * (1 / 600 + 1 / 600))  
    print(f"ICC {icc:.3f}: design effect {d:5.2f}, MDE {m * 100:.2f} points")
```

La corrélation intra-classe est l'entrée que personne n'a — En R

Vary the ICC across the plausible range and report the MDE for each.

La corrélation intra-classe est l'entrée que personne n'a

CCI	Effet de plan	EMD, 24 écoles
0,01	1,49	2,82 pts
0,03	2,47	3,63 pts
0,065	4,19	4,73 pts
0,10	5,90	5,61 pts
0,20	10,80	7,59 pts

La corrélation intra-classe est l'entrée que personne n'a

- L'EMD va de 2,8 à 7,6 points sur la plage plausible des CCI — Un calcul de puissance citant une seule CCI sans. . .
- Où en trouver une — Un passage antérieur de la même enquête, une étude publiée dans le même secteur et le même pays, ou. . .

Ce qu'une étude sous-dimensionnée a le droit de conclure

- « Aucun effet supérieur à X n'a été détecté. » — Avec un EMD de 4,9 points et des différences de différences observées. . .
- « L'étude n'était pas conçue pour détecter un effet de l'ampleur qui compte. » — Si le programme serait jugé valable à. . .
- « Il faut plus d'unités, et en voici le nombre. » — Le tableau ci-dessus, avec la ligne que devrait viser une. . .
- Trois affirmations qui ne sont pas permises — « Le programme n'a eu aucun effet » — les données ne peuvent pas l'étayer

La puissance a posteriori n'existe pas — En Python

```
observed = -0.0096
se = 0.0129
print(f"observed effect {observed * 100:+.2f} points, SE {se * 100:.2f}")
```

La puissance a posteriori n'existe pas — En R

Do not compute power from the observed effect. It is a rearrangement of p.

La puissance a posteriori n'existe pas

- La puissance calculée à partir de l'effet observé est une fonction monotone de votre valeur p — elle n'ajoute aucune. . .
- La quantité utile est l'intervalle de confiance — comme l'a établi le cours de statistiques, et l'EMD calculé depuis le. . .

Rapportez-le en entier — Exemple (suite)

Statistical power, school feeding evaluation

Design 15 treated and 9 control schools, ~50 children each.

Inputs outcome SD 0.143, ICC 0.065 (estimated from this study's own empty mixed model), 5% two-sided, 80% power.

Design effect 4.19. Minimum detectable effect 4.88 percentage points.

The observed difference-in-differences is -0.96 points (95% CI -3.5 to +1.6). Effects larger than about 3.5 points in either direction are ruled out; smaller effects are not, and the design could not have detected them.

This is reported as a null result about large effects. It is not evidence that the programme has no effect.

Sensitivity: across ICCs from 0.01 to 0.20 the minimum detectable effect

Rapportez-le en entier — Exemple (suite)

ranges from 2.8 to 7.6 points. The ICC used is estimated from 24 clusters and is itself imprecise.

A future evaluation targeting a 3-point effect requires about 60 schools. Adding children to the existing 24 does not help: doubling to 100 per school moves the minimum detectable effect from 4.73 to 4.45 points.

- Le dernier paragraphe est celui sur lequel un programme peut agir — et c'est la raison de calculer la puissance après. . .

- Tous les protocoles de ce cours ont été reconstitués après coup.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/impact-evaluation-methods/
ie-06-the-effect-you-could-have-detected](https://data-analysis.cassion.dev/fr/cours/impact-evaluation-methods/ie-06-the-effect-you-could-have-detected)