

Évaluation d’une approche d’augmentation de données par génération de données synthétiques pour l’estimation précoce de l’effet traitement dans un essai clinique de survie : étude de simulation rétrospective.



CONTEXTE

Le recrutement des patients constitue l’étape la plus longue et la plus coûteuse des essais cliniques, retardant souvent l’accès à des thérapies efficaces. Des méthodes robustes d’augmentation des données par génération de données synthétiques pourraient permettre d’anticiper les résultats d’un essai avant la fin de l’inclusion, de consolider des analyses intermédiaires ou d’exploiter au mieux les essais cliniques n’ayant pu atteindre leur cible de recrutement.

OBJECTIFS

Évaluer la capacité d’une approche d’augmentation de données par la méthode Avatar à reproduire l’estimation finale d’un essai clinique de survie à partir de cohortes partiellement recrutées, et identifier le niveau minimal d’inclusion associé à des performances proches de celles observées à 90 % de recrutement.

JEU DE DONNÉES EXEMPLE

- Essai thérapeutique ACTG 175 dans le VIH – Jeu de données en open access (AIDS Clinical Trial Group)

- Design : essai clinique randomisé, contrôlé, en double aveugle, en bras parallèles, zidovudine (n = 619) comparé à zidovudine + didanosine (n = 513)

- Outcome : délai jusqu’à une diminution ≥ 50 % du taux de CD4

- Analyse initiale : modèle de Cox

- Résultats : hazard ratio de 0,50 (IC95 % : 0,39–0,63) Beta (Log HR) = -0,7037

Références

Hammer SM, Katzenstein DA, Hughes MD, Gundacker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, Niu M, Hirsch MS, Merigan TC. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. AIDS Clinical Trials Group Study 175 Study Team. N Engl J Med. 1996 Oct 10;335(15):1081-90. doi: 10.1056/NEJM199610103351501. PMID: 8813038. <https://search.r-project.org/CRAN/refmans/speff2trial/html/ACTG175.html>

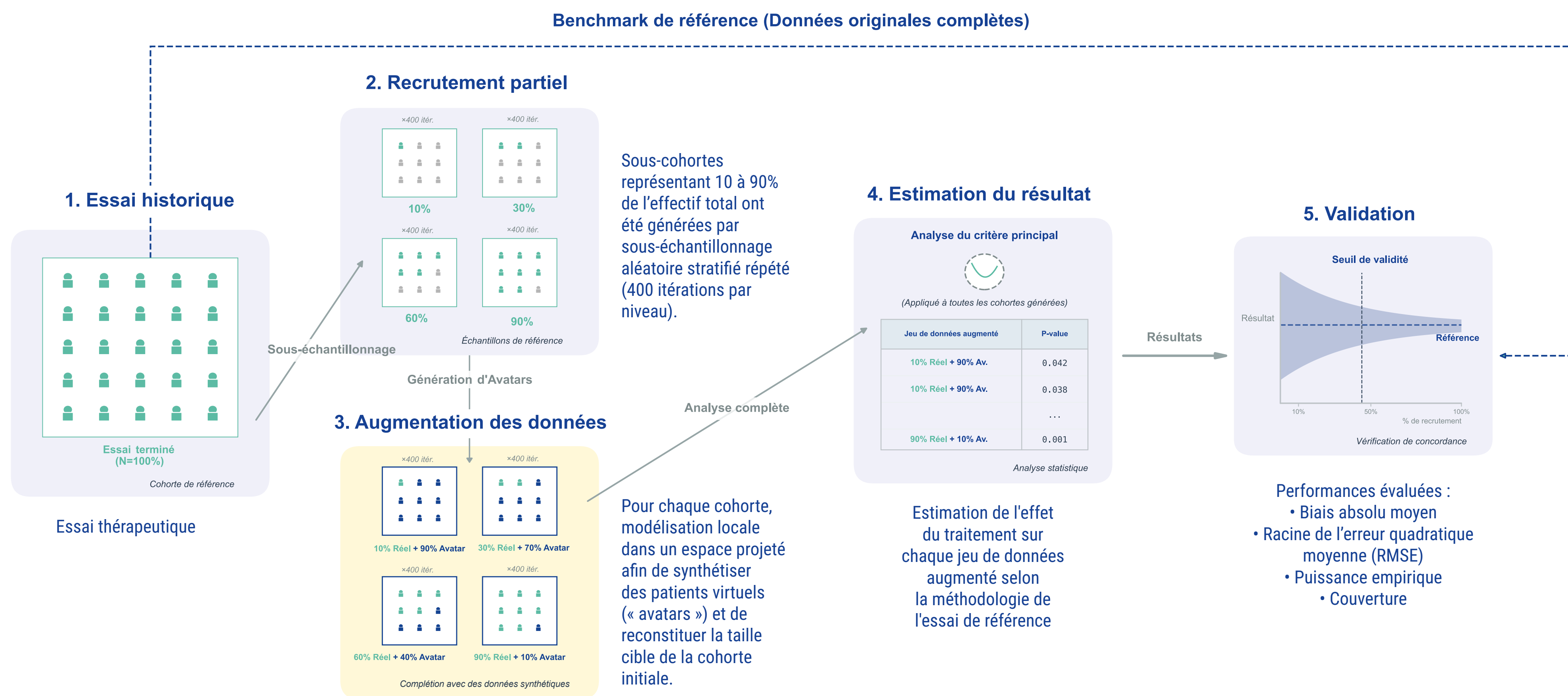
<https://kmcconeghy.github.io/jumble/reference/ACTG175.html>
<https://r-packages.io/datasets/ACTG175>

CONCLUSION

Dans cette étude de simulation rétrospective, l’augmentation de données par la méthode Avatar a permis de rapprocher progressivement les estimations intermédiaires de l’analyse finale à mesure que la proportion de données observées augmentait. Une stabilisation des performances était observée dès 50% d’inclusion. Ces premiers résultats suggèrent que la génération de données synthétiques pourrait constituer un levier majeur pour sécuriser les analyses intermédiaires et optimiser le succès des essais cliniques. Bien qu’une validation sur des contextes thérapeutiques variés reste requise, cette approche offre de nouvelles perspectives pour accélérer la recherche clinique.

MÉTHODES

Prédiction précoce des résultats d’essais cliniques par augmentation Avatar



Méthodologie de génération d’avatar

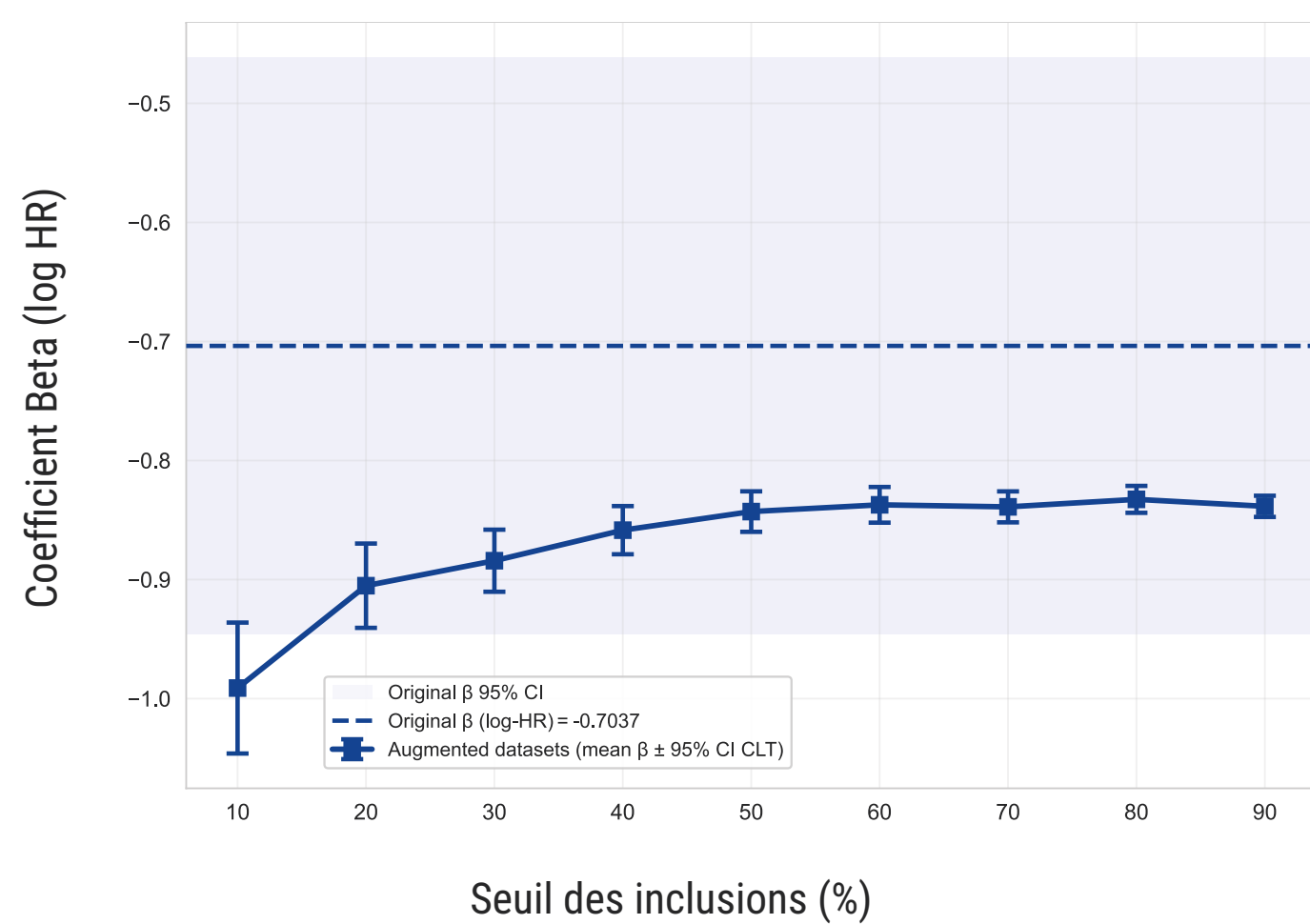
Avatar est une méthode de génération de données synthétiques centrée sur l’individu et préservant la confidentialité, qui fonctionne en plusieurs étapes :

1. Projection des données originales dans un espace continu standardisé
2. Identification des k plus proches voisins pour chaque individu
3. Génération d’un « avatar » synthétique via une perturbation locale pour chaque individu

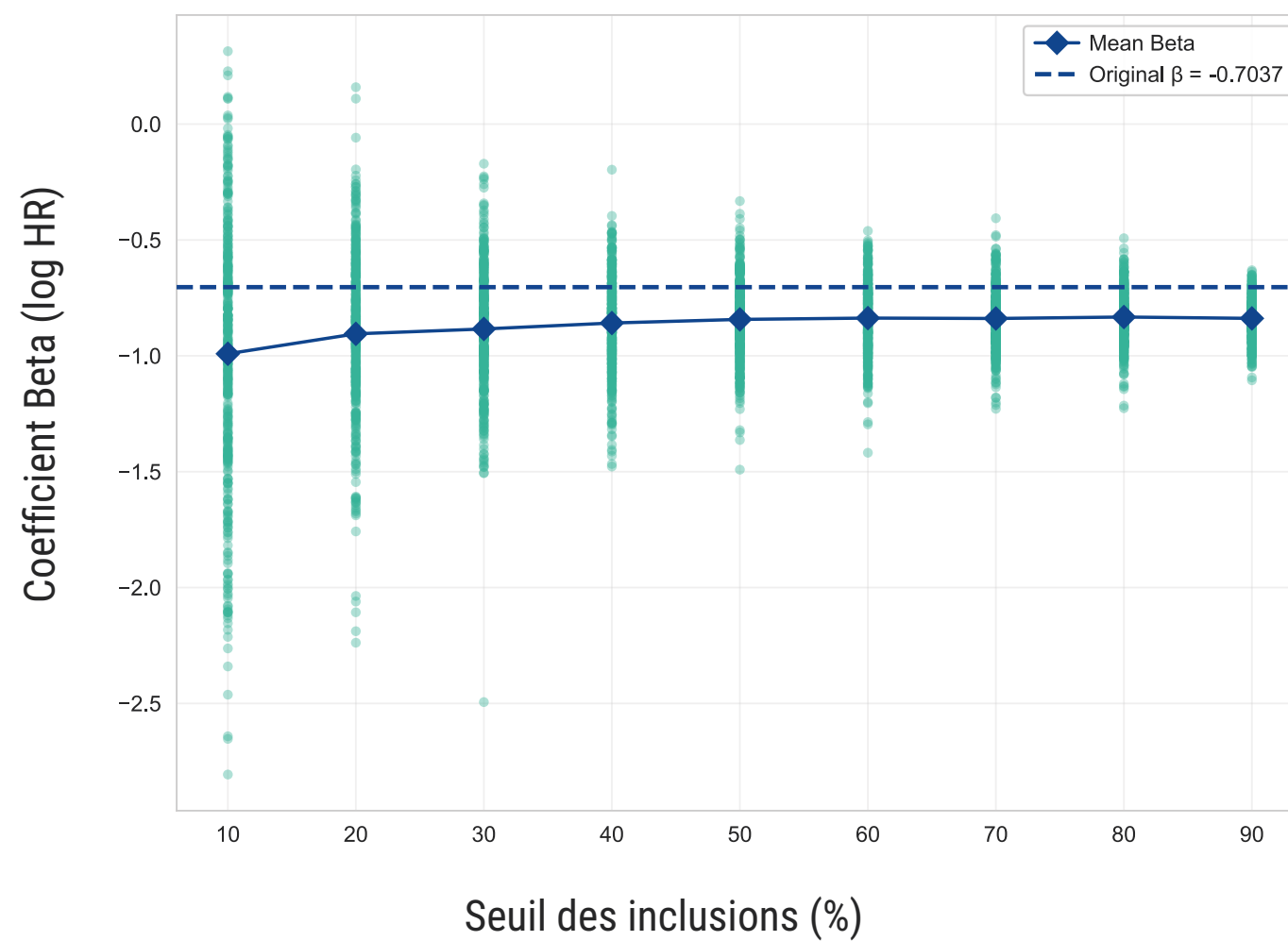
RÉSULTATS

Estimation du résultat

Coefficient Beta (log HR)

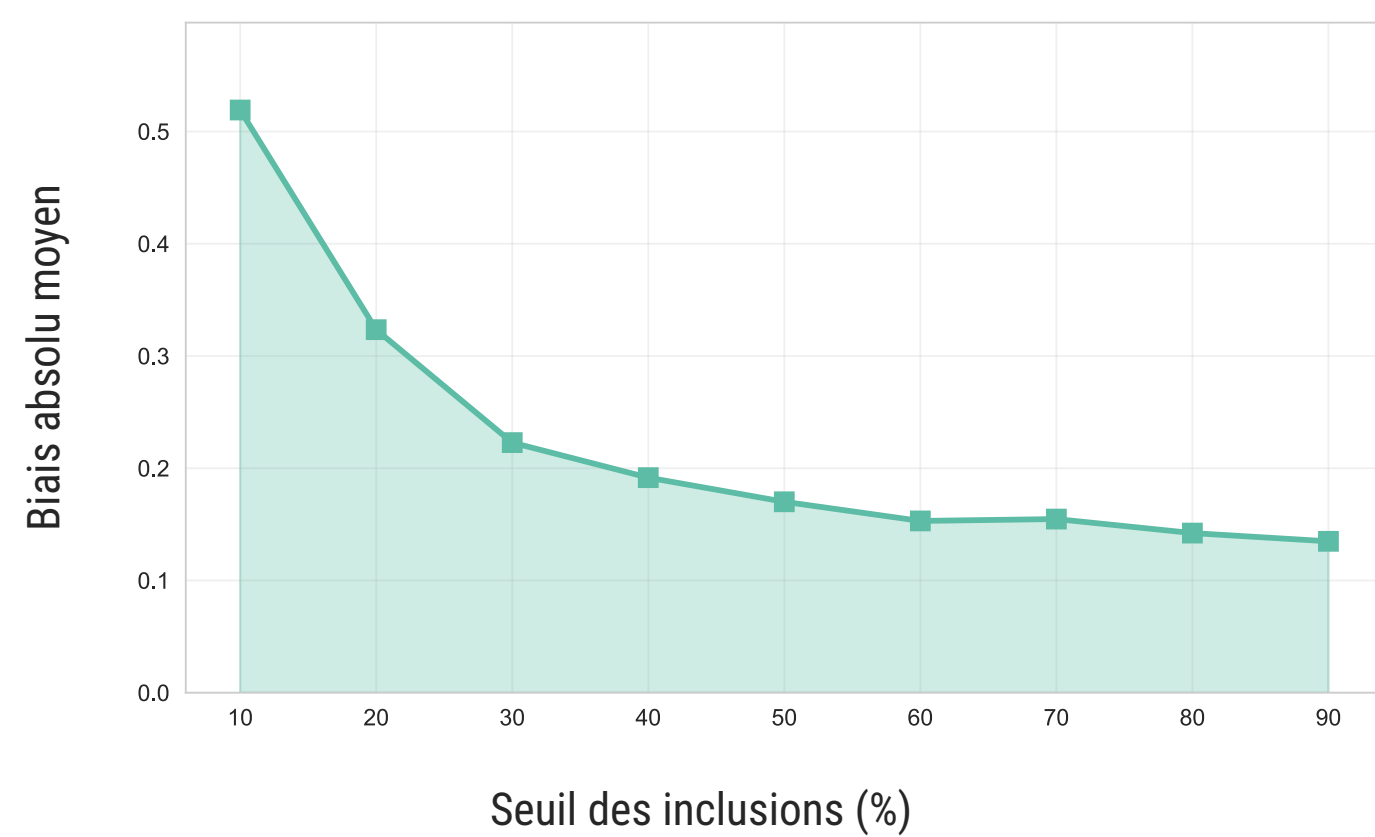


Variabilité de beta (400 itérations par seuil d’inclusion)

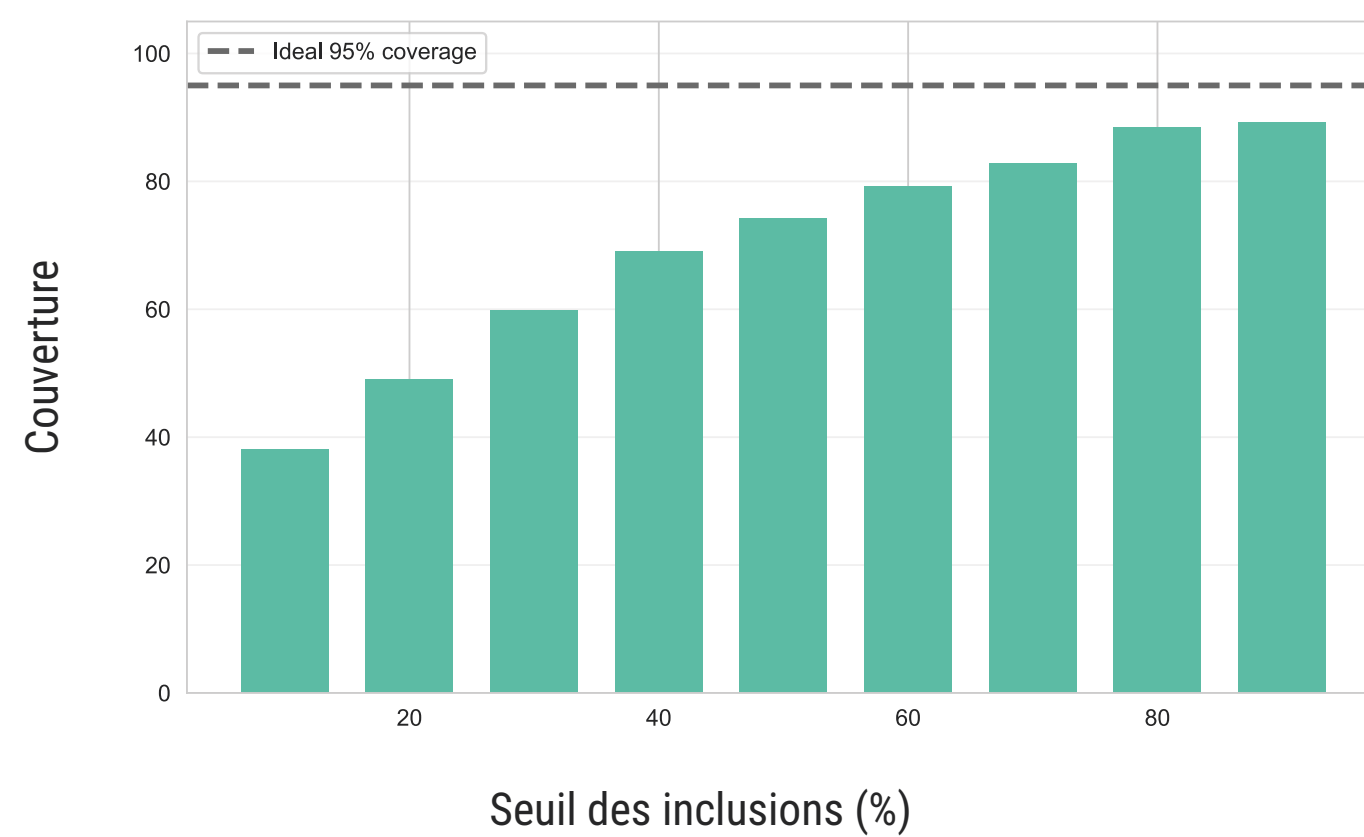


Validation des performances

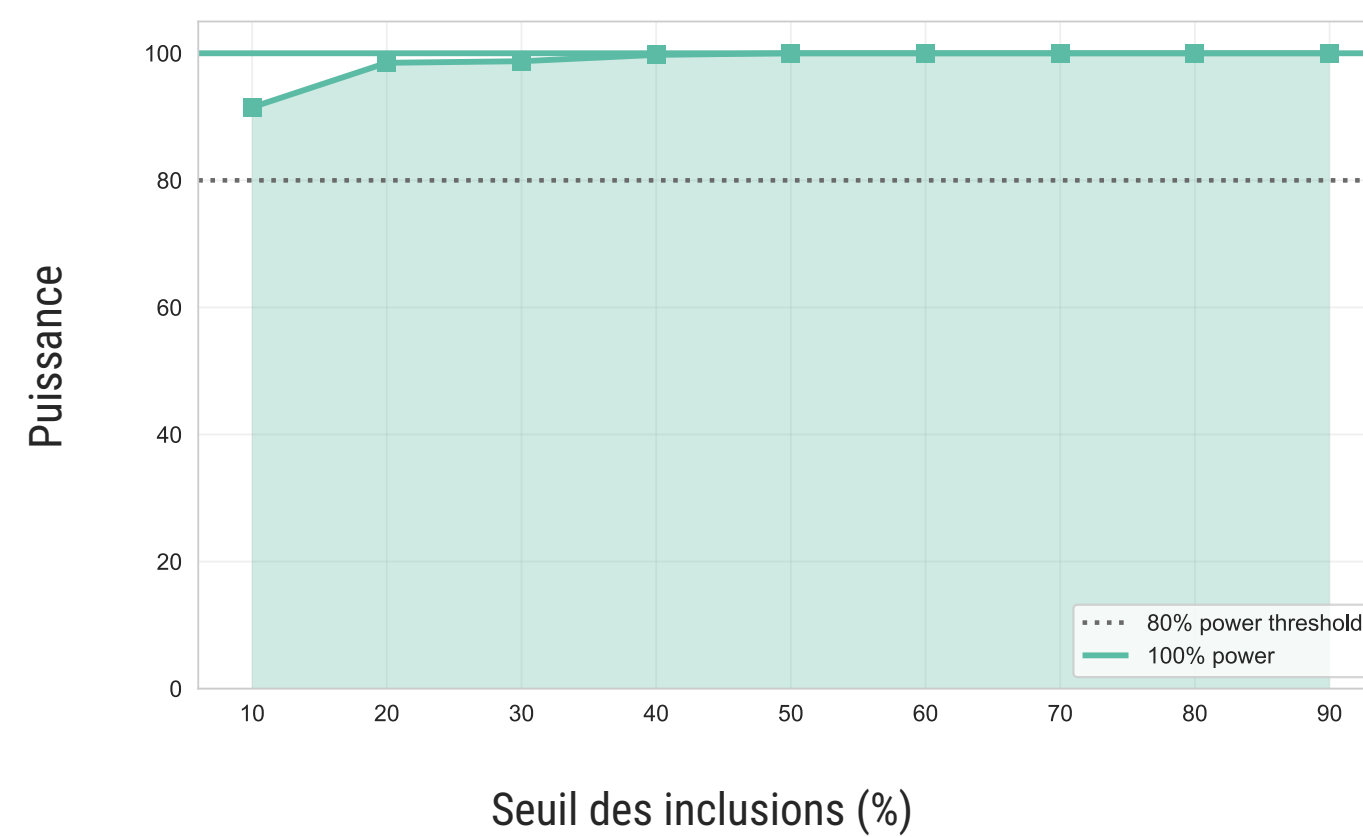
Biais absolu moyen



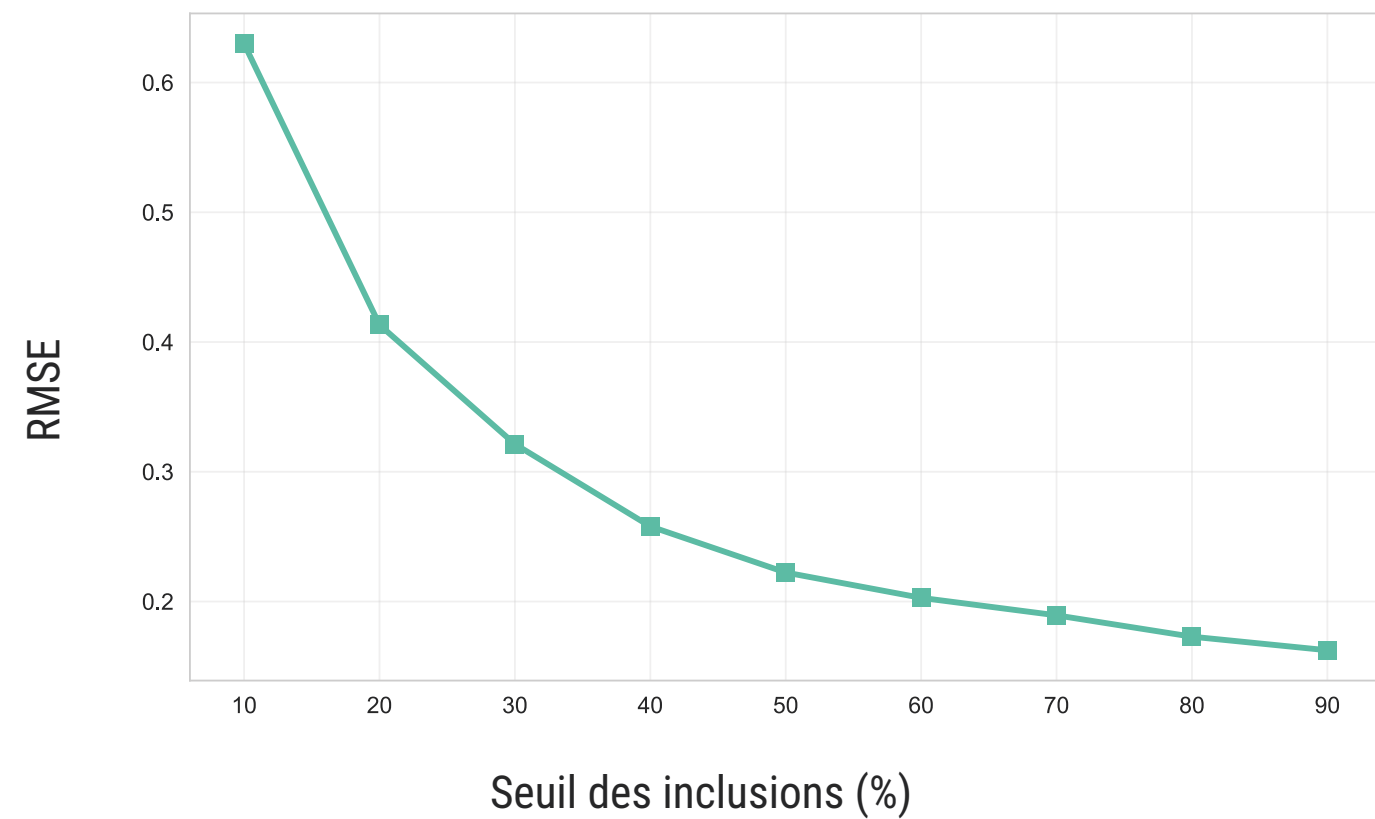
Couverture



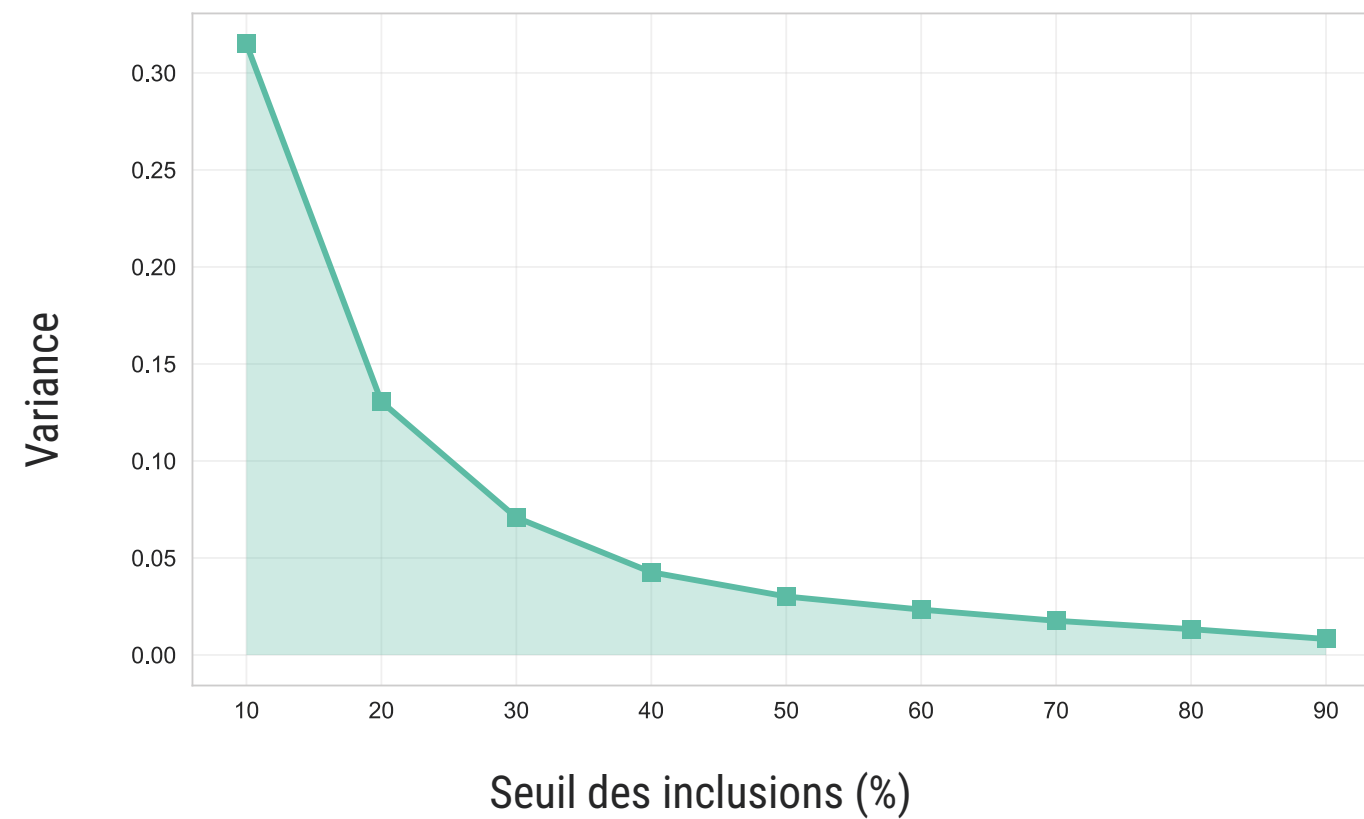
Puissance empirique



RMSE



Variance



- **Biais absolu moyen** : Moyenne des erreurs absolues (0 = absence de biais)
- **Couverture** : % IC contenant la vraie valeur (attendu à 95%)
- **Puissance empirique** : Capacité à détecter un effet significatif
- **RMSE** : Moyenne des erreurs quadratiques (0 = bonnes performances)
- **Variance** : Variance moyenne des estimateurs (plus bas, estimations plus précises)

Globalement, les performances se stabilisent à partir de 50% de recrutement.