

# BIAIS DE COLLISION ET PARADOXE DE L'OBÉSITÉ

Vivian Viallon & Marine Dufournet

*Univ. Lyon, Université Claude Bernard Lyon1, Ifsttar, UMRESTTE, UMR T\_9405, F-69373, LYON.*

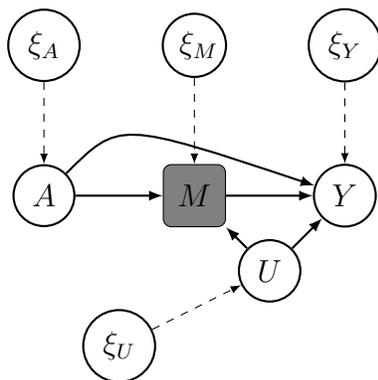
**Résumé.** Le paradoxe de l'obésité a été rapporté dans plusieurs études observationnelles : l'obésité protégerait d'une mortalité précoce chez des patients souffrant d'une maladie chronique, telle que le diabète. Des arguments causaux ont récemment été avancés pour expliquer ce paradoxe. Le conditionnement sur un médiateur, le diabète ici, crée un biais de collision, si bien que le côté protecteur de l'obésité observé chez les patients diabétiques n'a pas de valeur causale. Récemment, Sperrin et al. (2016) ont relancé le débat prétendant que le biais était vraisemblablement faible et ne pouvait expliquer à lui seul le paradoxe. Cependant, des failles existent dans les arguments de Sperrin et al., qui remettent grandement en cause leurs conclusions. Nous évaluons le biais entre (i) l'association entre l'obésité et la mortalité précoce parmi les patients diabétiques  $\Delta_{AS}$  et (ii) l'effet causal considéré par Sperrin et al. Dans le cadre des modèles causaux structuraux, nous expliquons pourquoi ce biais peut être bien plus important que ce que ces auteurs rapportent. Nous considérons également des effets causaux alternatifs et étudions leur différence avec  $\Delta_{AS}$ . Des exemples numériques sont présentés pour illustrer la magnitude des biais correspondants sous des scénarios réalistes.

**Mots-clés.** Causalité, Biais de collision, paradoxe de l'obésité

**Abstract.** The “obesity paradox” has been reported in several observational studies, where obesity was shown to be associated to a decreased mortality in individuals suffering from a chronic disease, such as diabetes. Causal arguments have recently been given to explain this apparently paradoxical fact: conditioning on a mediator, diabetes in this example, creates a collider bias and the observed protective “effect” of obesity among diabetic patients has no causal value. Recently, Sperrin et al. (2016) relaunched the debate and claimed that the resulting bias was likely to be weak and therefore unlikely to be the main explanation for the obesity paradox. However, a number of issues in their work make their conclusions questionable. We study the bias between (i) the association between obesity and early death among diabetic patients  $\Delta_{AS}$  and (ii) the causal effect considered by Sperrin et al. Under the usual framework of structural causal models, we explain why this bias can be much higher than what these authors reported. We further consider alternative causal effects of potential interest and study their difference with  $\Delta_{AS}$ . Numerical examples are presented to illustrate the magnitude of the corresponding biases under realistic scenarios.

**Keywords.** Causality, Collider bias, Obesity paradox

Figure 1: Le modèle causal considéré par Sperrin et al. (2016)



## 1 Introduction

Le paradoxe de l'obésité a été rapporté dans plusieurs études observationnelles, où il était observé une association négative entre obésité et mortalité précoce chez des patients souffrant d'une maladie chronique, telle que le diabète ou l'insuffisance cardiaque. Des explications biologiques ont été avancées, avant que des arguments causaux ne soient récemment fournis pour élucider ce paradoxe : la maladie chronique étant sur un chemin entre l'obésité et la mortalité précoce, elle agit comme un médiateur. Or le conditionnement sur un médiateur crée généralement un biais, dit de collision. Ainsi, l'effet protecteur observé de l'obésité parmi les patients diabétiques, par exemple, n'a pas de valeur causale.

Dans un article récemment publié dans la revue *Epidemiology*, Sperrin et al. (2016) relancent le débat sur la véracité du paradoxe. Ces auteurs ont tenté d'évaluer la magnitude du biais de collision sous des scénarios raisonnables. D'après leurs analyses, celle-ci est typiquement faible et il est donc peu probable que le biais de collision puisse expliquer à lui seul le paradoxe de l'obésité.

Cependant, une erreur sur l'expression du biais s'est glissée dans l'analyse de Sperrin et al., ce qui remet grandement en cause leurs conclusions. Les outils développés récemment autour de l'inférence causale reposent pour l'essentiel sur des résultats probabilistes relativement simples, quoique parfois contre-intuitifs lorsqu'ils sont appliqués sur les variables contrefactuelles. L'erreur commise par Sperrin et al. en est un bel exemple et la description détaillée de cette erreur, outre son intérêt pour le paradoxe de l'obésité, permet d'illustrer ces "pièges" sur un exemple simple.

## 2 Quantités d'intérêt dans le paradoxe de l'obésité

Sperrin et al. fondent leur arguments sur une forme simple du modèle causal décrivant le paradoxe de l'obésité ; voir la figure 1. Les quatre variables  $A$ ,  $M$ ,  $Y$  et  $U$  sont supposées binaires, et correspondent à l'obésité, au diabète, à la mortalité précoce et à un facteur de confusion binaire (qui peut ne pas être observé). Les variables  $\xi_X$ , pour  $X \in \{A, Y, M, U\}$ , correspondent aux perturbations associées à chacune des variables  $A, Y, M, U$ . Ces perturbations sont supposées mutuellement indépendantes. Pour simplifier l'exposé, nous considérons dans un premier temps des effets causaux, et associations, mesurés sur l'échelle additive. En particulier, l'association entre  $A$  et  $Y$  sachant que  $M = 1$  est

$$\Delta_{AS} = E[Y|M = 1, A = 1] - E[Y|M = 1, A = 0].$$

Cette quantité peut être estimée à partir d'un échantillon de patients diabétiques, même si  $U$  n'est pas observée. Comme indiqué ci-dessus, plusieurs études ont rapporté une estimation négative pour  $\Delta_{AS}$ , conduisant au paradoxe. Cependant,  $\Delta_{AS}$  doit pouvoir être relié à un effet causal pour que le paradoxe soit réel. Parce qu'on se concentre sur la population des patients diabétiques, *i.e.* tels que  $M = 1$ , Sperrin et al. ont alors tenté de mesurer le biais entre  $\Delta_{AS}$  et l'effet causal de  $A$  sur  $Y$ , conditionné sur l'évènement  $M = 1$ , qu'ils définissent comme suit

$$\Delta_{CE} = E[Y^{A=1}|M = 1] - E[Y^{A=0}|M = 1].$$

Ici,  $Y^{A=1}$  et  $Y^{A=0}$  sont deux variables contrefactuelles, ou résultats potentiels, que l'on aurait pu observer dans les mondes contrefactuels  $\Omega^{A=1}$  et  $\Omega^{A=0}$  qui auraient suivi les interventions  $do(A = 1)$  et  $do(A = 0)$ , respectivement ; voir Pearl (2000).

Sperrin et al. tentent alors d'obtenir une expression de  $\Delta_{CE}$  qui ne dépende que des distributions des variables observées. Ils établissent alors, de manière erronée, que  $\Delta_{CE}$  vaut

$$\begin{aligned} \Delta_{Sp} = & \sum_u \{E[Y|M = m, A = 1, U = u] \\ & - E[Y|M = m, A = 0, U = u]\} P[U = u|M = m], \end{aligned}$$

en prétendant que, pour  $a \in \{0, 1\}$ ,

$$\begin{aligned} E[Y^{A=a}|M = m] &= \sum_u E[Y^{A=a}|M = m, U = u] P[U = u|M = m] \\ &= \sum_u E[Y|M = m, A = a, U = u] P[U = u|M = m]. \end{aligned} \quad (1)$$

Or l'égalité (1) n'est garantie que si  $E[Y^{A=a}|M = m, U = u] = E[Y|M = m, A = a, U = u]$ , qui n'est elle-même garantie que si  $Y^{A=a} \perp\!\!\!\perp A \mid (M, U)$ . Cette indépendance

conditionnelle n'est généralement pas vérifiée sous le modèle de la figure 1, précisément à cause du biais de collision :  $M$  étant sur un chemin causal entre  $A$  et  $Y$ , l'ensemble  $\{M, U\}$  ne satisfait pas le critère back-door; voir Pearl (2000). Le conditionnement sur  $M$  crée une association entre les perturbations  $\xi_A$  et  $\xi_M$ , et par suite entre les perturbations  $\xi_A$  et  $\xi_Y$ , générant le biais de collision.

Ainsi,  $\Delta_{Sp}$  est généralement différent de  $\Delta_{CE}$  sous le modèle de la figure 1, et la raison de cette différence est précisément le biais de collision, qui fait que  $E[Y^{A=a}|M = m, U = u]$  et  $E[Y|M = m, A = a, U = u]$  sont typiquement différentes.

En fait la quantité  $\Delta_{CE}$  n'est pas identifiable sous le modèle considéré par Sperrin et al., au sens où elle ne peut pas être exprimée en fonction de la distribution des variables  $(A, M, U, Y)$  sans hypothèse supplémentaire sur le modèle causal. En spécifiant la forme des fonctions structurelles, ou génératrices, ainsi que la distribution des perturbations dans ce modèle, on peut néanmoins obtenir une formule analytique pour  $\Delta_{CE}$  et donc évaluer le biais  $\Delta_{AS} - \Delta_{CE}$ . En reprenant les analyses numériques menées par Sperrin et al., mais avec la bonne formule pour le biais  $\Delta_{AS} - \Delta_{CE}$ , la figure 2 établit que ce biais peut-être sévère, ce qui contredit les conclusions de Sperrin et al.

D'autre part, la pertinence de l'effet causal  $\Delta_{CE}$  peut être remise en cause dans le contexte présent, où  $M$  agit comme un médiateur de l'effet de  $A$  sur  $Y$ . En particulier, l'effet direct contrôlé de  $A$  sur  $Y$  en  $M = 1$ ,  $\Delta_{CDE} = E[Y^{A=1, M=1}] - E[Y^{A=0, M=1}]$  peut être plus pertinent; voir VanderWeele et Vansteelandt (2010). Sous le modèle de la figure 1, il vient

$$\begin{aligned} \Delta_{CDE} = & \sum_u \{E[Y|A = 1, M = 1, U = u] \\ & - E[Y|A = 0, M = 1, U = u]\}P(U = u). \end{aligned}$$

On montre alors facilement que la différence entre  $\Delta_{AS}$  et  $\Delta_{CDE}$  est seulement due à la différence entre  $P[U = u]$  et  $P[U = u|M = m, A = a]$ . En particulier, en l'absence du facteur de confusion  $U$ ,  $\Delta_{AS}$  et  $\Delta_{CDE}$  sont égaux. Le biais entre  $\Delta_{AS}$  et  $\Delta_{CDE}$  peut donc s'interpréter comme un biais de confusion. Nos résultats numériques établissent qu'il est généralement limité sous le modèle considéré par Sperrin et al.; voir la figure 2. Il n'en reste pas moins que ce modèle n'exclut pas la possibilité d'une association négative  $\Delta_{AS} < 0$  conjointement à un effet direct contrôlé positif  $\Delta_{CDE} > 0$  (résultats non rapportés ici).

### 3 Illustrations numériques

Nous décrivons ici le modèle que nous avons considéré, qui est cohérent avec la figure 1 et qui correspond à une spécification du modèle considéré par Sperrin et al. Soit  $\mathbb{I}[\cdot]$  la fonction indicatrice. Les perturbations  $\xi_A$ ,  $\xi_U$ ,  $\xi_M$  et  $\xi_Y$  sont générées selon une loi uniforme sur  $[0, 1]$ . On définit les variables  $A = \mathbb{I}[\xi_A \leq 1/2]$  et  $U = \mathbb{I}[\xi_U \leq 1/2]$ , qui sont

donc deux variables indépendantes de loi  $B(1/2)$ . Introduisons maintenant la fonction sigmoïde  $\text{expit}(x) = (1 + \exp(-x))^{-1}$ , et posons, pour tout  $(a, u, m) \in \{0, 1\}^3$  et des paramètres réels  $\alpha_0, \alpha_A, \alpha_U, \alpha_{AU}, \beta_0, \beta_A, \beta_U$  et  $\beta_M$ ,

$$\begin{aligned} p_M(a, u) &= \text{expit}(\alpha_0 + \alpha_A a + \alpha_U u + \alpha_{AU} a u) \\ p_Y(a, m, u) &= \text{expit}(\beta_0 + \beta_A a + \beta_U u + \beta_M m). \end{aligned}$$

Finalement, nous définissons les variables  $M$  et  $Y$  par

$$\begin{aligned} M &= f_M(A, U, \xi_M) = \mathbb{1}[\xi_M \leq p_M(A, U)], \\ Y &= f_Y(A, M, U, \xi_Y) = \mathbb{1}[\xi_Y \leq p_Y(A, M, U)]. \end{aligned}$$

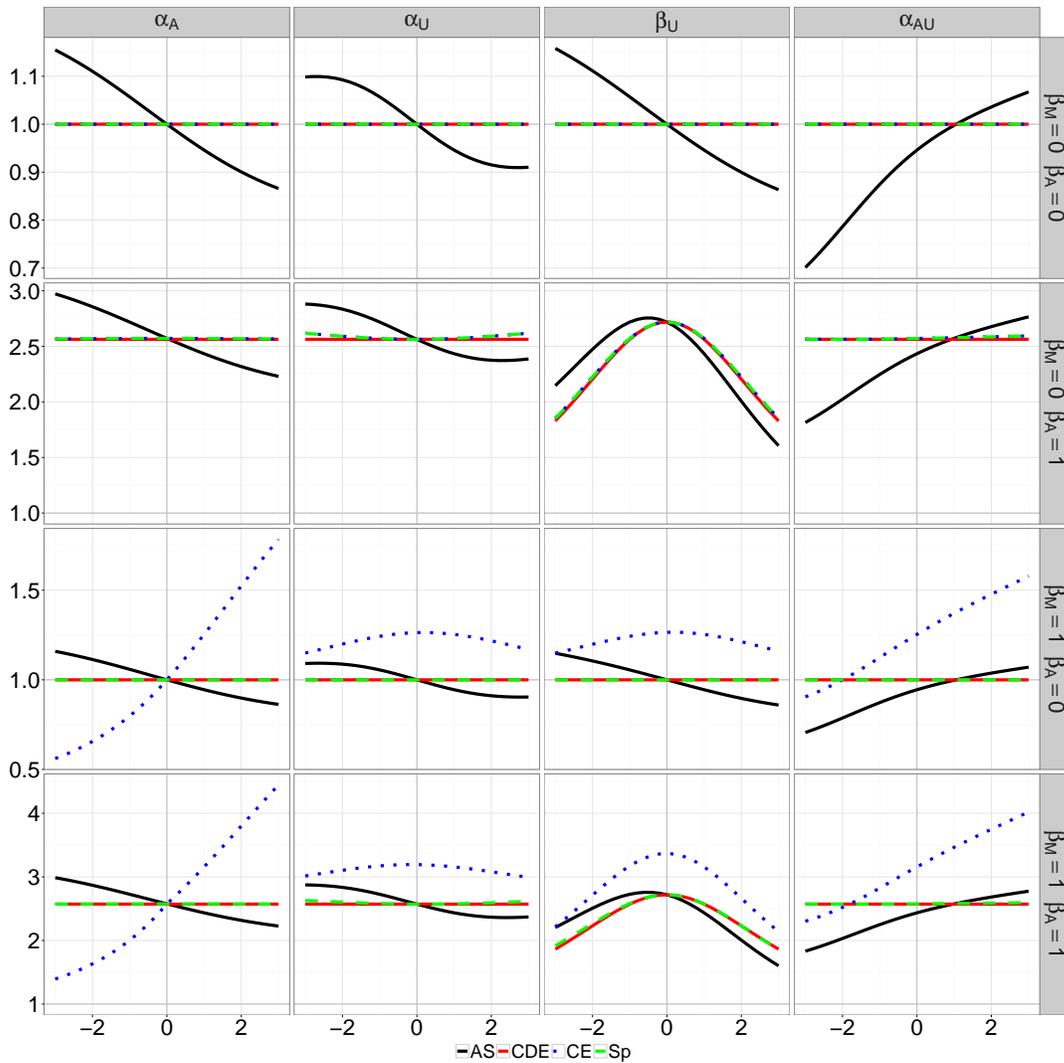
Pour toute combinaison des paramètres  $\alpha_0, \dots, \beta_M$ , du modèle, on peut aisément calculer  $\Delta_{AS}$ ,  $\Delta_{Sp}$  et  $\Delta_{CDE}$ . D'autre part, en posant  $x \wedge y = \min(x, y)$ , on peut montrer que

$$\begin{aligned} P(Y^{A=a} = 1 | M = 1) &= \frac{1}{4P(M = 1)} \sum_{\substack{i_A \in \{0,1\} \\ i_U \in \{0,1\}}} \left( p_Y(a, 1, i_U) [p_M(i_A, i_U) \wedge p_M(a, i_U)] \right. \\ &\quad \left. + p_Y(a, 0, i_U) \{p_M(i_A, i_U) - [p_M(i_A, i_U) \wedge p_M(a, i_U)]\} \right), \end{aligned}$$

et en déduire l'expression de  $\Delta_{CE}$ . La figure 2 illustre les différences entre  $\Delta_{AS}$ ,  $\Delta_{CE}$ ,  $\Delta_{Sp}$  et  $\Delta_{CDE}$  pour différentes combinaisons des paramètres, chacun de ces effets étant mesurés sur l'échelle de l'odds-ratio ici. Les deux premières lignes correspondent aux résultats présentés par Sperrin et al. : lorsque  $\beta_M = 0$  (*i.e.* lorsque  $M$  n'est pas un médiateur), le biais de collision disparaît,  $\Delta_{CE} = \Delta_{Sp} = \Delta_{CDE}$  et le biais entre  $\Delta_{AS}$  et  $\Delta_{CE}$  se réduit à un biais de confusion, typiquement faible. Cependant, et contrairement à ce que Sperrin et al. prétendent, la situation où  $\beta_M \neq 0$  est bien différente, puisque le biais de collision est alors actif, quelque soit la valeur de  $\beta_A$ . En particulier, les deux dernières lignes de la figure 2 montrent que le biais entre  $\Delta_{AS}$  et  $\Delta_{CE}$  peut être important et qu'il est tout à fait possible d'avoir  $\Delta_{AS} < 0$  (association négative entre obésité et mortalité chez les patients diabétiques) et  $\Delta_{CE} > 0$  (effet causal positif de l'obésité sur la mortalité, chez ces mêmes patients). La comparaison entre  $\Delta_{AS}$ , et  $\Delta_{CDE}$  montre que le biais entre ces deux quantités est typiquement faible, et que l'association observée chez les patients diabétiques est fournit une approximation sans doute légèrement biaisée seulement de l'effet direct contrôlé.

Pour résumer, et contrairement aux conclusions de Sperrin et al., le biais de collision peut induire une différence sensible entre  $\Delta_{AS}$  et  $\Delta_{CE}$ , et peut expliquer à lui seul le paradoxe de l'obésité. Le biais entre  $\Delta_{AS}$  et  $\Delta_{CDE}$  est quant à lui uniquement dû au facteur de confusion  $U$  et est typiquement beaucoup plus faible.

Figure 2: Résultats de notre étude numérique. Les odds-ratios causaux et observationnels sont présentés pour  $(\beta_A = 0, \beta_M = 0)$ ,  $(\beta_A = 1, \beta_M = 0)$ ,  $(\beta_A = 0, \beta_M = 1)$ , et  $(\beta_A = 1, \beta_M = 1)$ , et en faisant varier les autres paramètres  $\alpha_A, \alpha_U, \beta_U$  and  $\alpha_{AU}$ . Dans chaque colonne, sur l'axe des abscisses, un des paramètres varie entre  $-3$  to  $3$  (colonne de gauche:  $\alpha_A$ , colonne suivante :  $\alpha_U$ , colonne suivante :  $\beta_U$ , et colonne de droite :  $\alpha_{AU}$ ), et les autres paramètres sont fixés à une valeur par défaut (1 pour  $\alpha_A, \alpha_U, \beta_U$  et 0 pour  $\alpha_{AU}$ ).



## Bibliographie

- [1] Pearl, J. (2000), *Causality: models, reasoning, and inference*, Cambridge University Press, New York.
- [2] Sperrin, M. Clandlish, J. et Badrick, E. (2016), Collider bias is only a partial explanation for the obesity paradox, *Epidemiology*, 27, 525–30
- [3] VanderWeele, T. J. et Vansteelandt, S. (2010), Odds Ratios for Mediation Analysis for a Dichotomous Outcome, *American Journal of Epidemiology*, 172, 1339–48.