

# Apport des données télématiques pour comprendre le risque en assurance automobile

Arthur Charpentier<sup>1</sup>

<sup>1</sup> Université du Québec à Montréal & AXA JRI

Séminaire du Réseau de recherche en sécurité routière, Juin 2022



RÉSEAU DE RECHERCHE  
EN SÉCURITÉ ROUTIÈRE  
du Québec

# Rapide présentation

**Arthur Charpentier**

Université du Québec à Montréal

 @freakonometrics

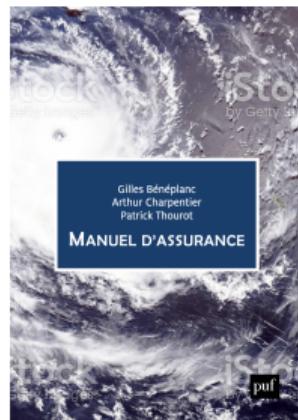
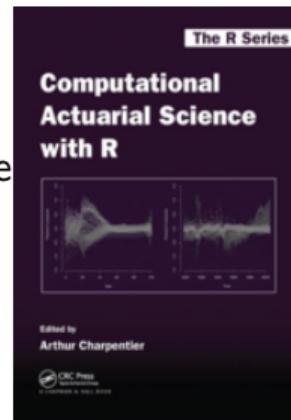
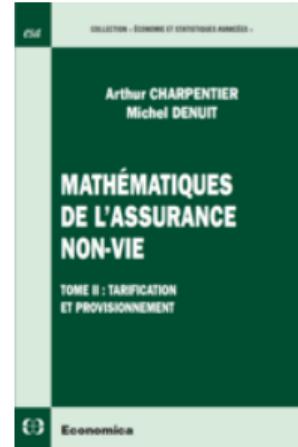
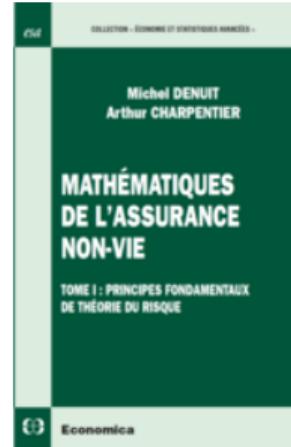
 freakonometrics

 [freakonometrics.hypotheses.org](https://freakonometrics.hypotheses.org)

Modèle prédictifs, économie de l'assurance

Apprentissage automatique

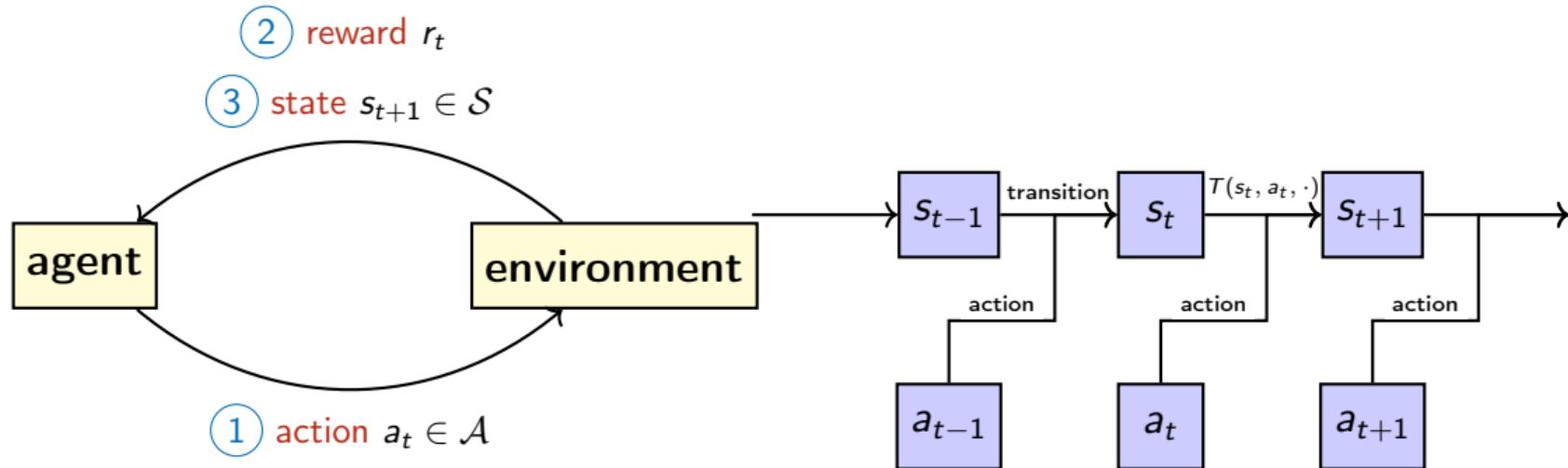
Apprentissage par renforcement, recherche de la stratégie optimale  $\pi$



# Apprentissage par renforcement, exploration / exploitation

Note: apprentissage par renforcement

$$\max_{\pi} \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s \right], \text{ où } \begin{cases} \pi(a|s) = \mathbb{P}(a_t = a \mid s_t = s) \\ P_a(s'|s) = \Pr(s_{t+1} = s' \mid s_t = s, a_t = a) \end{cases}$$



# Discrimination, proxy, corrélation et causalité

- ▶ “at the core of insurance business lies discrimination between risky and non-risky insureds”, Avraham (2017)
- ▶ gender, Verbelen, Antonio & Claeskens (2017)
- ▶ credit scoring, Kabler (2004), Arya *et al.* (2013), Morris *et al.* (2017) ou Kiviat (2019)

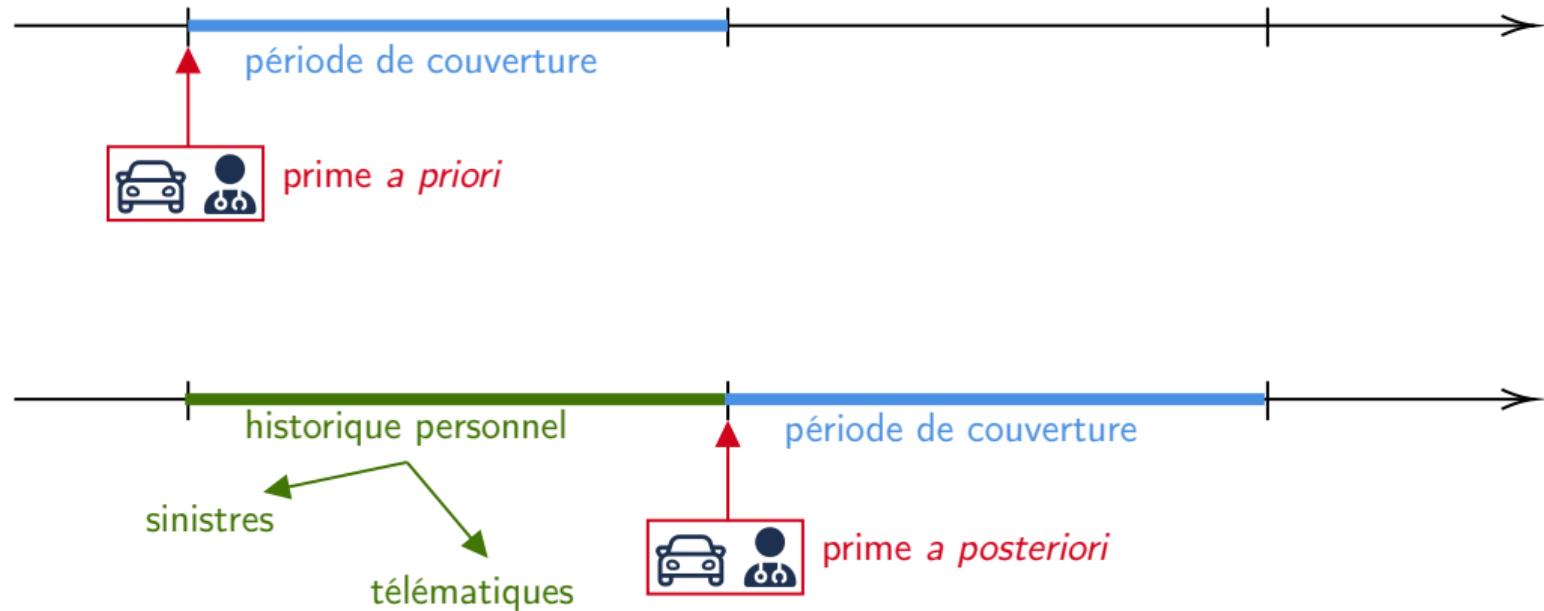


## Publications récentes

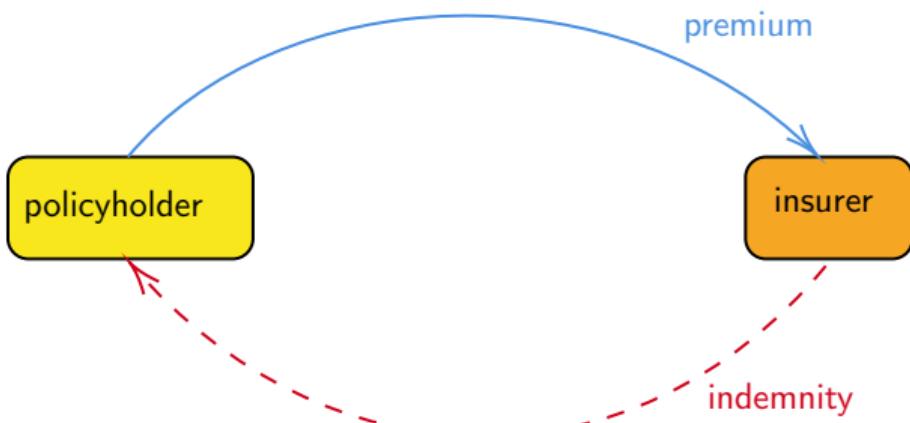
- ▶ Charpentier, A., Elie, R. & Remlinger, C. (2021) Reinforcement Learning in Economics and Finance. *Computational Economics* 
- ▶ Bénéplanc, G., Charpentier, A. & Thourot, P. (2022) Manuel d'Assurance. *Presses Universitaires de France*
- ▶ Barry, L. & Charpentier, A. (2022) The Fairness of Machine Learning in Insurance: New Rags for an Old Man?, *PARI Chair Working Paper* 
- ▶ Charpentier (2022) Assurance: Discrimination, Biais et Équité *Institut Louis Bachelier* 
- ▶ Grari, V., Charpentier, A., V. Lamprier & M. Detyniecki (2022) A fair pricing model via adversarial learning *submitted* 
- ▶ Belz, E., Charpentier, A. & Geoffard, P.Y. (2022) To sue or not to sue ? *forthcoming*

# Tarification actuarielle I

Notion d'**inversion du cycle de production** en tarification assurantielle.

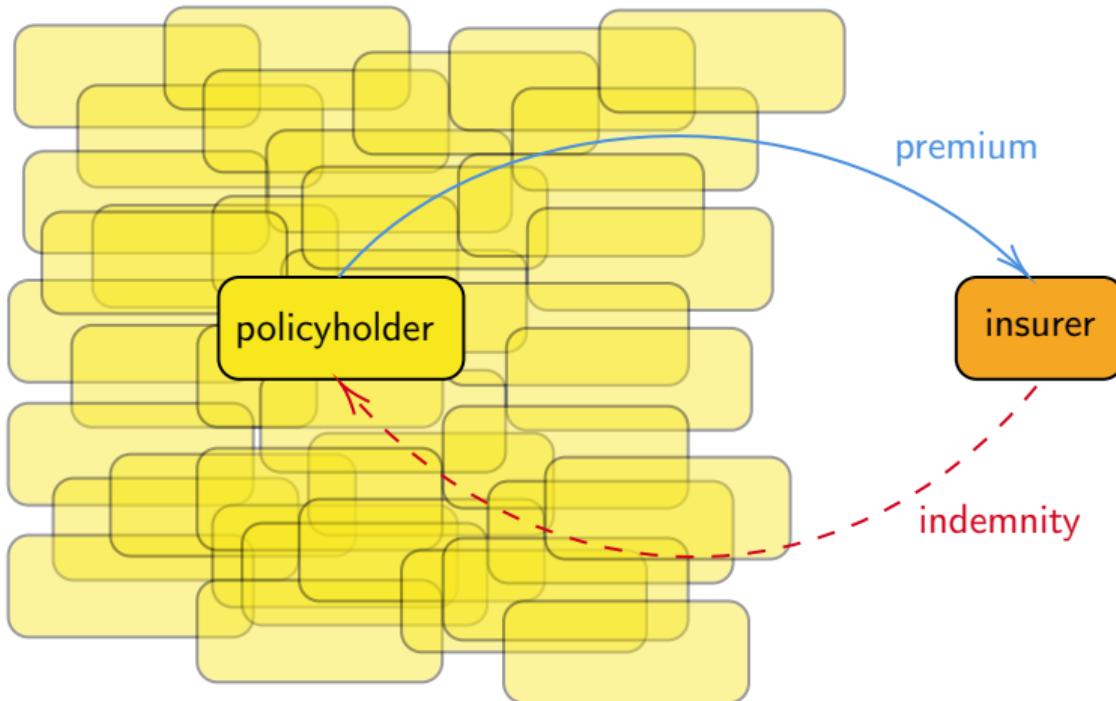


## Tarification actuarielle II



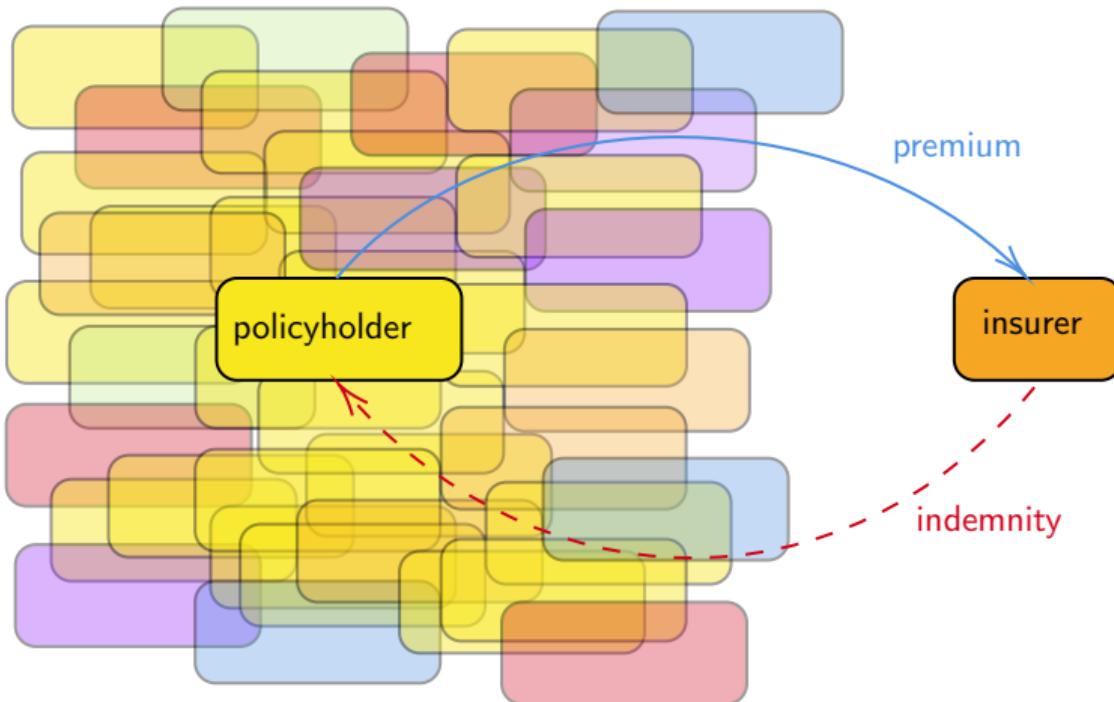
## Tarification actuarielle III

- ▶ Insurance is the contribution of the many to the misfortune of the few



## Tarification actuarielle IV

- ▶ Insurance is the contribution of the many to the misfortune of the few



## Tarification actuarielle V

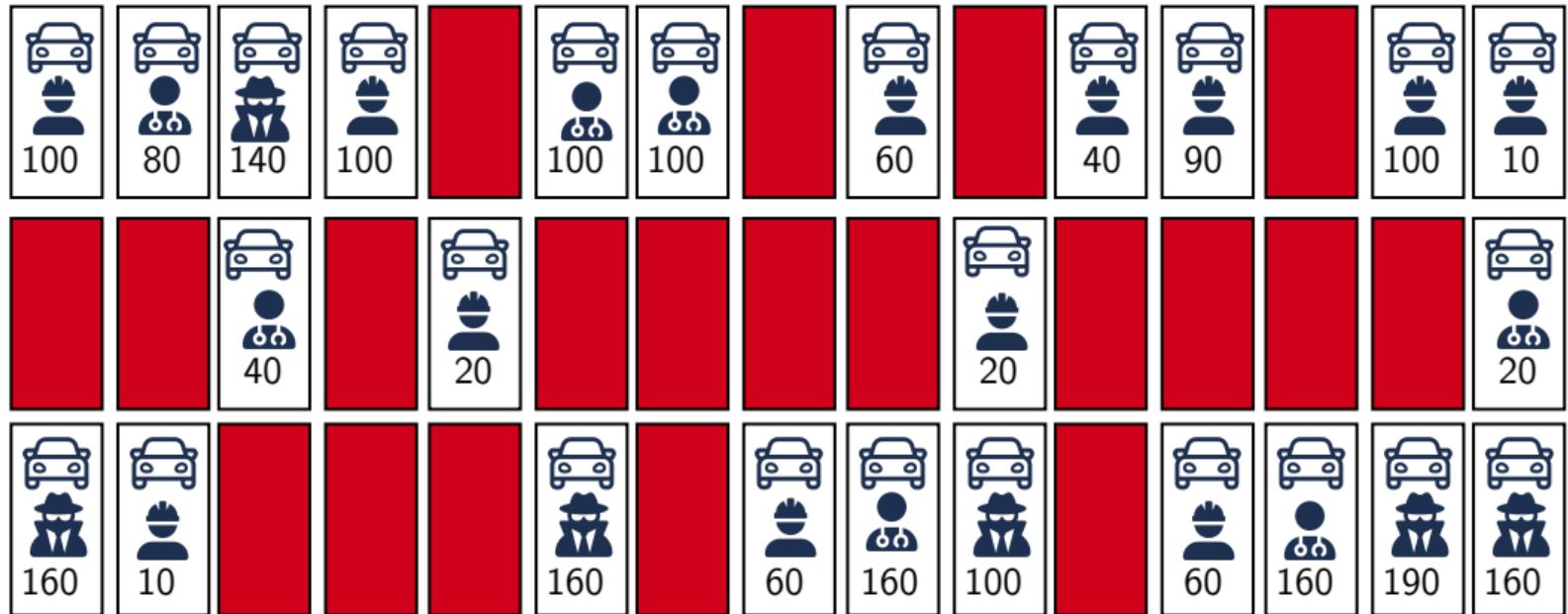
100	80	140	100	90	100	100	30	60	100	40	90	100	100	100	10
20	60	40	120	20	20	120	150	20	20	20	60	20	80	20	20
160	10	70	160	110	160	160	160	60	160	100	80	60	160	190	160

Tarification actuarielle VI

100	80	140	100	90	100	100	30	60	100	40	90	100	10	
20	60	40	120	20	20	120	150	20	20	20	60	20	20	
160	10	70	160	110	160	160	60	160	100	80	60	160	190	160

 ?  $\mathbb{E}[Y]$  ←  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \approx 85$

## Tarification actuarielle VII



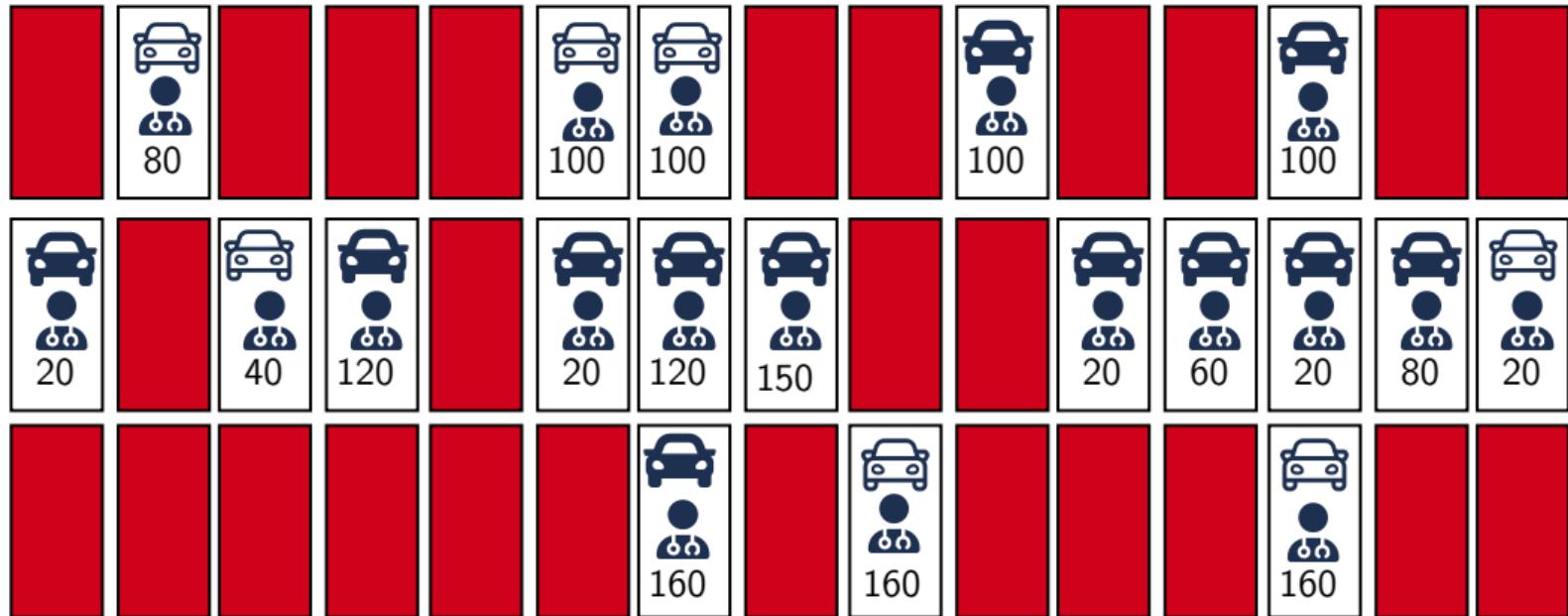
?

$\mathbb{E}[Y | \text{car}]$



$\bar{y} \approx 95$  (référence + 10)

## Tarification actuarielle VIII



?

$$\mathbb{E}[Y | \text{person}]$$



$$\bar{y} \approx 82 \quad (\text{référence} - 3)$$

## Tarification actuarielle IX

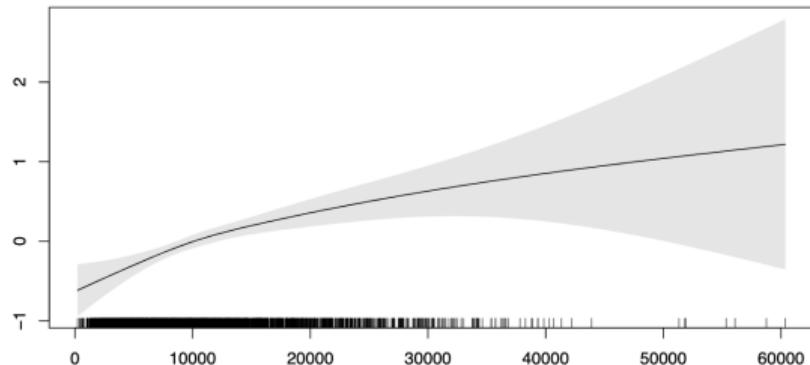
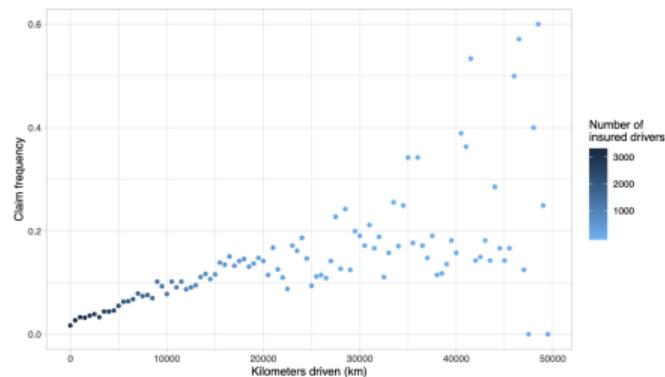
- ▶ Tarification *a priori*,  $\mathbb{E}[Y|X_1 = x_1, \dots, X_k = x_k]$   
possibles  $x$ : âge du conducteur principal, ancienneté du permis de conduire, genre du conducteur, modèle de voiture, puissance, vitesse maximale, code postal, type de carburant, etc
- ▶ Tarification *a posteriori*,  $\mathbb{E}[Y_{t+1}|Y_t = y_t, Y_{t-1} = y_{t-1}, \dots, Y_{t-k} = y_{t-k}]$   
(*experience rating*)  
marche lorsque  $Y$  correspond à la fréquence annuelle de sinistres  
(possibles discussions suivant le degré de responsabilité)

Modèle bayésien classique: si  $Y_t \sim \mathcal{P}(\lambda)$  où *a priori*  $\Lambda \sim \mathcal{G}(\alpha, \beta)$ ,  $\mathbb{E}[Y_{t+1}] = \alpha/\beta$

$$\mathbb{E}(Y_{t+1} | y_1, \dots, y_t) = \frac{\alpha + \sum_{j=1}^t y_j}{\beta + t} = \frac{\alpha + \text{nombre de sinistres}}{\beta + \text{nombre d'années}}$$

# Données télématiques I

- ▶ Utilisation de données télématiques (observées a posteriori)  
difficulté de tenir compte de l'exposition (kilométrage annuel), Litman (2005, 2011)  
*Mileage Rate Factor* (MRF) ou *Per-Mile Premiums* (PMP)  
*Pay-As-You-Drive* (PAYD) ou *Pay-How-You-Drive* (PHYD)



(picture Boucher, Côté & Guillen (2017))

## Données télématiques II

- ▶ Information sur le véhicule (OBD II) ou sur le conducteur ?



(picture various and on the right [Girma, Yan & Homaifar \(2019\)](#))

- ▶ Quelle information ? des métadonnées ?

distance parcourue, nombre de trajets, types de routes (*motorways, urban areas, abroad* ou *other*, en % du temps, ou du kilométrage), moments de conduite (22h-6h, 6h- 9h30, 9h30-16h, 16h-19h, 19h-22h), jour (semaine ou fin de semaine)

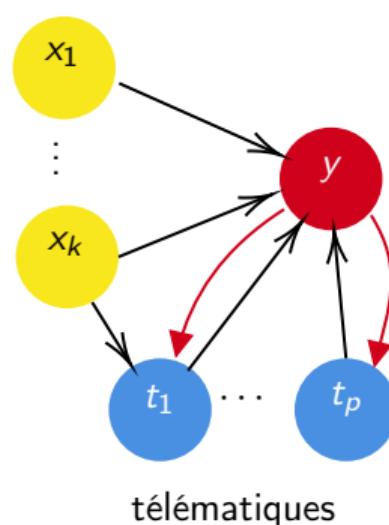
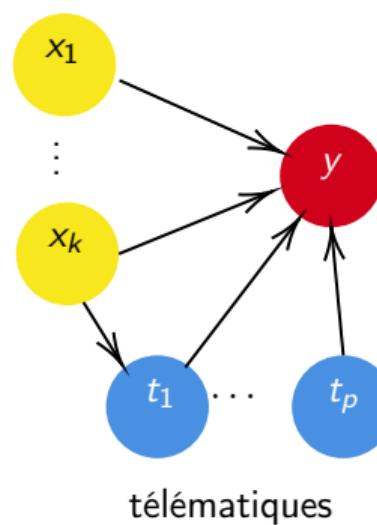
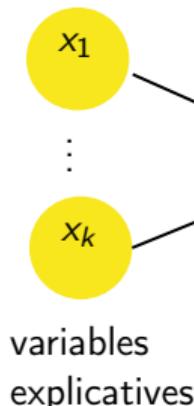
## Données télématiques III

- ▶

Quelle information ? des **scores de conduite** ?

scores de freinage (*braking score*), score d'accélération (*accelaration score*), score de virage (*turn score*)

Les actuaires conduisent des pseudo-modèles causaux

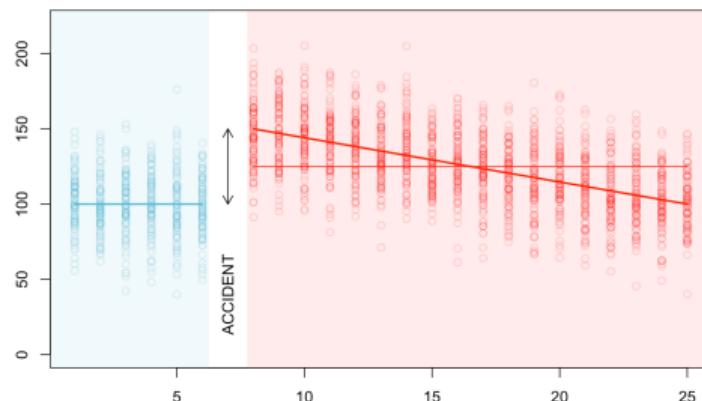
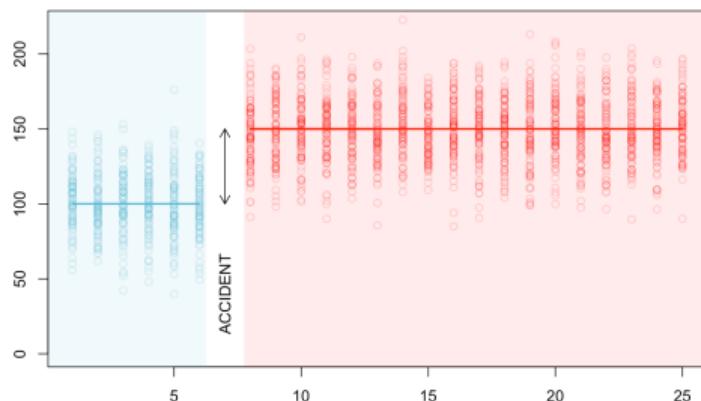


# Modèles de rupture sur données télématiques I

En prenant les données des personnes qui ont eu un accident, on veut regarder des ruptures avant/après l'accident (date  $t_0$ )

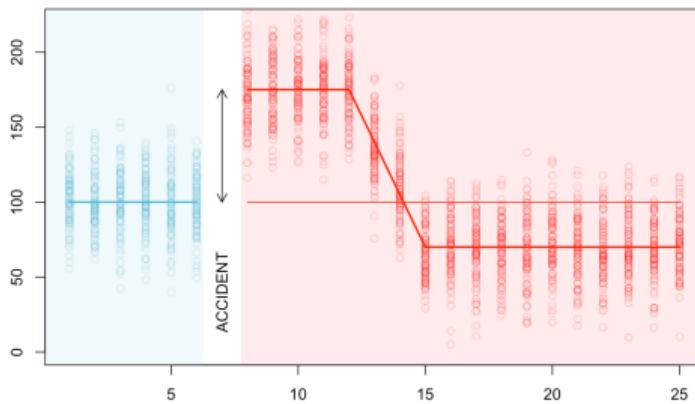
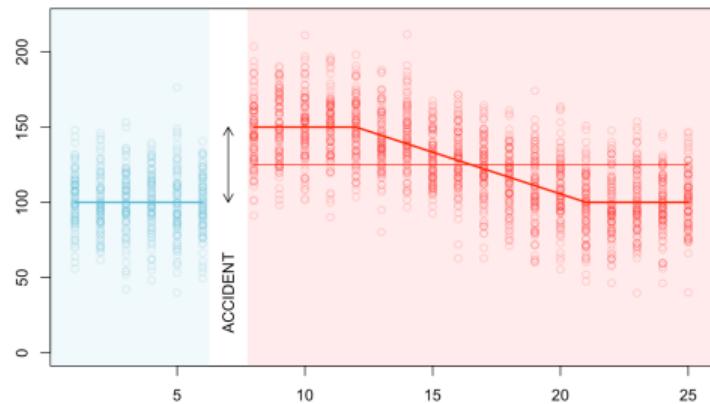
$$y_{i,t} = \alpha + \mu_i + \beta 1_{t>t_0} + \varepsilon_{i,t} \text{ où } 1_{t>t_0} = \begin{cases} 1 \text{ après un accident} \\ 0 \text{ avant un accident} \end{cases}$$

$$y_{i,t} = \alpha + \mu_i + \beta 1_t + \gamma(t - t_0)_+ + \varepsilon_{i,t}$$



# Modèles de rupture sur données télématiques II

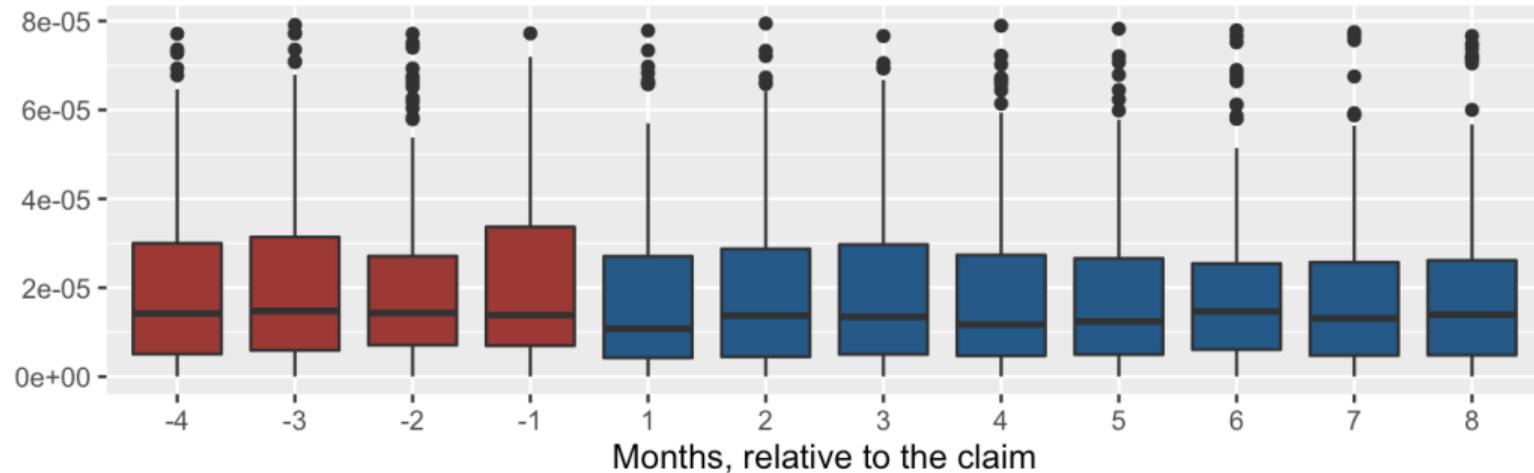
- ▶ y-a-t-il un changement ?  $\beta \neq 0$  ?
- ▶ ce changement est-il temporaire ? a-t-on un retour à la situation avant ?  
à quelle vitesse ? revient-on au même niveau qu'avant ?



## Modèles de rupture sur données télématiques III

Note: les vrais graphiques ont l'allure suivante, par exemple braking

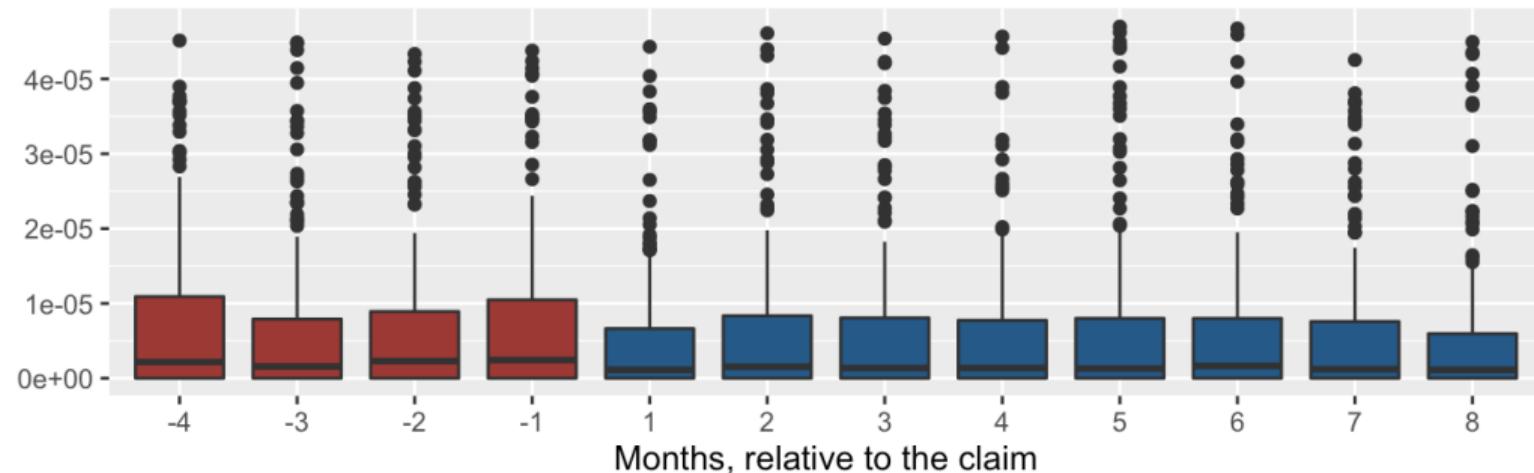
tlm\_TotalBraking/tlm\_TotalDriveDistance



## Modèles de rupture sur données télématiques IV

Note: les vrais graphiques ont l'allure suivante, par exemple acceleration

tlm\_TotalAcceleration/tlm\_TotalDriveDistance



## Modèles de rupture sur données télématiques V

En supposant un modèle linéaire (pour simplifier l'écriture), pour un assuré  $i$ , à la date  $t$ , pour une variable télématique donnée  $x_{i,t}$

►  $y_{i,t} = \alpha + \beta 1_t + \varepsilon_{i,t}$

où  $1_t = \begin{cases} 1 & \text{après un accident} \\ 0 & \text{avant un accident} \end{cases}$

il y a un changement si  $\beta \neq 0$

E.g. Trip\_count:  $-20.61\%^{***}$

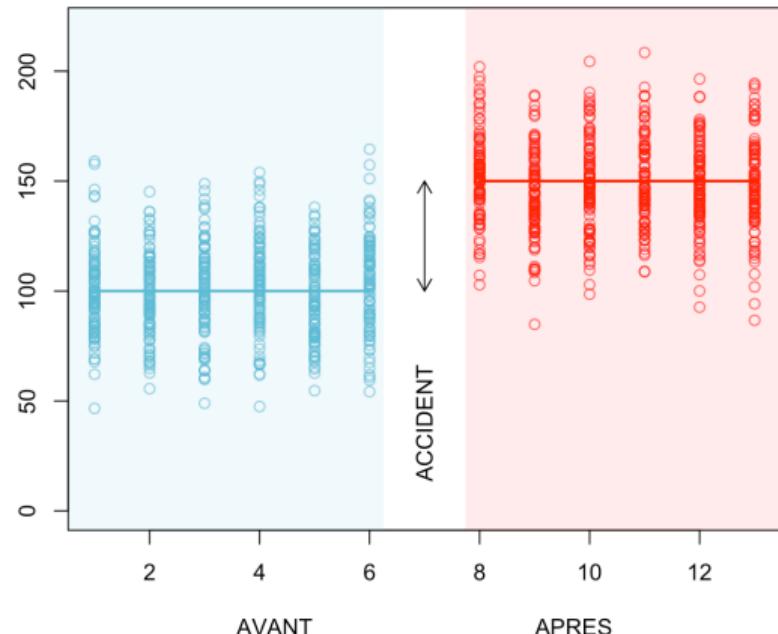
heavy\_acceleration:  $-31.41\%^*$

medium\_acceleration:  $-31.14\%^{***}$

heavy\_braking:  $-38.32\%^{***}$

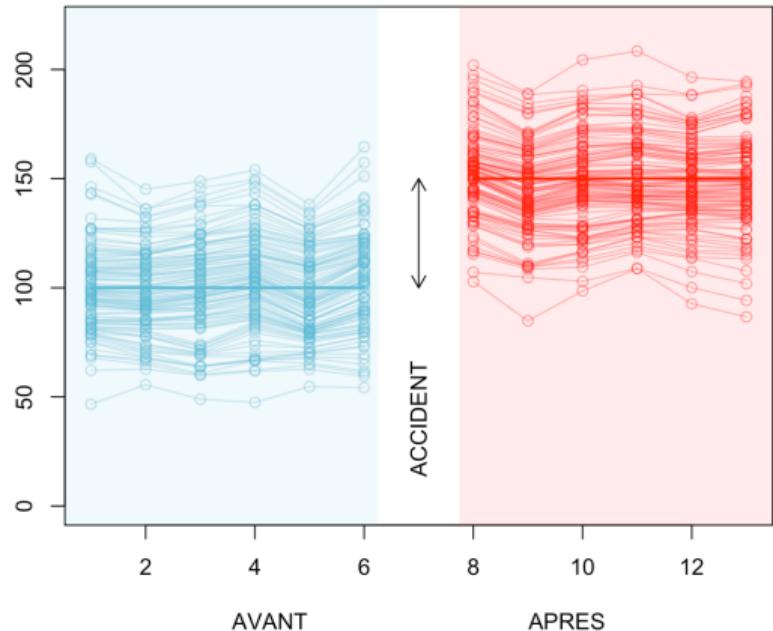
night\_time:  $-15.82\%^{***}$

total\_distance:  $-22.52\%^{***}$



# Modèles de rupture sur données télématiques VI

- ▶  $y_{i,t} = \alpha + \beta 1_t + \mu_i + \varepsilon_{i,t}$   
pour tenir compte d'un effet individuel  
(conducteur par conducteur)  
il y a un changement si  $\beta \neq 0$



## Modèles de rupture sur données télématiques VII

- ▶  $y_{i,t} = \alpha + \mu_i + \beta \mathbf{1}_t + \sum_{m=1}^{12} \gamma_m \text{mois}(i, t) + \varepsilon_{i,t}$  pour tenir compte du mois de l'année
- ▶  $y_{i,t} = \alpha + \mu_i + \beta \mathbf{1}_t + \sum_{m=1}^{12} \gamma_m \text{mois}(i, t) + \delta \text{age}_i + \varepsilon_{i,t}$  pour tenir compte de l'âge

E.g. medium\_acceleration:  $-0.0015 (-131.868\%***, p = 3 \cdot 10^{-5})$

heavy\_braking:  $-0.0014 (-138.55\%***, p = 3 \cdot 10^{-6})$

heavy\_cornering:  $-0.0027 (-120.55\%***, p = 0.003)$

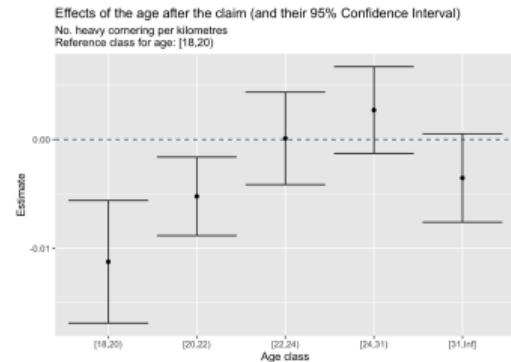
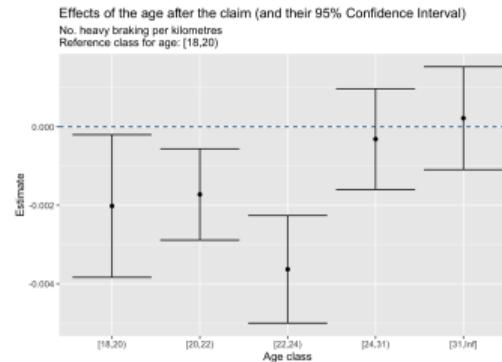
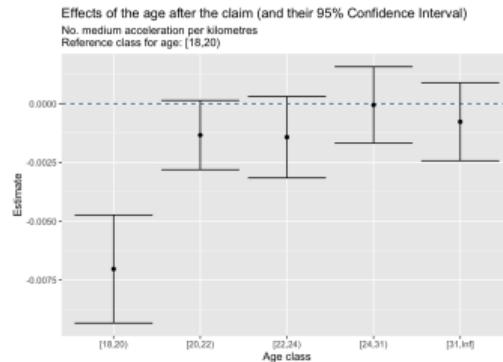
lane\_change:  $-9.34 \cdot 10^{-5} (-128.531\%*, p = 0.05)$

total\_trips:  $-26.39 (-123.36\%***, p = 1 \cdot 10^{-34})$

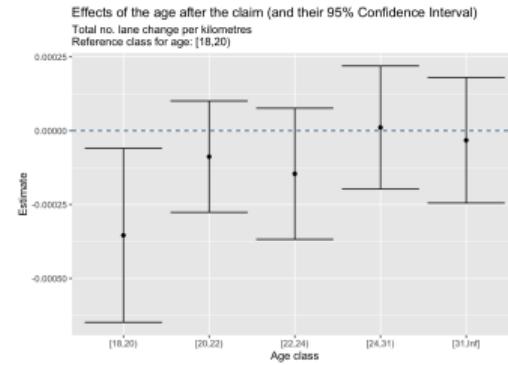
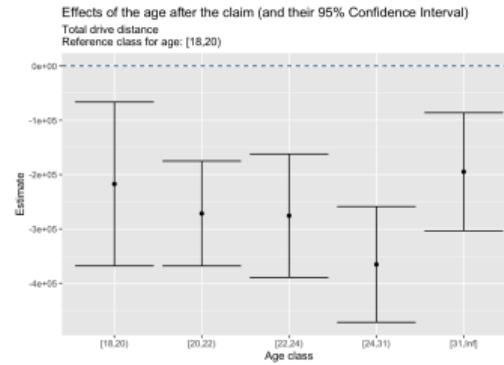
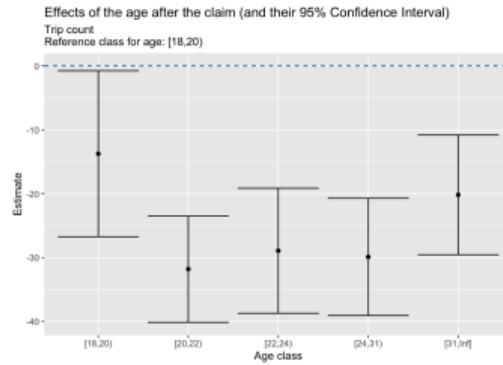
total\_distance:  $-270.93 (-126.75\%***, p = 2 \cdot 10^{-27})$

# Modèles de rupture sur données télématiques VIII

On peut aussi regarder un effet croisé entre la rupture et l'âge (classe d'âge)

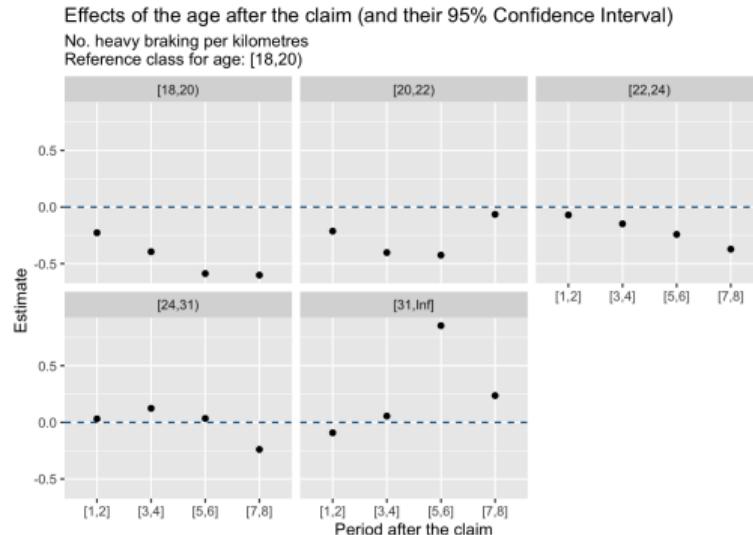
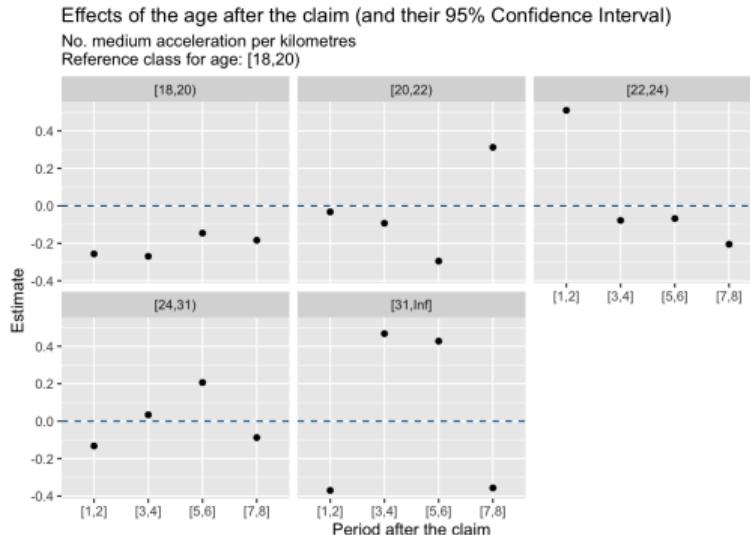


# Modèles de rupture sur données télématiques IX



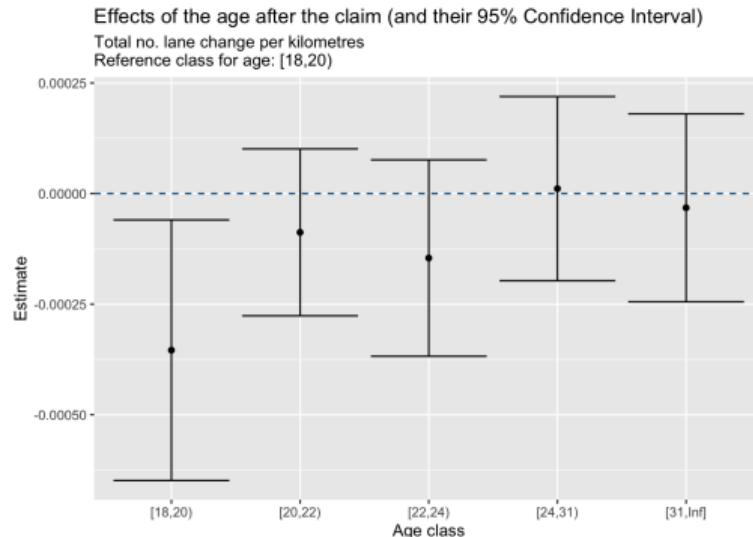
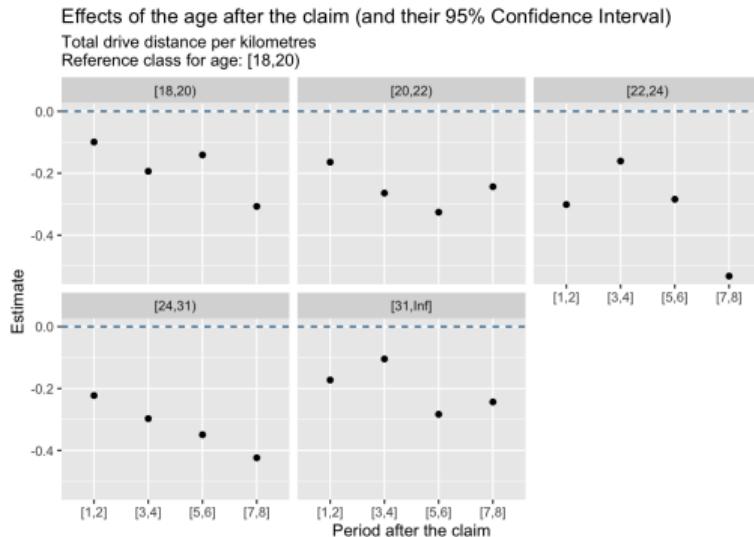
# Modèles de rupture sur données télématiques X

Et regarder l'évolution de  $\beta$  (2 premiers mois, 2 mois suivants, etc)

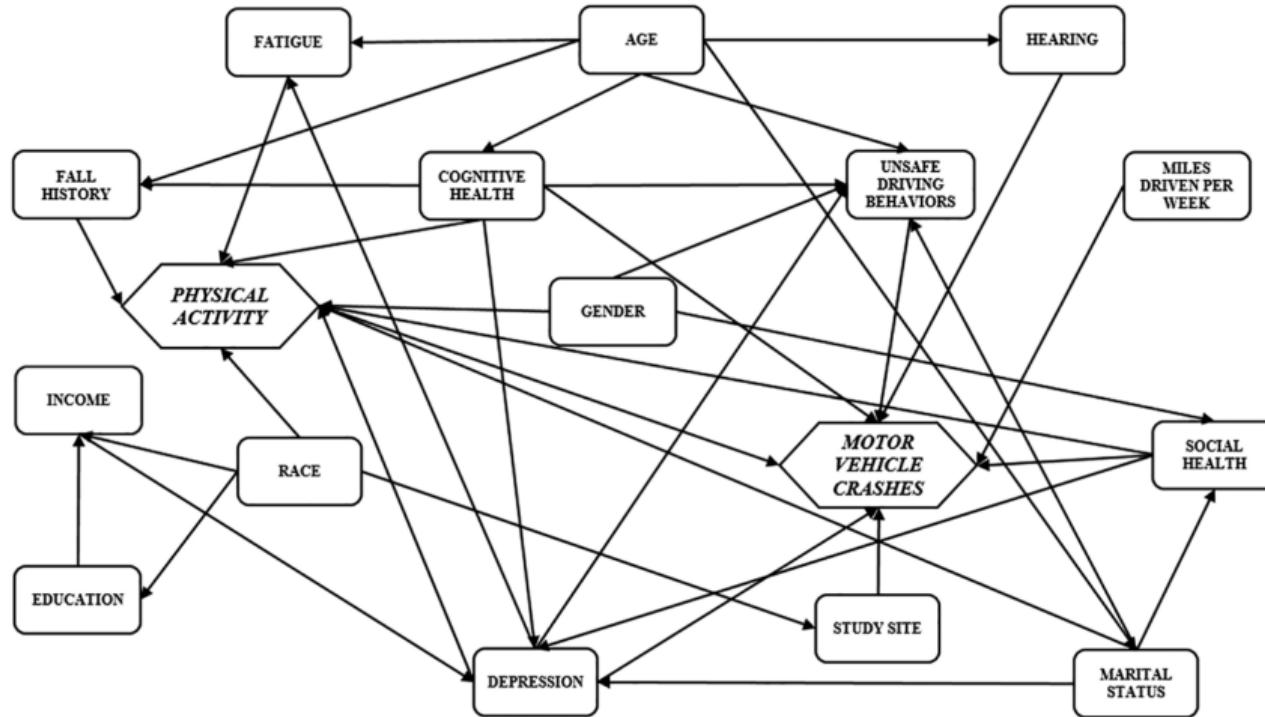


# Modèles de rupture sur données télématiques XI

Et regarder l'évolution de  $\beta$  (2 premiers mois, 2 mois suivants, etc)



# Quelle place pour l'analyse causale ? I



(picture Talwar, Mielenz, Hill *et al.* (2019))