

Causal inference to formalize responsibility analyses in road safety epidemiology

Marine Dufournet^{1,2,3}, Emilie Lanoy^{4,5}, Jean-Louis Martin^{1,2,3}, Vivian Viallon^{1,2,3}

Affiliation des auteurs :

¹ Université de Lyon, F-69622, Lyon, France

² IFSTTAR, UMRESTTE, UMR_T9405, Unité Mixte de Recherche Epidémiologique et de Surveillance Transport Travail Environnement-69675 Bron

³ Université Claude Bernard Lyon 1, F-69373 Lyon

⁴ Gustave Roussy, Université Paris-Saclay, Service de biostatistique et d'épidémiologie, Villejuif, F-94805

⁵ Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, F-94085

Correspondance :

Marine DUFOURNET

Téléphone : +33 4 78 65 68 50

e-mail : marine.dufournet@ifsttar.fr

Thème 5: Approche statistique de la causalité

Mots clés : inférence causale, DAG, biais de sélection, analyse en responsabilité, sécurité routière

Depuis plusieurs années, les modèles causaux structuraux (SCM) fournissent différents outils afin d'identifier et d'estimer des effets causaux sur des données dites observationnelles. Dans le cadre de la sécurité routière, l'estimation d'effets causaux est d'autant plus complexe que les données disponibles ne concernent généralement que les accidentés. Pour contourner ce problème d'absence de témoins, il est maintenant classique d'utiliser des « analyses en responsabilité » et de comparer le risque d'accident chez les responsables vs les non-responsables de l'accident. Au traditionnel biais de confusion rencontrés généralement en épidémiologie s'ajoute donc un biais de sélection extrême.

L'objectif de nos travaux est de déterminer quels effets causaux sont identifiables et estimables à partir des analyses en responsabilité, selon les données d'accidentés disponibles. Pour cela, nous utilisons les outils de l'inférence causale (DAG, contrefactuels) et les récents résultats sur l'identifiabilité des effets causaux en présence de biais de sélection. Nous considérons comme exemple l'estimation de l'effet causal de l'alcool sur la responsabilité de l'accident dans deux cas. Le premier, lorsque toutes les données d'accidents sont disponibles. Le second, qui est le plus fréquent en pratique, lorsque les données concernent uniquement les accidents sévères (corporels ou mortels par exemple).

Nous montrons que, dans le premier cas, l'effet causal de l'alcool sur la responsabilité peut être approximé par un autre effet causal identifiable, sous certaines hypothèses. Lorsque les données disponibles dépendent de la sévérité de l'accident, l'effet causal de l'alcool sur la responsabilité n'est pas identifiable, et ne peut être approximé. Nous présentons des résultats numériques afin d'illustrer nos résultats théoriques et donner un premier ordre de grandeur du biais induit par les analyses en responsabilité. Enfin, nous montrons que les effets d'association estimés dans le deuxième cas n'ont de validité ni externe ni interne.

Nos résultats fournissent une description formelle des analyses en responsabilité et donnent une première quantification du biais induit par ces analyses. La définition formelle que nous fournissons de l'association estimée dans ces analyses n'est pas seulement utile dans le cadre de la sécurité routière. Elle permet également d'expliquer d'autres biais de sélection, comme celui en jeu dans le cadre du paradoxe de l'obésité.