

Formalisation de pratiques épidémiologiques pour l'inférence causale

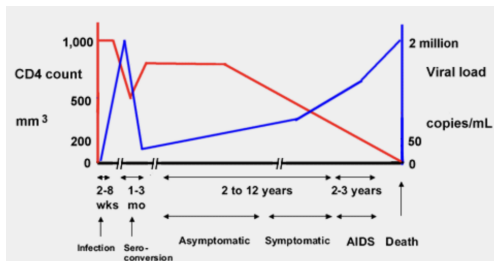
Clément Bérenfeld², Iqraa Meah¹, François Petit³ et Charlotte Voinot²

1. MIA Paris-Saclay
2. Inria Montpellier, équipe PreMeDICaL
3. INSERM, Paris

Statistiques au sommet de Rochebrune – 2026

Question épidémiologique

Meilleur moment pour démarrer traitement contre VIH?



Autres exemples :

- Meilleur moment pour arrêter un traitement
- Effet d'une période + ou - longue d'arrêt un traitement

Source: <https://i-base.info/ttfa/section-2/14-how-cd4-and-viral-load-are-related/>

Règles de traitement dynamiques

Notations :

- $A_t \in \mathcal{A}_t$ (souvent $\mathcal{A}_t = \{0, 1\}$) indique la prise de traitement;
- $X_t \in \mathcal{C}_t$ covariable (souvent $\mathcal{C}_t = \mathcal{C}$);
- $Y_t \in \mathbb{R}$ ou $Y_t \in \{0, 1\}$ outcome d'intérêt;

Règle de traitement:

$$q = \{q_t : \mathcal{X}_t \times \overline{\mathcal{H}}_{t-1} \rightarrow \mathcal{A}_t, t \in \{1, \dots, \tau - 1\}\},$$

avec $\overline{\mathcal{H}}_{t-1} = (X_{t-1}, A_{t-1}, \dots, X_1, A_1, X_0)$,

Objectifs: comparaison de règle $\psi(Y_\tau(q^i)) - \psi(Y_\tau(q^j))$

Exemple d'observations

Ind	Time (months)	CD4 cell count	A_t	Y_t (event indicator)
1	0	550	0	0
1	1	342	1	0
1	2	270	1	0
1	3	200	1	1
2	0	580	0	0
2	1	480	0	0
2	2	450	1	0
2	3	470	1	0
3	0	600	0	0
3	1	510	0	0
3	2	320	0	1

Cadre pour la comparaison de règles de traitement

- “démarrer traitement quand $CD4 \leq 350$ ” = q^1 (delayed start)
- “démarrer traitement quand $CD4 \leq 500$ ” = q^2 (early start)

Cadre idéal : Essais clinique randomisé (données expérimentales)
Limites pratiques liées à l'éthiques, la faisabilité, le coût
⇒ utilisation de données observationnelles

Cadre : émulation d'essais cible [Hernán and Robins, 2016]
Disclaimer : émulation d'essais cible \neq inference causale

Exemple d'observations

Ind	Time (months)	CD4 cell count	A_t	Y_t (event indicator)
1	0	550	0	0
1	1	342	1	0
1	2	270	1	0
1	3	200	1	1
2	0	580	0	0
2	1	480	0	0
2	2	450	1	0
2	3	470	1	0
3	0	600	0	0
3	1	510	0	0
3	2	320	0	1

- Retard d'un mois dans démarrage de traitement pour Ind 2
- Ind 3 mort avant de commencer son traitement ?

Retour sur le contexte épidémiologique

Essais clinique randomisé

Plan d'analyse de données clair et définit en amont

- Période de grâce
- Intention de traitement connue permet de gérer la non adhérence

Données observationnelles :

- Période de grâce non définie
- Intention rarement enregistrée

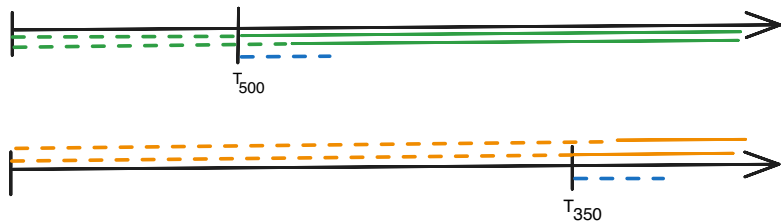
Essais cible émulé

Plan d'analyse posthoc :

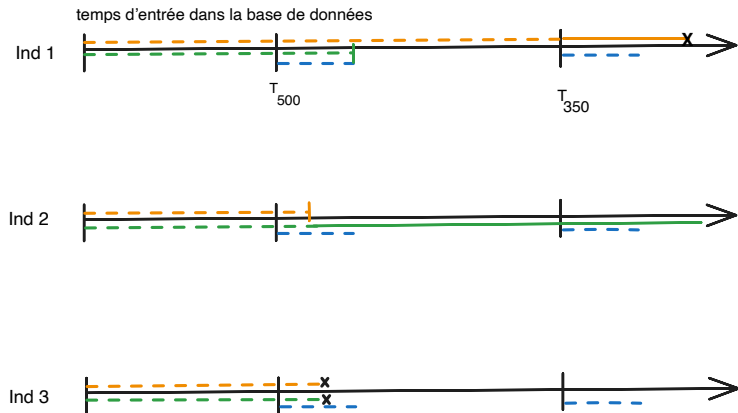
cohérence épidémio vs possibilités avec données

Point de vue en RCT avec période de grâce

temps d'entrée dans l'essai



Clone-censor weight (CCW)

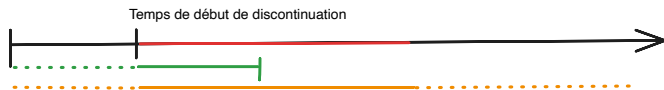


Les croix indiquent des évènements et les marques verticales des censures artificielles.

Bias de sélection corrigé par pondération.

Clone-censor weight (CCW)

[Cain et al., 2010] formalisme pour les périodes de grâce
Motivations : généralisation pour différents cas épidémiologiques
Exemple : [Dumas et al., 2025] comparaison de longueur de
discontinuation de traitement pour le cancer du sein



Les marques verticales indiquent des censures artificielles.

Formalisme

- Règle de traitement stochastique :
 $q = \{q_t : \mathcal{X}_t \times \overline{\mathcal{H}}_{t-1} \rightarrow \mathcal{P}(\mathcal{A}_t), t \in [\tau - 1]\},$
- Espace de trajectoire admissibles $\Omega \subset \overline{\mathcal{H}}$
 $\overline{H}_t \in \Omega_t \Rightarrow \overline{H}_s \in \Omega_s, \forall s < t$
- Espace de noyaux admissibles :
 $\mathcal{Q}(\Omega_s) = \{q_s | q_s(A_s | X_s, \overline{H}_{s-1}) > 0 \Leftrightarrow \overline{H}_s \in \Omega_s\}.$
- Noyau "naturel" : $q_t^{\text{nat}} \in \mathcal{Q}(\Omega_s)$

$$q_t^{\text{nat}}(a | X_t, \overline{H}_{t-1}) := \frac{\mathbb{P}(A_t = a | X_t, \overline{H}_{t-1})}{\sum_{a \in \mathcal{A}_t(X_t, \overline{H}_{t-1})} \mathbb{P}(A_t = a | X_t, \overline{H}_{t-1})} \mathbf{1}_{\{a \in \mathcal{A}_t(X_t, \overline{H}_{t-1})\}}$$

$$\text{avec } \mathcal{A}_t(X_t, \overline{H}_{t-1}) = \{ a \in \mathcal{A}_t \mid (\overline{H}_{t-1}, X_t, a) \in \Omega_t \}$$

Résultats dans le cadre de la survie

$$\hat{S}^{\text{CCW}}(t) := \prod_{s \leq t} \left(1 - \frac{D_s^{\text{CCW}}}{N_s^{\text{CCW}}} \right)$$

avec $N_t^{\text{CCW}} := \sum_{i=1}^n W_{i,t-1} R_{i,t}^{\Omega}$, $D_t^{\text{CCW}} := \sum_{i=1}^n W_{i,t-1} R_{i,t}^{\Omega} \Delta_{i,t}$,

et $W_{i,t}^{\text{CCW}} := \prod_{s=0}^t \frac{1_{\{a \in \mathcal{A}_t(X_t, \bar{H}_{t-1})\}}}{\sum_{a \in \mathcal{A}_t(X_t, \bar{H}_{t-1})} \mathbb{P}(A_t = a | X_t, \bar{H}_{t-1})} \times \frac{1}{p_{s+1}^C(\bar{H}_{i,s})}$

est un estimateur consistant de $S^{q^{\text{nat}}}(t) := \mathbb{P}(Y_t(q^{\text{nat}}) = 0)$,
sous réserve que

$$q_t^{\text{nat}}(a | X_t, \bar{H}_{t-1}) > 0 \Rightarrow \mathbb{P}(A_t = a | X_t, \bar{H}_{t-1}) > 0 \quad \text{et}$$

$$q_t^{\text{nat}}(A_t | X_t, \bar{H}_{t-1}) > 0 \Rightarrow p_{t+1}^C(\bar{H}_t) > 0.$$

En fait, $W_{i,t}^{\text{CCW}} := \prod_{s=0}^t \frac{1_{\{a \in \mathcal{A}_t(X_t, \bar{H}_{t-1})\}}}{\sum_{a \in \mathcal{A}_t(X_t, \bar{H}_{t-1})} \mathbb{P}(A_t = a | X_t, \bar{H}_{t-1})} \times \frac{1}{p_{s+1}^C(\bar{H}_{i,s})} =$

$$\prod_{s=0}^t \frac{q_s^{\text{nat}}(A_{i,s} | X_{i,s}, \bar{H}_{i,s-1})}{\mathbb{P}(A_s = A_{i,s} | X_s, \bar{H}_{i,s-1})} \times \frac{1}{p_{s+1}^C(\bar{H}_{i,s})}$$

Prochaines étapes

Prochaines étapes :

- Résultats de variance avec le bootstrap non-paramétrique
- Résultats sur données simulées
- Essayer différentes définition de Ω sur données réelles ?
↪ analyse de sensibilité

Prochaines étapes

Prochaines étapes :

- Résultats de variance avec le bootstrap non-paramétrique
- Résultats sur données simulées
- Essayer différentes définition de Ω sur données réelles ?
↪ analyse de sensibilité

Merci pour votre attention !





Cain, L. E., Robins, J. M., Lanoy, E., Logan, R., Costagliola, D., and Hernán, M. A. (2010).

When to start treatment? a systematic approach to the comparison of dynamic regimes using observational data.

The international journal of biostatistics, 6(2).



Dumas, E., Jochum, F., Coussy, F., Hamy, A.-S., Majdling, A., Houzard, S., Le Bihan-Benjamin, C., Reyat, F., Gougis, P., and Stensrud, M. J. (2025).

Explaining the relationships between age, endocrine therapy persistence, and risk of recurrence in hormone receptor-positive early breast cancer: A nationwide cohort study.

Journal of Clinical Oncology, 0(0):JCO.24.01131.

PMID: 40043224.



Hernán, M. A. and Robins, J. M. (2016).

Using big data to emulate a target trial when a randomized trial is not available.

American journal of epidemiology, 183(8):758-764.