

Développement d'une règle de traitement individuelle optimale en présence d'observation irrégulière du résultat

Janie Coulombe

Travail conjoint avec Erica E.M. Moodie, Susan M. Shortreed et Christel Renoux

janie.coulombe@umontreal.ca

Département de mathématiques et de statistique,
Université de Montréal

Présenté au Département de mathématiques, Université de Sherbrooke
25 novembre 2022

Motivation

Contexte:

- ▶ Dossiers médicaux électroniques du Royaume-Uni (*Clinical Practice Research Datalink* (CPRD)).
- ▶ Cohorte de nouveaux utilisateurs d'antidépresseurs avec un diagnostic de dépression dans l'année précédente.

Contexte:

- ▶ Dossiers médicaux électroniques du Royaume-Uni (*Clinical Practice Research Datalink* (CPRD)).
- ▶ Cohorte de nouveaux utilisateurs d'antidépresseurs avec un diagnostic de dépression dans l'année précédente.
- ▶ Inhibiteurs sélectifs de la recapture de la sérotonine (ISRS) ciblés: **Citalopram** et **fluoxétine** (31,120 patients).
- ▶ Effets indésirables incluent les changements dans l'appétit et le poids.

Maintenant, supposons que citalopram ait le meilleur effet marginal sur le poids (c'est-à-dire qu'en moyenne, il cause moins de variations de poids détrimementales).

Question:

- Peut-on faire mieux que de toujours prescrire l'antidépresseur qui a le "meilleur" effet causal marginal sur le poids?

- ▶ On souhaite développer une règle de traitement individuelle optimale, de type

“Traiter avec citalopram si l’espérance du résultat ($E[Y(t)|\dots]$) est optimisée sous citalopram, et sinon, avec fluoxétine”

- ▶ Le but: Minimiser les changements de poids “détrimentaux”.
- ▶ On utilise une fonction de l’indice de masse corporelle (IMC) comme résultat.

Comment?

Idée: A partir des données (caractéristiques des patients et autres), on peut en apprendre sur les profils qui réagissent mieux à tel ou tel médicament.

Comment?

Idée: A partir des données (caractéristiques des patients et autres), on peut en apprendre sur les profils qui réagissent mieux à tel ou tel médicament.

(Notez, cela correspond à étudier des effets conditionnels de traitement, plutôt que l'effet marginal).

Comment?

Idée: A partir des données (caractéristiques des patients et autres), on peut en apprendre sur les profils qui réagissent mieux à tel ou tel médicament.

(Notez, cela correspond à étudier des effets conditionnels de traitement, plutôt que l'effet marginal).

On voudra estimer ce qu'on appelle la **modification d'effet** du traitement par les caractéristiques des patients.

Comment?

Idée: A partir des données (caractéristiques des patients et autres), on peut en apprendre sur les profils qui réagissent mieux à tel ou tel médicament.

(Notez, cela correspond à étudier des effets conditionnels de traitement, plutôt que l'effet marginal).

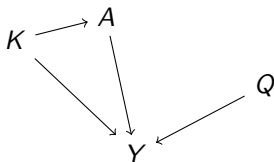
On voudra estimer ce qu'on appelle la **modification d'effet** du traitement par les caractéristiques des patients.

Ex.:

- ▶ Au-dessus de 40 ans, citalopram devient préférable à fluoxétine
- ▶ Les patients qui ont déjà reçu un diagnostic d'anxiété généralisée ont moins de variations de poids avec fluoxétine
- ▶ Etc.

Modification d'effet (=interaction avec le traitement)

Supposons qu'on s'intéresse à l'effet causal de A sur Y et qu'on a:



Certains modèles possibles pour Y :

$$\mathbb{E}[Y|A, K, Q] = \beta_0 + \beta_a A + \beta_k K + \beta_q Q, \quad (\text{A})$$

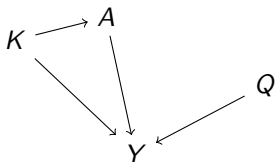
$$\mathbb{E}[Y|A, K, Q] = \beta_0 + \beta_a A + \beta_k K + \beta_q Q + \beta_{int} A \times K, \quad (\text{B})$$

$$\mathbb{E}[Y|A, K, Q] = \beta_0 + \beta_a A + \beta_k K + \beta_q Q + \beta_{int} A \times Q, \quad (\text{C})$$

$$\mathbb{E}[Y|A, K, Q] = \beta_0 + \beta_a A + \beta_k K + \beta_q Q + \beta_{int1} A \times K + \beta_{int2} A \times Q, \quad (\text{D})$$

Modification d'effet (=interaction avec le traitement)

Supposons qu'on s'intéresse à l'effet causal de A sur Y et qu'on a:



Certains modèles possibles pour Y :

$$\mathbb{E}[Y|A, K, Q] = \beta_0 + \beta_a A + \beta_k K + \beta_q Q, \quad (\text{A})$$

$$\mathbb{E}[Y|A, K, Q] = \beta_0 + \beta_a A + \beta_k K + \beta_q Q + \beta_{int} A \times K, \quad (\text{B})$$

$$\mathbb{E}[Y|A, K, Q] = \beta_0 + \beta_a A + \beta_k K + \beta_q Q + \beta_{int} A \times Q, \quad (\text{C})$$

$$\mathbb{E}[Y|A, K, Q] = \beta_0 + \beta_a A + \beta_k K + \beta_q Q + \beta_{int1} A \times K + \beta_{int2} A \times Q, \quad (\text{D})$$

Modèles (B), (C) et (D) permettent à l'effet de A sur Y de varier selon K , ou Q , ou les deux. C.-à-d. que

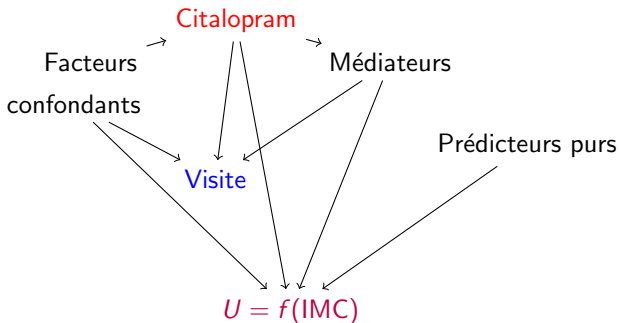
$$\mathbb{E}[Y|A = 1, K, Q] - \mathbb{E}[Y|A = 0, K, Q]$$

dépend des valeurs de K et Q .

Notre règle dépendra donc des effets conditionnels de traitement (qui doivent être estimés de façon convergente).

Défis particuliers dans notre contexte: L'IMC est mesuré à **répétition** et de façon **irrégulière**, et il y a de la confusion.

Le diagramme causal postulé à chaque temps t :



Avec $Visite(t) = 1$ correspondant à l'observation de $U(t) = f(IMC)$.

L'indicateur de visite agit comme facteur de collision. Or, la condition sur un facteur de collision crée des associations trompeuses entre ses ancêtres (Greenland, 2003).

Méthodes

Notation (1)

On s'intéresse à un échantillon aléatoire de patients indexés par $i = 1, \dots, n$.

Soient:

- ▶ $A_i(t)$ et $Y_i(t)$ le traitement binaire et le résultat continu de l'individu i au temps t
- ▶ Les facteurs confondants dénotés par $\mathbf{K}_i(\mathbf{t})$
- ▶ Les prédicteurs purs de $Y_i(t)$, dénotés par $\mathbf{Q}_i(\mathbf{t})$
- ▶ Les prédicteurs de $Y_i(t)$ sans le traitement, $\mathbf{X}_i^\beta(\mathbf{t}) = [\mathbf{1} \quad \mathbf{K}_i(\mathbf{t}) \quad \mathbf{Q}_i(\mathbf{t})]$
- ▶ Les modificateurs d'effet $\mathbf{X}_i^\psi(\mathbf{t})$
- ▶ Soit $\mathbf{X}(\mathbf{t}) = [\mathbf{X}^\beta(\mathbf{t}) \quad \mathbf{X}^\psi(\mathbf{t})]$

Notation (2)

Pour modéliser les temps d'observation de $Y_i(t)$:

- ▶ $dN_i(t)$ l'indicateur de visite et d'observation de $Y_i(t)$
- ▶ Par définition, $N_i(t) = \int_{s=0}^t \sum_{j=1}^{F_i} \mathbb{I}(s = T_{ij})$ pour T_{ij} les temps de visite du patient i
- ▶ C_i le temps passé dans l'étude, pour le patient i
- ▶ $\xi_i(t) = \mathbb{I}(C_i \geq t)$ l'indicateur d'être encore dans l'étude au temps t
- ▶ On assume que $dN_i(t) \perp Y_i(t) | \mathbf{V}_i(t)$ pour $\mathbf{V}_i(t)$ un ensemble de variables affectant les visites.

Voyons d'abord le cas plus simple (règle qui n'est pas doublement robuste).

Postulons:

$$\mathbb{E}[Y_i(t)|A_i(t), \mathbf{X}_i(t)] = g(\mathbf{X}_i^\psi(t)) + A_i(t)\psi' \mathbf{X}_i^\psi(t).$$

C'est le **blip** qui nous intéresse (comment l'effet de traitement varie dans diverses strates de la population) et non la fonction $g(\cdot)$. On se concentre sur l'estimation des coefficients du blip.

Sous le modèle postulé pour le résultat, la règle suivante:

**“Traiter avec citalopram si l'espérance de l'issue
est maximisée sous citalopram, et sinon, avec fluoxétine”**

devient alors

“Traiter avec citalopram si $\hat{\psi}'\mathbf{X}_i^{\psi}(t) \geq 0$, et sinon, avec fluoxétine.”

Si on peut estimer ψ de façon convergente (et robuste? A venir plus tard...) alors on a une règle avec de bonnes propriétés.

Hypothèses causales

- Soient $Y_{i0}(t)$ et $Y_{i1}(t)$ deux issues potentielles sous les interventions 0 et 1 (Neyman, 1923; Rubin, 1974)
- Paramètre d'intérêt: effet conditionnel de traitement

$$\text{ECT} = \mathbb{E}[Y_{i1}(t) - Y_{i0}(t) | \mathbf{X}_i^\psi(t)]$$

- Données (si $\mathbf{X}^\psi(t) = X^\psi(t)$):

ID	$dN_i(t)$	$A_i(t)$	$X_i^\psi(t)$ (groupe âge)	$Y_i(t)$	$Y_{i0}(t)$	$Y_{i1}(t)$
1	1	1	18-25	25	-	25
2	1	0	26-35	22	22	-
...						
$n-1$	1	0	18-25	23	23	-
n	1	1	26-35	15	-	15

Hypothèses causales

À partir des **hypothèses causales** suivantes, on peut estimer des fonctions de $Y_{i0}(t)$ et $Y_{i1}(t)$:

- Interchangeabilité conditionnelle

$$A_i(t) \perp \{Y_{i0}(t), Y_{i1}(t)\} \mid \mathbf{K}_i(\mathbf{t}), \mathbf{V}_i(\mathbf{t}), dN_i(t)$$

- Positivité de l'intervention et de l'observation

$$0 < \mathbb{P}(A_i(t) \mid \mathbf{K}_i(\mathbf{t}), \mathbf{X}_i^\psi) < 1$$

$$0 < \mathbb{P}(dN_i(t) \mid \mathbf{V}_i(\mathbf{t})) < 1$$

- Consistence de l'issue

$$Y_i(t) = A_i(t)Y_{i1}(t) + (1 - A_i(t))Y_{i0}(t)$$

Pourquoi?

Estimateur doublement pondéré

On utilise le concept de “pseudo-population”:

$$\begin{aligned} & \mathbb{E}[Y_{i1}(t) - Y_{i0}(t) | \mathbf{X}_i^\psi(\mathbf{t})] \\ &= \mathbb{E}_{\mathbf{V}}[\mathbb{E}_{\mathbf{K}}[\mathbb{E}[Y_{i1}(t) | \mathbf{K}_i(\mathbf{t}), \mathbf{V}_i(\mathbf{t}), \mathbf{X}_i^\psi(\mathbf{t})]]] - \mathbb{E}_{\mathbf{V}}[\mathbb{E}_{\mathbf{K}}[\mathbb{E}[Y_{i0}(t) | \mathbf{K}_i(\mathbf{t}), \mathbf{V}_i(\mathbf{t}), \mathbf{X}_i^\psi(\mathbf{t})]]] \\ &= \mathbb{E}_{\mathbf{V}}[\mathbb{E}_{\mathbf{K}}[\mathbb{E}[\textcolor{blue}{Y}_{i1}(\textcolor{blue}{t}) | \mathbf{K}_i(\mathbf{t}), \textcolor{blue}{A}_i(\textcolor{blue}{t}) = 1, \mathbf{V}_i(\mathbf{t}), \textcolor{red}{dN}_i(\textcolor{red}{t}) = 1, \mathbf{X}_i^\psi(\mathbf{t})]]] \\ &\quad - \mathbb{E}_{\mathbf{V}}[\mathbb{E}_{\mathbf{K}}[\mathbb{E}[\textcolor{red}{Y}_{i0}(\textcolor{red}{t}) | \mathbf{K}_i(\mathbf{t}), \textcolor{red}{A}_i(\textcolor{red}{t}) = 0, \mathbf{V}_i(\mathbf{t}), \textcolor{red}{dN}_i(\textcolor{red}{t}) = 1, \mathbf{X}_i^\psi(\mathbf{t})]]] \end{aligned}$$

par interchangeabilité conditionnelle et positivité

$$\begin{aligned} &= \mathbb{E}_{\mathbf{V}}[\mathbb{E}_{\mathbf{K}}[\mathbb{E}[\textcolor{blue}{Y}_i(\textcolor{blue}{t}) | \mathbf{K}_i(\mathbf{t}), A_i(t) = 1, \mathbf{V}_i(\mathbf{t}), dN_i(t) = 1, \mathbf{X}_i^\psi(\mathbf{t})]]] \\ &\quad - \mathbb{E}_{\mathbf{V}}[\mathbb{E}_{\mathbf{K}}[\mathbb{E}[\textcolor{red}{Y}_i(\textcolor{red}{t}) | \mathbf{K}_i(\mathbf{t}), A_i(t) = 0, \mathbf{V}_i(\mathbf{t}), dN_i(t) = 1, \mathbf{X}_i^\psi(\mathbf{t})]]] \end{aligned}$$

par consistance.

Estimateur doublement pondéré

On peut ensuite montrer pour chaque partie de l'expression:

$$\begin{aligned} & \mathbb{E}_{\mathbf{V}}[\mathbb{E}_{\mathbf{K}}[\mathbb{E}[Y_i(t)|\mathbf{K}_i(\mathbf{t}), A_i(t) = 1, \mathbf{V}_i(\mathbf{t}), dN_i(t) = 1, \mathbf{X}_i^\psi(\mathbf{t})]]] \\ &= \sum_{\mathbf{V}} \sum_{\mathbf{K}} \frac{\mathbb{P}(A_i(t) = 1|\mathbf{K}_i(\mathbf{t}))\mathbb{P}(dN_i(t) = 1|\mathbf{V}_i(\mathbf{t}))}{\mathbb{P}(A_i(t) = 1|\mathbf{K}_i(\mathbf{t}))\mathbb{P}(dN_i(t) = 1|\mathbf{V}_i(\mathbf{t}))} \\ & \quad \times \mathbb{E}[Y_i(t)|\mathbf{K}_i(\mathbf{t}), A_i(t) = 1, \mathbf{V}_i(\mathbf{t}), dN_i(t) = 1, \mathbf{X}_i^\psi(\mathbf{t})]\mathbb{P}(\mathbf{K} = \mathbf{K}_i(\mathbf{t}))\mathbb{P}(\mathbf{V} = \mathbf{V}_i(\mathbf{t})) \\ &= \dots \\ &= \mathbb{E} \left[\frac{\mathbb{I}(A_i(t) = 1, dN_i(t) = 1) Y_i(t)}{\mathbb{P}(A_i(t) = 1|\mathbf{K}_i(\mathbf{t}))\mathbb{P}(dN_i(t) = 1|\mathbf{V}_i(\mathbf{t}))} | \mathbf{X}_i^\psi(\mathbf{t}) \right], \end{aligned}$$

et similairement pour $A_i(t) = 0$, menant à

$$\mathbb{E} \left[\frac{\mathbb{I}(A_i(t) = 0, dN_i(t) = 1) Y_i(t)}{\mathbb{P}(A_i(t) = 0|\mathbf{K}_i(\mathbf{t}))\mathbb{P}(dN_i(t) = 1|\mathbf{V}_i(\mathbf{t}))} | \mathbf{X}_i^\psi(\mathbf{t}) \right]$$

Estimateur doublement pondéré

- ▶ On modélise les temps d'observation à partir du modèle à taux proportionnel:

$$\mathbb{E}[dN_i(t)|\mathbf{V}_i(\mathbf{t})] = \xi_i(t) \exp \{ \gamma' \mathbf{V}_i(\mathbf{t}) \} \lambda_0(t) dt$$

avec γ estimé avec Andersen et Gill (1982). Le modèle ci-haut sert à calculer un poids inverse à l'intensité de visite (IIV) dénoté par $\varphi_i(\gamma, \mathbf{V}_i(\mathbf{t}))$ (Lin et al., 2004).

- ▶ Ce poids est combiné à un poids inverse à la probabilité de traitement (IPT):

$$e_i(\omega, \mathbf{K}_i(\mathbf{t})) = \frac{\mathbb{I}(A_i(t) = 1)}{\mathbb{P}(A_i(t) = 1 | \mathbf{K}_i(\mathbf{t}); \omega)} + \frac{\mathbb{I}(A_i(t) = 0)}{\mathbb{P}(A_i(t) = 0 | \mathbf{K}_i(\mathbf{t}); \omega)}.$$

Pour construire une règle doublement robuste, on incorpore maintenant un modèle “**sans traitement**” (*treatment-free*) en plus du blip $\psi' \mathbf{X}_i^\psi(t)$:

$$\mathbb{E}[Y_i(t)|A_i(t), \mathbf{X}_i(t)] = f \left\{ \mathbf{X}_i^\beta(t); \beta \right\} + A_i(t) \psi' \mathbf{X}_i^\psi(t).$$

- La fonction $f \left\{ \mathbf{X}_i^\beta(t); \beta \right\}$: Un premier essai pour ajuster pour la confusion (on ajoutera en plus les poids IPT)

Doublement robuste - intuition

- Soit on trouve un poids (inverse) relié au mécanisme de prescription, qui rend $A_i(t)$ et $\mathbf{X}_i^\beta(t)$ indépendants lorsque le poids est bien spécifié

Doublement robuste - intuition

- ▶ Soit on trouve un poids (inverse) relié au mécanisme de prescription, qui rend $A_i(t)$ et $\mathbf{X}_i^\beta(t)$ indépendants lorsque le poids est bien spécifié
- ▶ Soit $f\left\{\mathbf{X}_i^\beta(t); \beta\right\}$ est bien spécifié et on a l'indépendance entre $Y_i(t)$ et $\mathbf{X}_i^\beta(t)$ dans

$$\mathbb{E}[Y_i(t)|A_i(t), \mathbf{X}_i(t)] = \underbrace{f\left\{\mathbf{X}_i^\beta(t); \beta\right\}}_{\text{contient les facteurs confondants}} + A_i(t)\psi'\mathbf{X}_i^\psi(t)$$

de sorte que le blip (en bleu) peut être estimé.

Doublement robuste - intuition

- ▶ Soit on trouve un poids (inverse) relié au mécanisme de prescription, qui rend $A_i(t)$ et $\mathbf{X}_i^\beta(t)$ indépendants lorsque le poids est bien spécifié
- ▶ Soit $f\{\mathbf{X}_i^\beta(t); \beta\}$ est bien spécifié et on a l'indépendance entre $Y_i(t)$ et $\mathbf{X}_i^\beta(t)$ dans

$$\mathbb{E}[Y_i(t)|A_i(t), \mathbf{X}_i(t)] = \underbrace{f\{\mathbf{X}_i^\beta(t); \beta\}}_{\text{contient les facteurs confondants}} + A_i(t)\psi'\mathbf{X}_i^\psi(t)$$

de sorte que le blip (en bleu) peut être estimé.

Dans les deux cas, les coefficients ψ peuvent être estimés de façon convergente même si l'un des deux modèles (du poids ou du modèle sans traitement) n'est pas correct.

Estimateur doublement pondéré

- L'équation d'estimation à résoudre (extension des dWOLS proposés par Wallace et Moodie (2015)):

$$U(\beta, \psi; \hat{\gamma}, \hat{\omega}) = \sum_{i=1}^n \int_0^{\tau} \varphi_i(\hat{\gamma}, \mathbf{V}_i(\mathbf{t})) e_i(\hat{\omega}, \mathbf{K}_i(\mathbf{t})) \\ \times \left[\frac{\frac{\partial f\{\mathbf{X}_i^{\beta}(t); \beta\}}{\partial \beta}}{A_i(t) \mathbf{X}_i^{\psi}(t)} \right] \left[Y_i(t) - f\{\mathbf{X}_i^{\beta}(t); \beta\} - A_i(t) \psi' \mathbf{X}_i^{\psi}(t) \right] dN_i(t) = \mathbf{0}.$$

- La variance asymptotique peut être calculée (Newey et McFadden, 1994):

$$\sqrt{n}(\hat{\psi}_{\text{DW}} - \psi_0) \xrightarrow{d} N(\mathbf{0}, \Sigma_{\mathbf{A}}),$$

pour $\Sigma_{\mathbf{A}}$ un bloc dans

$$\Sigma = \mathbf{G}_{\beta, \psi}^{-1} \mathbb{E} \left[\left\{ \mathbf{g}(\mathbf{o}; \beta_0, \psi_0, \omega_0, \gamma_0) - \mathbf{G}_{\omega, \gamma} \mathbf{M}^{-1} \mathbf{m}(\mathbf{o}; \omega_0, \gamma_0) \right\}^{\otimes 2} \right] \mathbf{G}_{\beta, \psi}^{-1}.$$

Autres hypothèses

Note 1: On fait l'hypothèse que l'effet de traitement est aigu (pas de délai, pas d'effets rémanents) (Dong et al., 2021):

Note 2: Tel que décrit dans Wallace et Moodie (2015), le fait

1. d'utiliser un poids qui satisfait la “*balancing condition*” (comme IPT ou les poids de chevauchement) et
2. de bien spécifier le blip $\psi' \mathbf{X}_i^\psi(t)$,

en plus de satisfaire les autres hypothèses mentionnées plus tôt, mène à un estimateur doublement robuste. Dans notre cas, on ajoute un poids IIV qui doit absolument être correctement spécifié.

Étude de simulation

$$\mathbf{K}_i = \{K_{1i}, K_{2i}, K_{3i}\} \sim \{N(1, 1), \text{Bern}(0, 55), N(0, 1)\}$$

$$A_i \sim \text{Bern}(p_i) \text{ avec } p_i = \text{expit} \{0,5 + 0,55K_{1i} - 0,2K_{2i} - 1K_{3i}\}$$

$$Z_i(t) \sim \begin{cases} N(2, 1) & \text{si } A_i = 1 \\ N(4, 2) & \text{si } A_i = 0 \end{cases}$$

$$Q_i(t) \sim \text{Bern}(0.5)$$

$$Y_i(t) = \sqrt{t/100} - 2 A_i + 2.5 \{Z_i(t) - \mathbb{E}[Z_i(t)|A_i]\} + \\ 0,4K_{1i} + 0,05K_{2i} - 0,6K_{3i} + 0.5 \{A_i \times Q_i(t)\} - 1 \{A_i \times K_{1i}\} + \epsilon_i(t)$$

$$\text{où } \epsilon_i(t)|\phi_i \sim N(\phi_i, 0,01), \phi_i \sim N(0, 0,04)$$

$$dN_i(t) \sim \text{Poisson}(\lambda_i(t)) \text{ avec } \lambda_i(t) = \eta_i \exp \{\gamma_A A_i + \gamma_Z Z_i(t) + \gamma_{K2} K_2 + \gamma_{K3} K_3\} \\ \text{où } \eta_i \text{ un effet aléatoire, } (\gamma_A, \gamma_Z, \gamma_{K2}, \gamma_{K3}) \text{ les param. de dépendance}$$

Étude de simulation

Les visites simulées selon le **modèle à taux proportionnel**:

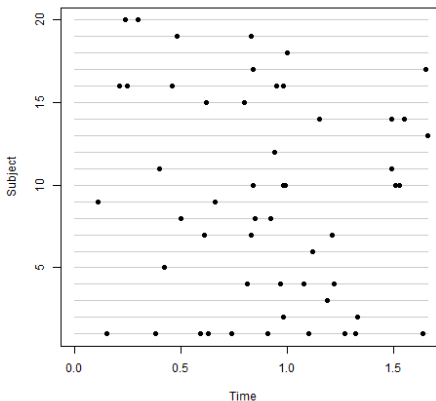


Diagramme abacus produit à partir du progiciel IrregLong (Pullenayegum, 2022)

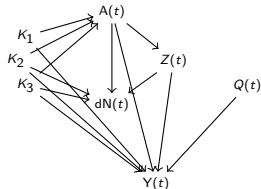
Estimateurs comparés

Pour évaluer la robustesse de la méthode, on compare 6 estimateurs:

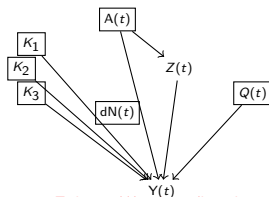
Estimateur	Modèle mal spécifié (x)			
	Visite (partiel ¹)	Visite (complet ²)	Traitement ³	Issue ⁴
$\hat{\psi}_{DW1}$				
$\hat{\psi}_{DW2}$	x			x
$\hat{\psi}_{DW3}$	x		x	
$\hat{\psi}_{DW4}$		x		
$\hat{\psi}_{OLS}$		x	x	
$\hat{\psi}_{IPT}$		x		

1. Ajusté pour les variables importantes (traitement et médiateur)
2. Non ajusté pour le médiateur, ajusté pour le traitement et K_2
3. Mauvais format (au carré) pour termes linéaires de K_1 et K_3
4. Manque l'ajustement pour K_2

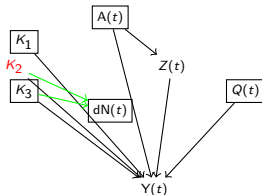
Diagrammes causaux correspondants



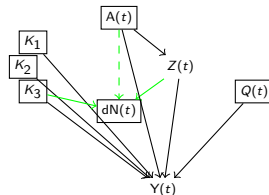
(a) Mécanisme de génération des données



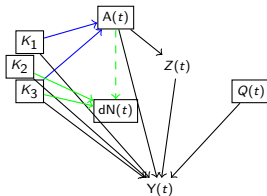
(b) Estimateur proposé



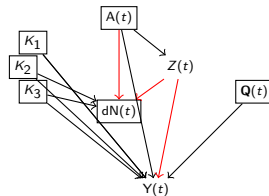
(c) DW2: Modèles observation et résultat incorrects



(e) DW4: Modèle observation incorrect ($dN_i(t) \sim A(t) + K_2$)



(d) DW3: Modèle observation et modèle traitement incorrects



(f) Estimateur IPT (biais stratification sur collision)

Note: Une boîte signifie qu'on conditionne sur la variable.

Résultats

Tableau 1: Résultats de l'étude de simulation ($M = 1000$ simulations) - taux d'erreur de la décision optimale pour le traitement

Taille éch.	Paramètres γ^v	No. temps obs. moyenne (EIQ)	Taux erreur					
			$\hat{\psi}_{DW1}$	$\hat{\psi}_{DW2}$	$\hat{\psi}_{DW3}$	$\hat{\psi}_{DW4}$	$\hat{\psi}_{OLS}$	$\hat{\psi}_{IPT}$
250	1	3 (1-3)	0.02	0.01	0.01	0.04	0.03	0.04
	2	3 (2-5)	0.05	0.06	0.05	0.16	0.15	0.16
	3	6 (3-9)	0.06	0.03	0.03	0.26	0.25	0.26
	4	10 (8-12)	0.01	0.01	0.00	0.01	0.00	0.01
500	1	3 (1-3)	0.01	0.01	0.01	0.03	0.03	0.03
	2	3 (1-5)	0.02	0.03	0.02	0.14	0.13	0.14
	3	6 (3-9)	0.04	0.02	0.02	0.25	0.25	0.25
	4	10 (8-12)	0.00	0.00	0.00	0.00	0.00	0.00

$v.1.$ (-2, -0.3, 0.2, -1.2); 2. (0.3, -0.6, -0.4, -0.3); 3. (0.4, -0.8, 1, 0.6); 4. (0, 0, 0, 0). Abbrév.: EQM, erreur quadratique moyenne; EIQ, étendue interquartile.

Application

Question

- ▶ Règle de traitement pour choisir entre citalopram (1) et fluoxétine (0) pour réduire les variations détrimmentales de l'IMC
- ▶ Définition de l'issue:

$$U(t) = 100 - [|IMC(t) - 22| - |IMC(0) - 22|] / IMC(0) \times 100$$

(enlève à 100 l'éloignement relatif au poids normal par rapport à l'entrée dans la cohorte - si l'IMC varie d'une façon détrimmentale, l'utilité diminue vers un minimum dans la cohorte).

- ▶ 31,120 patients et 48,388 mesures pour $U(t)$

Caractéristiques des patients à l'entrée dans la cohorte

Tableau 2: Caractéristiques des 31,120 patients, n (%), CPRD, 1998-2017

Variable	Traitement	
	Citalopram (n=18, 671)	Fluoxétine (n=12, 449)
Âge, moyenne (ES)	48.5 (18.1)	45.1 (16.5)
Sexe masculin	5965 (32)	3609 (29)
Indice de déprivation multiple, moyenne (ES)	3.0 (1.4)	3.1 (1.4)
Année d'entrée dans la cohorte		
1998-2005	3751 (20)	4896 (39)
2006-2011	10,279 (55)	5703 (46)
2012-2017	4641 (25)	1850 (15)
A été fumeur	11,586 (62)	8017 (64)
Abus d'alcool	1478 (8)	869 (7)
Maladie psychiatrique [†]	521 (3)	321 (3)
Anxiété	5956 (32)	2987 (24)
Médication		
Antipsychotiques	2836 (15)	1675 (13)
Autres psychotropiques [‡]	4476 (24)	2546 (20)
Hypolipidémiants	3360 (18)	1614 (13)
Nb. d'hospitalisations psychiatriques dans 6 mois précédents, moyenne (ES)	0.04 (0.24)	0.03 (0.34)

Abbrev.: CPRD, Clinical Practice Research Datalink; ES, erreur standard.

†. Indicateur d'un diagnostic d'autisme, OCD, trouble bipolaire ou schizophrénie.

‡. Inclut les benzodiazépines, anxiolytiques, barbituriques et hypnotiques.

Ratios de taux de visites

Tableau 3: Ratios de taux estimés (ICs 95% bootstrap) pour l'observation de l'issue $U(t)$, CPRD, 1998-2017

Variable	Ratio de taux (CI 95%)
Traitement au citalopram	0.9 (0.9, 0.9)*
Âge	1.0 (1.0, 1.0)
Sexe masculin	0.9 (0.9, 0.9)*
Indice de déprivation multiple	1.0 (1.0, 1.0)*
Année d'entrée dans la cohorte (réf.= <2006)	
2006-2011	0.9 (0.9, 1.0)*
2012-2017	0.9 (0.9, 0.9)*
A été fumeur	1.7 (1.6, 1.7)*
Abus d'alcool	1.0 (0.9, 1.1)
Maladie psychiatrique [†]	1.0 (0.9, 1.2)
Anxiété	1.0 (1.0, 1.0)
Médication	
Antipsychotiques	1.1 (1.0, 1.2)*
Autres psychotropiques [‡]	1.2 (1.2, 1.3)*
Hypolipidémiants	1.2 (1.2, 1.3)*
Nb. d'hospitalisations psychiatriques 6 mois précédents	1.0 (0.9, 1.0)

Abbrev.: IC, intervalle de confiance; CPRD, Clinical Practice Research Datalink.

*. IC significatif.

†. Indicateur d'un diagnostic d'autisme, OCD, trouble bipolaire ou schizophrénie.

‡. Inclut les benzodiazépines, anxiolytiques, barbituriques et hypnotiques.

Règle de traitement

Traiter avec citalopram si $-1.26 + 0.01 \times [\text{Age}] - 0.02 \times \mathbb{I}[\text{Sexe masculin}]$
 $+0.06 \times [\text{Indice déprivation multiple}] - 0.04 \times \mathbb{I}[\text{A été fumeur}]$
 $+0.46 \times \mathbb{I}[\text{Abus d'alcool}] + 1.36 \times \mathbb{I}[\text{Maladie psychiatrique}]$
 $+0.46 \times \mathbb{I}[\text{Anxiété}] - 0.66 \times \mathbb{I}[\text{Utilisation d'antipsychotiques}]$
 $+0.24 \times \mathbb{I}[\text{Autres psychotropiques}] + 0.06 \times \mathbb{I}[\text{Hypolipidémiants}] > 0$

Tableau 4: Comparaison des issues ajustées, CPRD, 1998-2017

Traitement	Issue ajustée moyenne (ES^\dagger)			
	$\hat{\psi}_{OLS}$	$\hat{\psi}_{IPT}$	$\hat{\psi}_{HIV}$	$\hat{\psi}_{DW1}$
Reçu	99.11 (0.001)	99.09 (0.001)	99.08 (0.001)	99.06 (0.001)
Optimal	99.11 (0.001)	99.12 (0.001)	99.08 (0.001)	99.11 (0.001)

Abbrev.: ES, erreur standard.

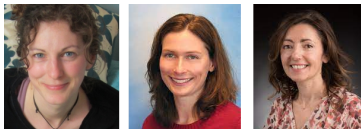
†. Basée sur les ES sommées et normalisées pour la variance de la moyenne des issues.

Discussion

Discussion

- ▶ Les règles de traitement individuelles permettent d'optimiser un choix de traitement selon les caractéristiques du patient
- ▶ Dans une analyse typique, on conditionne sur l'indicateur d'observation des variables. On doit étudier le diagramme causal!
- ▶ Les méthodes discutées s'appuient sur des hypothèses
 - causales
 - sur le mécanisme de génération des données
 - fenêtre temporelle pour l'effet de traitement
 - sur les modèles
- ▶ Extension à venir: Règle à plusieurs étapes et optimisation sur plusieurs points

Remerciements



Mmes Erica E. M. Moodie,
Susan M. Shortreed
et Christel Renoux

ainsi que Mr Éric Marchand pour
l'invitation.

Merci de votre attention!

Financement et support:



Contact:

janie.coulombe@umontreal.ca

Références

- Andersen, P. K., et Gill, R. D. (1982) Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10(4), pp. 1100-1120.
- Bian, Z., Moodie, E. E. M., Shortreed, S. M., et Bhatnagar, S. (2021) Variable selection in regression-based estimation of dynamic treatment regimes. *Biometrics*, à venir.
- Coulombe, J., Moodie, E. E. M., et Platt, R. W. (2021) Weighted regression analysis to correct for informative monitoring times and confounders in longitudinal studies. *Biometrics*, 77(1), pp. 162-174.
- Coulombe, J., Moodie, E. E. M., Shortreed, S. M., et Renoux, C. (2022) Estimating Individualized Treatment Rules in Longitudinal Studies with Covariate-Driven Observation Times. *arXiv:2202.09611v1*, pp. 1- 62.
- Dong, L., Moodie, E. E. M., Villain, L., et Thiébaud, R. Evaluating the use of generalized dynamic weighted ordinary least squares for individualized HIV treatment strategies. *arXiv:2109.01218v1*, pp. 1-38.
- Greenland, S. (2003) Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, 14(3), pp. 300-306.
- Lin, H., Scharfstein, D. O., et Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3), pp. 791-813.
- Newey, W. K., et McFadden, D. (1994) Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4, pp. 2111-2245.
- Neyman, J. S. (1923) On the application of probability theory to agricultural experiments. Essay on principles, section 9, *Statistical Science*, 5(14), pp. 465-472.
- Pullenayegum, E. Package 'IrregLong'. 2022.
- Wallace, M., et Moodie, E. E. M. (2015) Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71(3), pp. 636-644.
- Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), pp. 688-701.

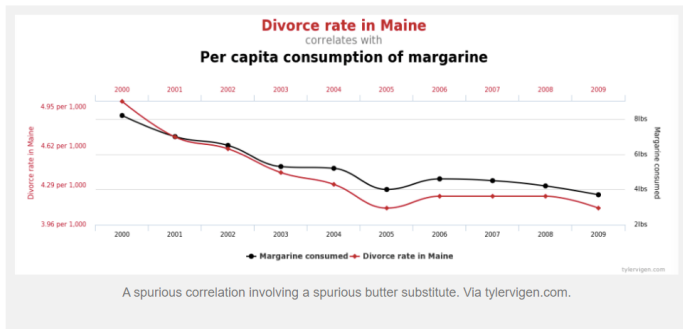
Blip évalué chez différents patients

Male sex (yes)	IMD (1 to 5)	Ever smoker (yes)	Alcohol abuse (yes)	Psychiatric diagnosis (yes)	Anxiety (yes)	Antipsy. drug (yes)	Psychotro. drug (yes)	Lipid lowering drug (yes)	Value blip function
0	1	0	0	0	0	1	0	0	-2.23
1	1	0	0	0	0	1	0	0	-2.07
0	3	0	0	0	0	1	1	1	-1.46
0	1	0	0	0	0	0	0	0	-1.32
1	3	0	0	0	0	1	1	1	-1.30
1	1	0	0	0	0	0	0	0	-1.16
0	3	1	0	0	0	0	0	0	-0.98
1	3	1	0	0	0	0	0	0	-0.82
0	5	0	0	0	0	0	0	0	-0.80
0	3	0	0	0	1	0	0	0	-0.71
1	5	0	0	0	0	0	0	0	-0.64
0	3	1	1	0	0	0	0	0	-0.56
1	3	0	0	0	1	0	0	0	-0.55
1	3	1	1	0	0	0	0	0	-0.40
0	5	1	1	0	0	0	0	0	-0.30
1	5	1	1	0	0	0	0	0	-0.14
0	3	0	1	1	0	0	0	0	0.67
0	3	1	1	1	0	0	0	0	0.75
1	3	0	1	1	0	0	0	0	0.83
0	3	1	0	1	1	0	0	1	0.89
1	3	1	1	1	0	0	0	0	0.91
0	3	0	1	1	1	0	0	0	1.02
0	5	1	1	1	0	0	0	0	1.01
1	3	1	0	1	1	0	0	1	1.05
1	5	1	1	1	0	0	0	0	1.17
1	3	0	1	1	1	0	0	0	1.18
1	5	1	1	1	1	0	1	1	2.03

Rappel: Facteur confondant

Divorce And Margarine

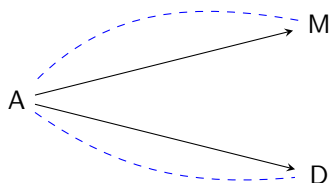
Posted on [April 10, 2017](#) by [annahaensch](#)



Source: [blogs.ams.org](#)

Facteur confondant A: Intérêt dans la relation M-D

Sans ajustement: M et D dépendants



Si modélisation de $M|A$ et ajustement (p. ex., poids IPT):

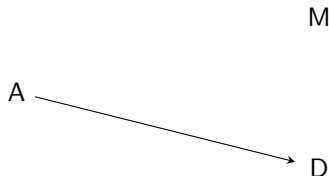


Tableau 5: Coefficients of the blip function (95% bootstrap CIs) for the optimal treatment rules as estimated by four alternative models: OLS which does not adjust for confounding or observation process, IPW which accounts only for confounding, IIV which accounts only for the observation process, and the proposed doubly-weighted estimator which accounts for both processes, CPRD, UK, 1998-2017, n=31,120 individuals.

Variable	$\hat{\psi}_{OLS}$	$\hat{\psi}_{IPT}$	$\hat{\psi}_{IIV}$	$\hat{\psi}_{DW}$
Intercept	-1.66 (-2.69, -0.46)	-1.38 (-2.62, -0.11)	-1.68 (-2.84, -0.58)	-1.45 (-2.66, -0.22)
Age	0.01 (-0.01, 0.03)	0.00 (-0.01, 0.03)	0.01 (-0.01, 0.03)	0.00 (-0.02, 0.03)
Male sex	-0.08 (-0.67, 0.55)	0.03 (-0.59, 0.62)	0.03 (-0.54, 0.65)	0.16 (-0.48, 0.76)
IMD	0.14 (-0.09, 0.31)	0.14 (-0.13, 0.32)	0.12 (-0.10, 0.31)	0.13 (-0.13, 0.31)
Ever smoker	0.23 (-0.38, 0.66)	0.13 (-0.47, 0.65)	0.21 (-0.41, 0.66)	0.08 (-0.50, 0.60)
Alcohol abuse	1.03 (-0.11, 2.24)	0.68 (-0.48, 1.88)	0.78 (-0.26, 1.99)	0.42 (-0.70, 1.60)
Psychiatric disease [†]	0.44 (-1.83, 2.10)	1.02 (-1.23, 2.84)	0.58 (-1.73, 1.93)	1.31 (-0.88, 3.05)
Anxiety	0.29 (-0.07, 1.12)	0.31 (-0.02, 1.21)	0.32 (-0.05, 1.15)	0.35 (0.00, 1.26)
Medication				
Antipsychotics	-0.73 (-1.56, 0.17)	-0.82 (-1.75, 0.10)	-0.78 (-1.61, 0.10)	-0.91 (-1.91, 0.03)
Other psychotropic drugs [‡]	0.03 (-0.81, 0.66)	0.07 (-0.73, 0.64)	0.22 (-0.49, 0.86)	0.30 (-0.47, 0.93)
Lipid lowering drugs	-0.16 (-0.73, 0.76)	0.04 (-0.70, 0.97)	-0.02 (-0.57, 0.95)	0.21 (-0.49, 1.23)

Abbreviations: CI, confidence interval; CPRD, Clinical Practice Research Datalink; UK, United Kingdom; IMD, Index of Multiple Deprivation.

[†]. An indicator for a diagnosis of either autism spectrum disorder, obsessive compulsive disorder, bipolar disorder, or schizophrenia.

[‡]. Which include benzodiazepine drugs, anxiolytics, barbiturates and hypnotics.