One-Stage Dynamic Treatment Regimes in Longitudinal Studies with Covariate-Driven Observation Times

Janie Coulombe
janie.coulombe@umontreal.ca
Department of Mathematics and Statistics
Université de Montréal

Joint work with Dr. Erica E.M. Moodie, Dr. Susan M. Shortreed, and Dr. Christel Renoux

JSM 2022, Washington DC
August 10, 2022

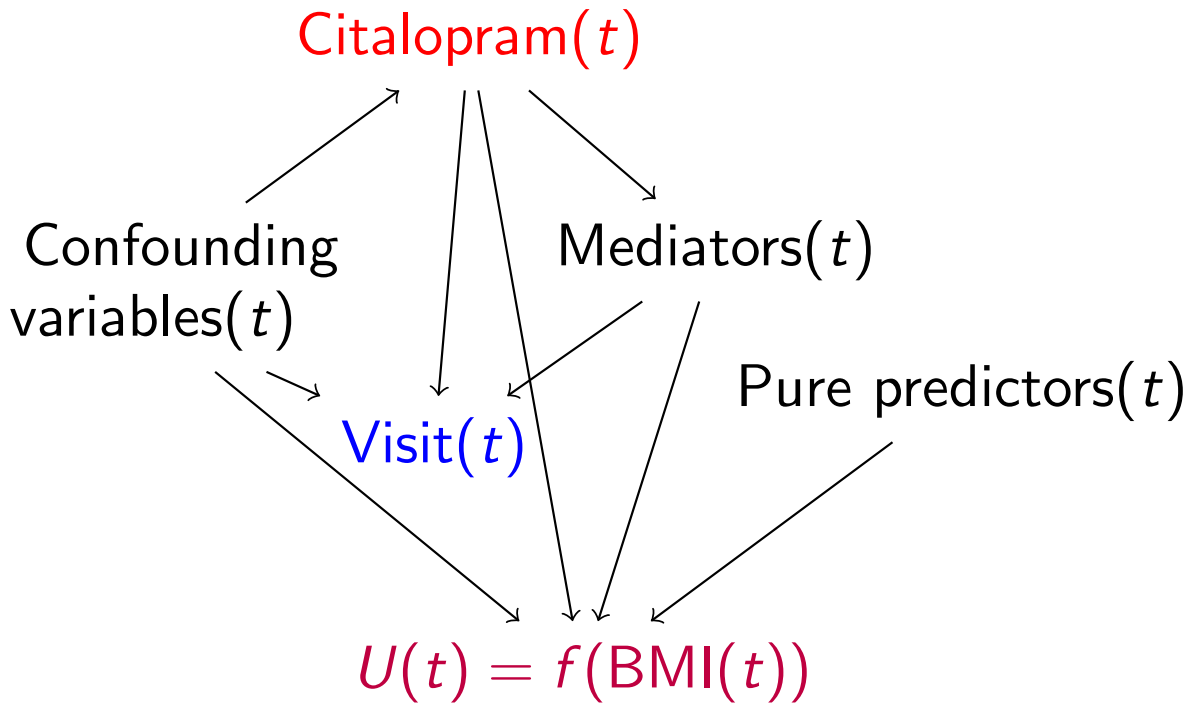
Context

- Cohort of new users of citalopram or fluoxetine (31120 patients) from the *Clinical Practice Research Datalink* (CPRD)
- We aim to develop an individualized treatment rule of the type

  "**Treat with citalopram if detrimental changes in weight are minimized under citalopram, treat with fluoxetine otherwise**"

- Repeated measures of a utility function $U(t)$, a function of BMI(t) ("informative monitoring")
- Confounding

The causal diagram that is postulated at each time t:



with $\text{Visit}(t) = 1$ corresponding to the observation of $U(t) = f(\text{BMI})$.

The individualized treatment rule will depend on treatment effect estimates.

Notation

Suppose a random sample of patients indexed by $i = 1, ..., n$. Let:

- $A_i(t)$ and $Y_i(t)$ the binary treatment and the continuous outcome, $\mathbf{K_i(t)}$ the confounding variables
- The set $\mathbf{X}^{\beta}(\mathbf{t}) = [\mathbf{1} \;\; \mathbf{K(t)} \;\; \mathbf{Q(t)}]$
- The effect modifiers $\mathbf{X}^{\psi}(\mathbf{t})$
- $\mathbf{X(t)} = \left[\mathbf{X}^{\beta}(\mathbf{t}) \;\; \mathbf{X}^{\psi}(\mathbf{t})\right]$
- $dN_i(t)$ the visit indicator ($=$ observation of $Y_i(t)$)
- C_i the follow-up time of patient i
- $\xi_i(t) = \mathbb{I}(C_i \geq t)$ the indicator for still being at risk
- We assume $dN_i(t) \perp Y_i(t) | \mathbf{V_i(t)}$ a set of variables

The outcome model includes a treatment-free model and a blip:

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

The following rule:

"Treat with citalopram if the outcome expectation is optimized under citalopram, and with fluoxetine otherwise"

becomes

"Treat with citalopram if $\hat{\boldsymbol{\psi}}'\mathbf{X}_i^{\psi}(t) \geq 0$, and with fluoxetine otherwise."

Potential outcome (Neyman, 1923; Rubin, 1974)

- Conditional exchangeability:

$$A_i(t) \perp \{Y_{i0}(t), Y_{i1}(t)\} \,|\mathbf{K_i(t)}, \mathbf{V_i(t)}, \mathbf{X}^{\psi}(\mathbf{t}), dN_i(t)$$

- Positivity:

$$0 < \mathbb{P}(A_i(t)|\mathbf{K_i(t)}, \mathbf{X}_{\mathbf{i}}^{\psi}) < 1$$

$$0 < \mathbb{P}(dN_i(t)|\mathbf{V_i(t)}) < 1$$

- Consistency:

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

Estimand: $\text{CTE} = \mathbb{E}[Y_{i1}(t) - Y_{i0}(t)|\mathbf{X}_{\mathbf{i}}^{\psi}(t)] =$

$$\mathbb{E}\left[\frac{\mathbb{I}(A_i(t)=1, dN_i(t)=1)Y_i(t)}{\mathbb{P}(A_i(t)=1|\mathbf{K_i(t)})\mathbb{P}(dN_i(t)=1|\mathbf{V_i(t)})}|\mathbf{X}_{\mathbf{i}}^{\psi}(\mathbf{t})\right] - \mathbb{E}\left[\frac{\mathbb{I}(A_i(t)=0, dN_i(t)=1)Y_i(t)}{\mathbb{P}(A_i(t)=0|\mathbf{K_i(t)})\mathbb{P}(dN_i(t)=1|\mathbf{V_i(t)})}|\mathbf{X}_{\mathbf{i}}^{\psi}(\mathbf{t})\right].$$

- We model observation times using the proportional rate model:

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

where γ are estimated using the Andersen and Gill model (1982).

This is used to compute inverse intensity of visit (IIV) weights $\varphi_i(\gamma, \mathbf{V_i(t)})$ (Lin et al., 2004).

- IIV weights are combined with inverse probability of treatment (IPT) weights:

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

- The estimating equation to solve (extension of the dWOLS proposed by Wallace and Moodie (2015)):

$$U(\boldsymbol{\beta}, \boldsymbol{\psi}; \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\omega}}) = \sum_{i=1}^{n} \int_0^{\tau} \varphi_i(\hat{\boldsymbol{\gamma}}, \mathbf{V_i(t)}) e_i(\hat{\boldsymbol{\omega}}, \mathbf{K_i(t)})$$

$$\times \begin{bmatrix} \frac{\partial f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\}}{\partial \boldsymbol{\beta}} \\ A_i(t)\mathbf{X}_i^{\psi}(t) \end{bmatrix} \left[Y_i(t) - f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\} - A_i(t)\boldsymbol{\psi}'\mathbf{X}_i^{\psi}(t) \right] dN_i(t) = \mathbf{0}.$$

- We can compute the asymptotic variance using theory on two-step estimators (Newey et McFadden, 1994)
- Note: It requires an acute treatment effect, no carryover effect of subsequent treatments (Dong et al., 2021)

Simulation study

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

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

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

$$Q_i(t) \sim Bern(0.5)$$

$$Y_i(t) = \sqrt{(t/100)} - 2\, A_i + 2.5\{Z_i(t) - \mathbb{E}[Z_i(t)|A_i]\} +$$
$$0{,}4K_{i1i} + 0{,}05K_{2i} - 0{,}6K_{3i} + 0.5\{A_i \times Q_i(t)\} - 1\{A_i \times K_{1i}\} + \epsilon_i(t)$$

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

$$dN_i(t) \sim Poisson(\lambda_i(t)),\ \lambda_i(t) = \eta_i \exp\{\gamma_A A_i + \gamma_Z Z_i(t) + \gamma_{K2} K_2 + \gamma_{K3} K_3\}$$

where η_i is a random effect, $(\gamma_A, \gamma_Z, \gamma_{K2}, \gamma_{K3})$ the dependence parameters

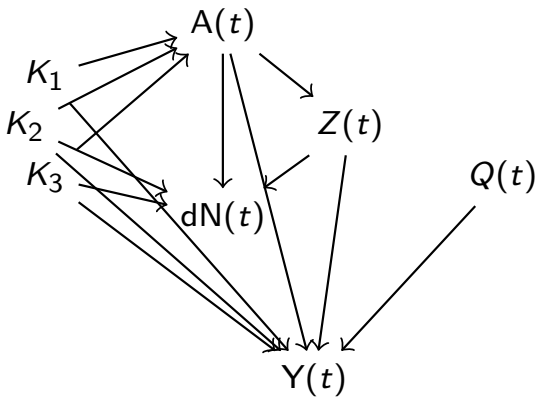
Estimators that were compared

To assess the method's robustness, we compare:

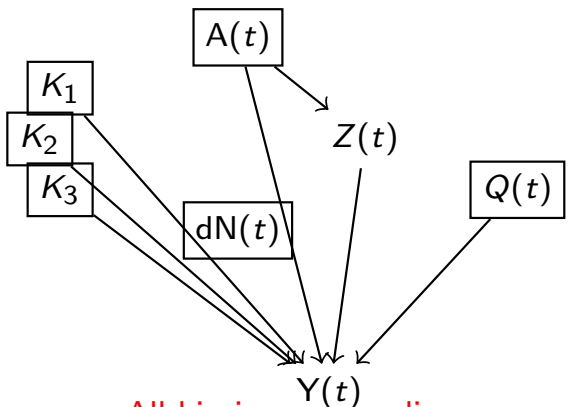
	Misspecified model (x)			
Estimator	Visit (partial[1])	Visit (full[2])	Treatment[3]	Outcome[4]
$\hat{\psi}_{DW1}$				
$\hat{\psi}_{DW2}$	x			x
$\hat{\psi}_{DW3}$	x		x	
$\hat{\psi}_{DW4}$		x		
$\hat{\psi}_{OLS}$		x	x	
$\hat{\psi}_{IPT}$		x		

1. Adjusted for the important variables (treatment and mediator)
2. Not adjusted for the mediator, adjusted for the treatment and K_2
3. Wrong functional form (squared terms) for the linear terms of K_1 and K_3
4. Misses the adjustment for K_2

Corresponding causal diagrams

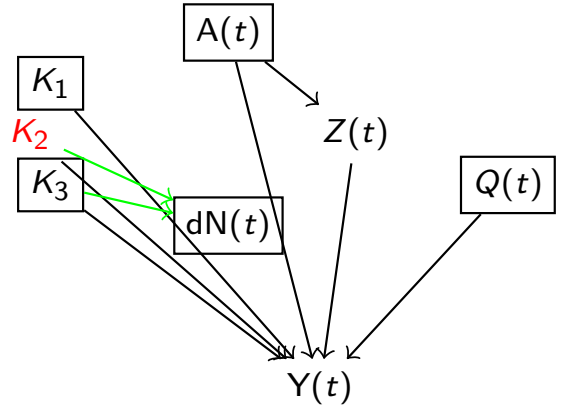


(a) Data generating mechanism

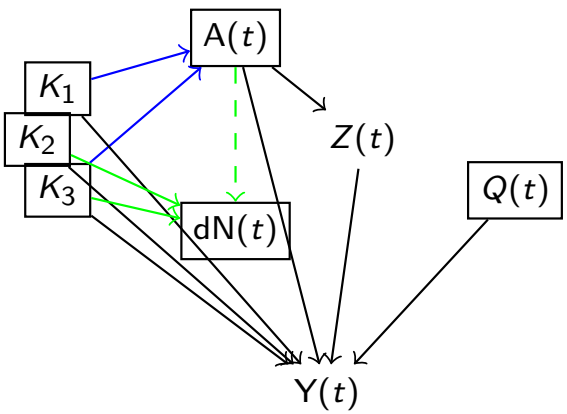


All biasing arrows disappear

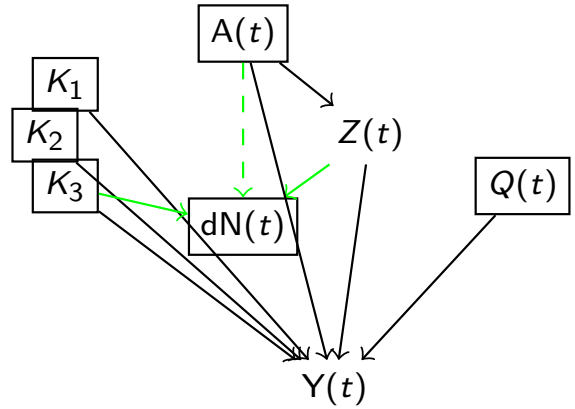
(b) Proposed estimator



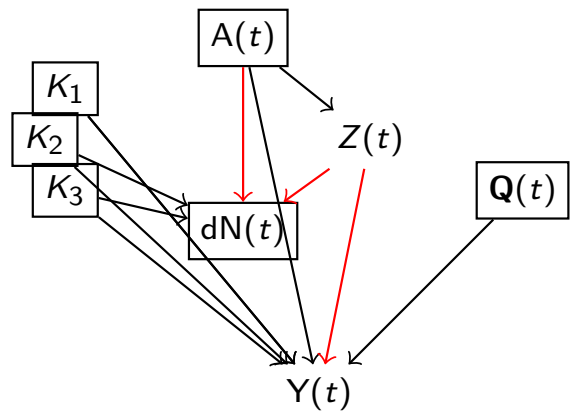
(c) DW2: Wrong observation and outcome models



(d) DW3: Wrong observation and treatment models



(e) DW4: Fully wrong observation model ($dN_i(t) \sim A(t) + K_2$)



(f) IPT estimator (prone to collider stratification bias)

Results

Table 1: Results of the simulation study ($M = 1000$ simulations) - error rate for the optimal treatment decision

Sample size	Parameters γ^{υ}	No. obs. times mean (IQR)	Error rate $\hat{\psi}_{DW1}$	$\hat{\psi}_{DW2}$	$\hat{\psi}_{DW3}$	$\hat{\psi}_{DW4}$	$\hat{\psi}_{OLS}$	$\hat{\psi}_{IPT}$
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

υ.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).
Abbrev.: IQR, interquartile range.

31K patients from the CPRD, 48K measurements

Outcome definition:

$$U(t) = 100 - 5 \times \mathbb{I}[\,\text{Detrimental change in BMI}(t)\text{ category}\,]$$
$$+ \mathbb{I}[\,\text{BMI}(0) < 18.5 \;\cup\; (18.5 \leq \text{BMI}(0) \leq 24.9 \;\cap\; \text{BMI}(t) < 20)\,] \times \{\%\text{ increase BMI}(t)\}$$
$$- \mathbb{I}[\,\text{BMI}(0) \geq 25 \;\cup\; (18.5 \leq \text{BMI}(0) \leq 24.9 \;\cap\; \text{BMI}(t) > 23.5)\,] \times \{\%\text{ increase BMI}(t)\}\,,$$

with $\mathbb{I}$ the indicator function. The rule is: *Treat with citalopram if*

$$-1.45 + 0.16 \times \mathbb{I}\,[\text{Male sex}] + 0.08 \times \mathbb{I}\,[\text{Has been smoker}] + 0.35 \times \mathbb{I}\,[\text{Anxiety}]$$
$$+0.13 \times [\text{Index Multiple Deprivation}] + 0.42 \times \mathbb{I}\,[\text{Alcohol abuse}] + 1.31 \times \mathbb{I}\,[\text{Psychiatric disease}]$$
$$-0.91 \times \mathbb{I}\,[\text{Use of antipsychotics}] + 0.30 \times \mathbb{I}\,[\text{Other psychotropics}] + 0.21 \times \mathbb{I}\,[\text{Lipid lowering drugs}] > 0.$$

Table 2: Comparison of fitted outcomes, CPRD, 1998-2017

	Mean fitted outcome (SE)			
Treatment	$\hat{\psi}_{OLS}$	$\hat{\psi}_{IPT}$	$\hat{\psi}_{IIV}$	$\hat{\psi}_{DW1}$
Received	98.2 (0.001)	98.2 (0.001)	98.3 (0.001)	98.3 (0.001)
Optimal	98.9 (0.001)	98.9 (0.001)	99.1 (0.001)	99.0 (0.001)

Abbrev.: SE, standard error.

Discussion

- In a standard analysis, we condition on the indicator of being observed (the causal diagram is a useful tool!)
- The methods discussed rely on several assumptions (causal, models, data generating mechanism, treatment effect acuteness)
- Extension: Rule for a sequence of treatments (multiple time points)

Acknowledgments



Drs
Erica E. M. Moodie
Susan M. Shortreed
Christel Renoux

Contact:
janie.coulombe@umontreal.ca

Funding and support:

NSERC CRSNG

compute canada | calcul canada

HEALTHY BRAINS HEALTHY LIVES

NIH National Institute of Mental Health

McGill UNIVERSITY

References

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.

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 Thiébaut, 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.

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.