

N° d'ordre NNT : 2017LYSE1244

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de

l'Université Claude Bernard Lyon 1

École Doctorale ED205

Ecole Doctorale Interdisciplinaire Sciences-Santé

Spécialité de doctorat : Biostatistique

Discipline : Epidémiologie

Soutenue publiquement le 01/12/2017, par :

Marine Dufournet

Quantification du biais de sélection en sécurité routière : apport de l'inférence causale

---

Devant le jury composé de :

Barbara Charbotel, Professeure (HDR), PUPH, Univ. Lyon	Présidente
Romain Pirrachio, Professeur (HDR), PUPH, Univ. Paris Descartes	Rapporteur
Grégoire Rey, Ingénieur de Recherche (HDR), CépiDc-INSERM, Univ. Paris-Sud	Rapporteur
Louis-Rachid Salmi, Professeur (HDR), PUPH, ISPED, Univ. Bordeaux	Rapporteur
Emilie Lanoy, Chercheure (HDR), INSERM CESP Oncostat, Paris Saclay	Examinatrice
Virginie Rondeau, Directrice de recherche (HDR), ISPED, Univ. Bordeaux	Examinatrice
Adeline Samson, Professeure des universités (HDR), Univ. Grenoble	Examinatrice
Vivian Viallon, Maître de conférences (HDR), IFSTTAR, Univ. Lyon	Directeur de thèse
Jean-Louis Martin, Chargé de recherche, IFSTTAR, Univ. Lyon	Encadrant

.

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Professeur Frédéric FLEURY**
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLÉE
Directrice Générale des Services	Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur G.RODE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. X. PERROT
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E.PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

.

Remerciements

Je tiens à remercier chaleureusement Vivian Viallon pour avoir dirigé mes recherches, m'avoir fait confiance, et avoir démultiplié mes connaissances et mes compétences durant ces trois ans. Ton intérêt et ton enthousiasme pour ces travaux ont été une véritable source de motivation. Un grand merci également pour tes encouragements durant la phase finale de rédaction.

J'aimerais également remercier Jean-Louis Martin pour ses conseils avisés, et pour avoir introduit une touche d'humour et de sagesse dans ce trio de travail.

Je remercie les membres du jury. Merci à mes rapporteurs Romain Pirrachio, Grégoire Rey et Louis-Rachid Salmi d'avoir accepté d'évaluer mon travail. Je remercie particulièrement Emilie Lanoy et Barbara Charbotel, en tant qu'examinatrices et membres de mon comité de suivi de thèse. Vos discussions et encouragements ont été une source à la fois de réconfort et de motivation. Merci également à Adeline Samson et Virginie Rondeau d'avoir accepté d'être examinatrice.

Je tiens à remercier toute l'équipe de l'UMRESTTE pour son accueil et son soutien tout au long de cette thèse. Merci Marie, ma fidèle acolyte du yoga, pour toutes les discussions intéressantes que j'ai pu avoir avec toi. Merci Marie-Pierre pour ton soutien si précieux. Un grand merci, Amandine, pour ton dynamisme et ton esprit de solidarité. Merci Dan, Aisha, Nassur, Anne-Marie, Sylviane, Amina, Emmanuèle, Hélène, Blandine, Anne-Sophie, Gilles, Martine...
Merci aux doctorants avec qui j'ai partagé des parties de tarot, un match de ping-pong ou une session de badminton : Anais, Setareh, Antoine, Simon, Rémi, Mathias, Cédric, Clémence, Nadim, Maxime, Grégory, Julien... et j'en oublie sûrement.

J'aimerais remercier Jean Lobry et Christelle Lopez de m'avoir accueilli au sein de l'équipe pédagogique du LBBE. Cette expérience d'enseignement a été une découverte pleine de succès. Merci à mes étudiants, notamment de L1, qui m'ont motivé, et souvent fait rire.

J'adresse un immense merci à mes amis. Merci d'être là pour partager un verre sur les quais, un molki place voltaire, des gambas au grum & gram, de belles randos, des tours de vélo à Miribel, des cours de dessin, un moment de danse improvisé, des voyages aux quatre coins du monde, ou tout simplement un coup de fil qui fait du bien. Bref, merci à toi, Daniele, Antoine, Laurine, Brice, Pauline, Pierre, Loïc, Jochen, Chloé, Francis, Mélanie, Stéphane, Elodie, Amandine, Nathalie, Clara, Océane, Nicolas, Laura, Clélie, Jeanne, Adrienne, Mathilde, les Vincent, Simon, Camille, Lucille, Martin, Matthieu, Nicolas, Solène, Laurent, Elisabeth, Nadège, Thomas, Samuel, Damien, François, Nidal, Yves...

Je remercie enfin ma famille, et en particulier mes parents pour leur soutien sans faille dans les périodes les plus éprouvantes. Merci à mon frère et sa compagne, mes tantes et oncles, ma cousine, et ma marraine. Merci à ma nièce et mon neveu d'avoir apporté une dose d'insouciance. J'ai une dernière pensée pour mes grands-parents, qui auraient été très fiers de leur petite fille.

Mille mercis donc à ceux qui ont compté, qui comptent aujourd'hui, et qui compteront toujours.

Résumé

Les principaux facteurs de l'insécurité routière sont connus, et l'enjeu réside aujourd'hui dans la mesure de l'effet d'un facteur, et la hiérarchisation de l'ensemble des causes intervenant dans la survenue de l'accident. Toutefois, les données disponibles concernent généralement que des accidentés. En l'absence de non-accidentés, l'épidémiologiste du risque routier se heurte à une sélection extrême. Une des solutions classiques est d'utiliser des analyses en responsabilité, et de mesurer l'effet causal d'un facteur sur le risque d'être responsable d'un accident. Néanmoins, la validité des analyses en responsabilité repose sur l'hypothèse, discutable, que les non-responsables sont représentatifs des circulants. L'objectif de cette thèse est donc de déterminer si les données disponibles d'accidentés permettent de fournir, via les analyses en responsabilité, des estimations des effets causaux sans biais, et notamment sans un biais de sélection résiduel.

Nous montrons dans cette thèse que, dès lors que l'inclusion dépend de la gravité de l'accident, et que le facteur étudié a un impact sur la vitesse, il est impossible d'estimer l'effet causal du facteur sur le risque d'être responsable de l'accident grave sans un biais de sélection résiduel. Ce résultat est tout d'abord démontré de manière formelle, grâce à l'utilisation des modèles causaux structuraux. Ces modèles sont fondés sur une structure graphique, le DAG, qui représente les différentes relations entre les variables. Ce DAG permet la description des variables réellement observées, mais également des variables contrefactuelles, variables observables dans un monde contrefactuel où l'on aurait fixé l'exposition à une certaine valeur. L'effet causal étant défini à partir de ces variables contrefactuelles partiellement observées, c'est la structure du DAG qui permet de déterminer si l'effet causal peut être estimé en fonction des variables observées. Or, la structure du DAG conduisant à la survenue d'un accident grave ne permet pas d'exprimer l'effet causal du facteur étudié sur la responsabilité de l'accident grave en fonction des distributions observées sur les accidentés graves. Conditionner les estimations sur les accidentés graves correspond à ajuster sur une variable du DAG appelée « collider », et ainsi à introduire un biais dit de collision. En générant un modèle relativement simple, nous donnons à nos résultats théoriques une illustration numérique. En effet, lorsque les données ne dépendent pas de la gravité de l'accident, ou que le facteur étudié n'a pas d'effet sur la vitesse, la mesure estimable à partir des analyses en responsabilité est une mesure sans biais de l'effet causal, sous certaines hypothèses de prévalences faibles. Lorsque l'inclusion dépend de la gravité de l'accident, il existe un biais et ce biais induit par les analyses en responsabilité est d'autant plus grand que l'intensité de la relation entre le facteur et la vitesse, et celle entre la vitesse et l'accident est grand. Les schémas d'étude présentés permettent d'approcher des situations où le facteur étudié serait l'alcool ou le cannabis. Dans le cas de l'alcool, il apparait que sous le modèle simple considéré, la mesure d'association estimable serait une sous-estimation de l'effet causal. En revanche, dans le cas du cannabis, la mesure d'association correspondrait à une sur-estimation de l'effet causal.

D'autre part, les outils de l'inférence causale nous ont permis de fournir une description formelle de la validité externe et interne, ainsi qu'une description formelle de la mesure d'association estimable via les analyses en responsabilité. Cette question de la validité interne d'une mesure se pose dans d'autres champs d'application que la sécurité routière. Elle se pose notamment dans le cas du paradoxe de l'obésité. Suite à des travaux d'autres auteurs, et contrairement à ce que ces auteurs prétendaient, nous avons montré que le conditionnement sur des malades chroniques, et donc le biais de collision, suffisait à expliquer à lui seul le paradoxe de l'obésité.

Abstract

Many factors associated with the risk and severity of road accidents are now widely considered as causal : alcohol, speed, usage of a mobile phone... Therefore, questions asked by decision-makers now mostly concern the magnitude of their causal effects, as well as the burden of deaths or victims attributable to these various causes of accident. One particularity of road safety epidemiology is that available data generally describe drivers and vehicles involved in road accidents only, or even severe road accidents only. This extreme selection precludes the estimation of causal effects. To circumvent this absence of « control » population of non-crash involved drivers, it is common to use responsibility analysis and to assess the causal effect of a given factor on the risk of being responsible for an accident among involved drivers. The underlying assumption is that non-responsible drivers represent a random sample of the general driving population that was « selected » to crash by circumstances beyond their control and therefore have the same risk factor profile as other drivers on the road at the same time. However, this randomness assumption is questionable. The objective of this thesis is to determine whether available data in road safety allow us to assess causal effects on responsibility without a residual selection bias.

We show that a good approximation of causal effect of a given factor on the risk of being responsible is possible only if the inclusion into the dataset does not depend on the severity of the accident, or if the given factor has no effect on speed. This result is shown by using the Structural Causal Model (SCM) framework. The SCM framework is based on a causal graph : the DAG (directed acyclic graph), which represents the relationships among variables. The DAG allows the description of what we observe in the actual world, but also what we would have observed in counterfactual worlds, if we could have intervened and forced the exposure to be set to a given level. Causal effects are then defined by using counterfactual variables, and it is the DAG's structure which determines whether causal effects are identifiable, or recoverable, and estimable from the distribution of observed variables. However, the assumptions embedded in the DAG which describes the occurence of a severe accident does not ensure that a causal odds ratios is expressible in terms of the observable distribution. Conditioning the estimations on involved drivers in a severe crash correspond to conditioning on a variable in the DAG called « collider », and to create a « collider bias ». We present numerical results to illustrate our theoretical arguments and the magnitude of the bias between the estimable association measure and some causal effects. Under the simple generative model considered, we show that, when the inclusion depends on the severity of the accident, the bias between the estimable association measure and causal effect is larger thn the relation between the exposure and speed, or speed and the occurrence of a severe accident is strong. Moreover, the presented designs allow us to describe some situations where the exposure could be alcohol or cannabis intoxication. In the case of alcohol, where alcohol and speed are positively correlated, the estimable associational effect underestimates the causal effect. In the case of cannabis, where cannabis and speed are negatively correlated, the estimable associational effect overestimates the causal effect.

On the other hand, we provide a formal definition of internal and external validity, and a counterfactual interpretation of the estimable quantity in the presence of selection bias, when causal effects are not recoverable. This formal interpretation of the estimable quantity in the presence of selection bias is not only useful in the context of responsibility analyses. It is for instance useful to explain the obesity paradox.

Table des matières

Introduction générale **1**
L'accident de la circulation, un enjeu de santé publique 1
Les facteurs favorisant la survenue de l'accident, et sa dimension multifactorielle 2
Données en sécurité routière : face au biais de sélection extrême, l'analyse en responsabilité 5
Objectifs des travaux de thèse 7

1 Causalité et modèles causaux structuraux **9**
1.1 De l'association à la causalité : approches philosophiques et scientifiques 10
1.1.1 Les grandes conceptions philosophiques de la causalité 10
1.1.2 Etablir la causalité en épidémiologie et biostatistique, les approches classiques 13
1.2 Approche formalisée de la causalité : introduction aux modèles causaux structuraux 16
1.2.1 DAG, réseau d'indépendances marginales et conditionnelles, loi jointe . . . 17
1.2.2 Equations structurelles, et définition de quantités causales 19
1.2.3 Identifiabilité d'une quantité causale dans les modèles markoviens 21
1.2.4 Identifiabilité d'une quantité causale dans les modèles non markoviens . . 24
1.2.5 Estimation d'une quantité causale 26
1.3 Analyse causale en médiation : décomposition d'un effet total 29
1.3.1 Approches classiques de la médiation 29
1.3.2 Définition des effets totaux, controlés, directs et indirects 31
1.3.3 Identifiabilité et estimation des effets directs et indirects 33

2 Formalisation des analyses en responsabilité **37**
2.1 Inférence causale en présence de sélection 38
2.1.1 Identifiabilité de $\mathbb{P}(Y_x = 1)$ en présence de sélection 40
2.1.2 Identifiabilité de l'odds ratio causal en présence de biais de sélection . . . 41
2.2 Application aux analyses en responsabilité 42
2.2.1 Description formelle de la notion de responsabilité 43
2.2.2 Application à l'analyse en responsabilité lorsque l'inclusion ne dépend pas de la gravité de l'accident 45
2.2.3 Application à l'analyse en responsabilité lorsque l'inclusion dépend de la gravité de l'accident 47
2.2.4 Validité externe et interne des mesures estimables 48

3 Sens et quantification du biais dans les analyses en responsabilité 55
3.1 Illustration numérique lorsque X est binaire . . . 56
3.1.1 Modèle causal considéré . . . 56
3.1.2 Résultats avec facteurs de confusion . . . 58
3.1.3 Résultats en présence d'un facteur de confusion W et d'interaction . . . 63
3.2 Illustration lorsque X est catégorielle . . . 69
3.2.1 Modèle causal considéré . . . 70
3.2.2 Résultats en l'absence de facteurs de confusion . . . 71
3.2.3 Résultats avec facteurs de confusion . . . 71

4 Biais de collision et paradoxe de l'obésité 75
4.1 Considérations théoriques . . . 76
4.1.1 Présentation du modèle causal . . . 76
4.1.2 Différence entre la mesure d'association et l'effet défini par Sperrin (2016) 79
4.1.3 Autres quantités d'intérêt : des effets directs contrôlés . . . 82
4.2 Illustrations numériques . . . 85
4.2.1 Mécanisme de génération des données . . . 85
4.2.2 Illustration de la différence entre $\mathbb{P}(Y_x = 1|M_x = 1)$ et $\mathbb{P}(Y_x = 1|M = 1)$. 86
4.2.3 Expression analytique de la quantité causale sous le mécanisme de génération de données considéré . . . 87
4.2.4 Résultats . . . 88

Discussion générale 93

Bibliographie 98

Annexes 107
.1 Preuve que $\mathcal{R}_x$ n'est pas indépendant de X sachant W et $\mathcal{S}$ dans le cas où seuls les accidents sévères sont disponibles . . . 107
.2 Illustration numérique lorsque X est catégorielle . . . 108
.3 Article soumis dans Journal of causal inference . . . 111
.4 Poster présenté à UK-Causal inference meeting . . . 138
.5 Lettre de recherche publiée dans Epidemiology . . . 140
.6 Réponse de Sperrin et al. . . . 144

Introduction générale

L'accident de la circulation, un enjeu de santé publique

En 2013, le nombre annuel de tués sur le réseau routier mondial a été estimé à 1,25 million (WHO, 2015), faisant de l'accident de la circulation routière la huitième cause de décès à l'échelle mondiale. Sur le territoire français, 3 461 personnes ont perdu la vie en 2015 (ONISR, 2015). Au nombre de décès s'ajoute un nombre de blessés : l'ONISR en décomptait plus de 70 000 en 2015. Cette statistique est toutefois à interpréter avec prudence car les blessés sont généralement sous-enregistrés. Le nombre de blessés toutes gravités confondues serait en réalité trois fois plus élevé (Amoros et al., 2008). D'autre part, les accidents de la circulation posent une question de coût pour l'ensemble de la société. En effet, derrière les blessés se profilent des coûts de prise en charge (hospitalisation, rééducation, prise en charge du handicap...), et des coûts de productivité (perte de productivité pour la victime mais également ses aidants...). La réduction du nombre d'accidents de la route constitue ainsi un enjeu majeur de santé publique. D'autant plus que les accidents de la route sont la première cause de mortalité chez les jeunes de 18-24 ans (WHO, 2015). Depuis 2010, les engagements pris en matière de sécurité routière lors de la « Déclaration de Moscou » sont d'ailleurs devenus résolution de l'ONU (ONU, 2009).

Les éléments de cadrage présentés font référence aux accidents de la route corporels. Selon la définition officielle, un accident corporel de la circulation routière correspond à un accident provoquant au moins une victime survenant sur une voie ouverte à la circulation publique et impliquant au moins un véhicule en mouvement. Un accident de la route peut toutefois être matériel, s'il n'inclut pas de victime. Dans le cadre de cette thèse, nous nous concentrerons sur les accidents corporels.

Afin de lutter contre l'insécurité routière, deux voies de recherche sont étudiées (Haddon Jr, 1972). La recherche en sécurité routière primaire consiste à réduire le nombre d'accidents corporels. Tandis que la recherche en sécurité secondaire consiste à réduire la gravité des accidents de la circulation, notamment en concevant des véhicules réduisant l'effet d'un choc (structure déformable, airbag, prétensionneur de ceinture de sécurité...). C'est sur l'axe de la sécurité primaire que s'inscrit ce projet doctoral. En effet, il s'agit de comprendre les mécanismes conduisant à la survenue de l'accident, en particulier en identifiant les facteurs intervenant dans la survenue de l'événement, puis à mesurer l'effet d'un facteur sur le risque d'accident. Dans ce sens, l'approche épidémiologique s'attache à déterminer les facteurs statistiquement associés à l'apparition de l'événement d'intérêt et mesurer la force de l'association entre un facteur donné et l'événement, ici l'accident.

Les facteurs favorisant la survenue de l'accident, et sa dimension multifactorielle

La plupart des facteurs favorisant la survenue de l'accident sont connus (alcool, vitesse, type de véhicule...) et font intervenir différentes parties : l'usager, le véhicule, l'infrastructure et plus largement l'environnement (INRETS, 2005). La matrice de Haddon (Haddon Jr, 1972) permet de référencer les différents facteurs selon la chronologie de l'événement et chacune des parties évoquées. Le tableau 1 donne un aperçu de facteurs intervenant fréquemment, classés selon la matrice de Haddon. Pour chacune des parties, plusieurs facteurs peuvent intervenir et interagir, faisant de l'accident de la route, et de sa gravité, un phénomène multifactoriel.

TABLE 1 – Exemples de facteurs de risque associés à chacune des cellules de la matrice de Haddon

	Usager	Véhicule	Environnement
Avant la collision (sécurité primaire)	Alcool, cannabis Vitesse, excès de vitesse Fatigue Usage du téléphone Temps de réaction	Usure des pneus/freins Freins ABS Système antipatinage Puissance du moteur Stabilité dans les virages	Etat et profil de la chaussée Météo Visibilité Densité du trafic Type de route
Pendant la collision (sécurité secondaire)	Age Sexe Port de la ceinture de sécurité Place dans le véhicule	Habitacle sécuritaire Ceinture, Airbag Poids et taille du véhicule	Présence de garde fou Objets fixes (arbres, lampadaires)
Après la collision (sécurité tertiaire)	Age Etat de santé Condition physique		Rapidité des secours Services pré-hospitaliers Système de soins adéquat

Cette thèse se place dans l'objectif de la quantification des effets des principaux facteurs intervenant dans la survenue de l'accident et de leur hiérarchisation. Il s'agit de se focaliser ainsi sur les facteurs intervenant avant la collision. Parmi ces facteurs, nous nous concentrons sur des facteurs comportementaux fréquemment rencontrés, à savoir la vitesse, la consommation d'alcool ou celle de cannabis.

La vitesse et le non-respect des limitations

La vitesse a un impact sur les performances du conducteur. Elle peut avoir un effet sur la survenue de l'accident, avec la difficulté accrue pour le conducteur de s'adapter aux situations rencontrées. Elle peut également provoquer plus facilement une perte de contrôle et augmenter la distance de freinage. La vitesse a également un effet sur des conséquences corporelles : plus la vitesse est élevée, plus l'accident est grave (ONISR, 2015). Ainsi, augmenter la vitesse moyenne de 10 % sur un réseau routier quel qu'il soit conduirait à augmenter le risque d'accident matériel de 10 %, le risque d'accident corporel de 20 %, et le risque d'accident mortel de 40 % (Nilsson, 2004, Elvik et al., 2004, Elvik, 2009).

S'agissant du non-respect des limitations de vitesse, selon les forces de l'ordre (ONISR, 2014), conduire à une vitesse excessive ou inadaptée à la situation serait la cause principale identifiée dans près de 26 % des accidents mortels. Des travaux de recherche ont également conclu que la

proportion d'accidents mortels attribuables à un dépassement des vitesses maximales autorisées sur les routes départementales était estimée en 2010 à 28 % (Viallon and Laumon, 2013). Cette proportion était de 45 % avant la mise en place des radars automatiques en 2001.
Toutefois, les petits excès de vitesse représentent la majorité des excès, quel que soit le réseau considéré. Par exemple, en 2009, l'OSNIR avait observé 24 % d'excès de vitesse sur les routes nationales et départementales limitées à 90 km/h. Ce taux passe à 7 % lorsqu'on ne considère que les excès de vitesse supérieurs à 10 km/h.

L'alcool
Sur le rôle de l'alcool dans la survenue de l'accident, les recherches bibliographiques montrent que l'absorption d'alcool à faible dose ($\leq$ 0,4 g/l) affecte essentiellement les processus attentionnels alors que les doses supérieures ($\geq$ 0,8 g/l) dégradent significativement le processus d'inhibition de la réponse et les capacités psychomotrices (Mignot and Mizzi, 2011). L'alcool reste en effet la substance la plus fréquemment retrouvée chez les conducteurs interpellés par les forces de l'ordre (Jones et al., 2007), les conducteurs décédés (Schwilke et al., 2006), les victimes de traumatisme routier (Michael Walsh et al., 2005), et les usagers ayant subi des contrôles routiers (Steinmeyer et al., 2001). Il reste toutefois difficile d'estimer la proportion de conducteurs circulant en étant alcoolisés. En revanche, il est connu que les conducteurs les plus concernés par une conduite sous l'emprise de l'alcool sont les hommes, surtout jeunes. Notons que l'alcoolémie au volant concerne toutes les générations hormis les plus âgés. L'alcoolémie est aussi fortement liée au motif de déplacement. Elle est plus fréquente lors des déplacements privés que lors des trajets domicile-travail ou des déplacements professionnels. Les conducteurs de deux-roues motorisés sont les usagers les plus fréquemment alcoolisés, les conducteurs de poids lourds le moins. Bien évidemment, l'alcoolémie au volant est davantage observée les nuits de week-end que les jours de semaine (Mignot and Mizzi, 2011).
La relation entre l'alcoolémie et la responsabilité a pu être observée sur les accidents de voiture et de piétons. L'étude SAM (Stupéfiants et Accidents Mortel) rapporte que le risque d'être responsable d'un accident mortel est multiplié par 8,5 [IC : 7,15 − 10,1] chez les conducteurs alcoolisés (Laumon et al., 2011). L'actualisation récente de l'étude SAM (ActuSAM) (Martin et al., 2017), l'odds ratio (OR) a été estimé à 16,2 [IC : 11,0 − 23,8]. Les auteurs justifient cet écart par la spécifité de la determination de la responsabilité dans ActuSAM, car les manœuvres inappropriées du conducteur ou les défaillances identifiées juste avant la collision sont plus systématiquement prises en compte qu'en employant la méthode de détermination utilisée dans SAM. Dans l'étude SAM, il est également remarqué que 60 % des accidents mortels impliquant un conducteur responsable alcoolisés, étaient des accidents à un seul véhicule sans piéton. Cela traduit le rôle significatif de l'alcool dans la réduction des capacités de contrôle du véhicule par le conducteur. Selon l'étude SAM, 31 % des accidents mortels seraient imputables à l'alcool. Comme le montre l'étude ActuSAM, cette proportion est constante dans le temps, alors que le nombre d'accidents a considérablement diminué.
S'agissant des faibles alcoolémies ($<$ 0,5 g/l), l'étude SAM n'a pas mis en évidence d'effet de celles-ci sur le risque d'être responsable. Par ailleurs, dans le cadre du projet européen DRUID (2008), aucune des trois études épidémiologiques n'a mis en évidence un sur-risque attaché aux faibles alcoolémie.

Le cannabis

Concernant le cannabis, sa consommation entraîne également des troubles de l'attention. L'étude SAM a montré que la conduite sous emprise de cannabis multipliait par 2 le risque d'être responsable d'un accident mortel. La mise en évidence d'un effet-dose a conforté l'hypothèse d'une relation causale. En 2011, 4 % des morts sur les routes françaises auraient été imputables à une conduite sous emprise du cannabis (Laumon et al., 2011). La moitié des victimes décédées avaient moins de 25 ans. Si, tous âges confondus, 3 % des conducteurs de l'étude étaient sous l'emprise du cannabis, les hommes de moins de 25 ans l'étaient bien davantage. Contrairement à l'alcool, la conduite sous cannabis n'est pas plus observée le week-end que la semaine, la nuit que le jour.

L'étude SAM a permis d'étudier l'effet de la combinaison alcool-cannabis sur la responsabilité de l'accident. Elle a montré que la conduite sous emprise à la fois d'alcool et de cannabis multipliait par 15 le risque d'être responsable. Ce risque est égal au produit des risques rattachés à l'alcool seul et au cannabis seul. Ainsi, l'enjeu majeur du cannabis au volant est son association avec l'alcool.

Quelques marqueurs de risque

Dans ce contexte multifactoriel où plusieurs éléments interviennent, la distinction entre un facteur causal, un facteur de risque ou encore un facteur intermédiaire n'est pas évidente.
Si tous les facteurs « causaux » sont des facteurs de risque, tous les facteurs de risque ne sont pas des facteurs « causaux ». Certaines caractéristiques de l'usager peuvent augmenter le risque d'accident, sans pour autant être qualifiées de causes. Ces caractéristiques définissent des populations à risque, dans lesquelles les véritables causes d'insécurité routière sont plus fréquentes. Le jeune homme, le conducteur âgé ou encore l'usager de deux roues motorisés correspondent par exemple à des populations à risque. Le jeune homme et l'usager de deux roues motorisés présente une plus grande inclinaison à la prise de risque, en termes notamment de consommation d'alcool et de cannabis pour le premier (OECD/ECMT Transport Research Centre et al., 2006), ou en termes de choix modal et de goût pour la vitesse pour le second (France et al., 2008). Chez le conducteur âgé, le vieillissement du système cognitif peut ralentir la vitesse du traitement de l'information routière, et le processus de décision en cas de danger (Vaa, 2003). Ces différentes catégories d'usagers sont alors plus exposées au risque d'accident. En présence d'autres facteurs « causaux », des causes potentielles peuvent également devenir des facteurs intermédiaires. C'est le cas par exemple de la vitesse, qui peut être à l'origine de l'accident, mais qui en présence de consommation d'alcool peut être vue comme un facteur intermédiaire (DRUID, 2008, Aarts and van Schagen, 2006).

Par ailleurs, certains facteurs de risque peuvent fausser l'association entre l'exposition et l'événement étudié. Ces facteurs sont couramment appelés « facteur de confusion » car ils sont à la fois liée à l'exposition étudiée et l'événement. Ces facteurs de confusion conduisent à biaiser les associations entre l'exposition et l'événement d'intérêt. Par exemple, l'âge peut être considéré comme un facteur de confusion puisqu'il influence à la fois la propension à consommer de l'alcool, et la probabilité d'avoir un accident en raison d'un manque d'expérience de conduite. La présence d'un facteur de confusion comme l'âge biaise l'estimation de l'effet causal de l'alcool sur la survenue de l'accident. En présence de facteurs de confusion, l'établissement d'une relation causale entre un facteur et un événement, mais également sa mesure, peut être complexe.

Données de sécurité routière : face au biais de sélection extrême, l'analyse en responsabilité

Afin d'identifier les déterminants d'un événement en épidémiologie, un des schémas d'étude le plus utilisé est l'étude de cohorte (Salmi et al., 2014). Néanmoins, une étude de cohorte est difficile à mettre en oeuvre étant donné que l'accident de la route est un événement rare. Dans ce cas, l'étude cas-témoins est plus adapté. Mais les données de sécurité routière ont la spécificité de ne concerner généralement que les accidentés, c'est-à-dire les conducteurs et usagers impliqués dans un accident mortel ou corporel. Il n'existe pas de groupe de référence constitué de non-accidentés, permettant ainsi une démarche comparative entre les cas (les accidentés) et les témoins (les non-accidentés). Au traditionnel biais de confusion rencontré généralement en épidémiologie s'ajoute en sécurité routière un biais de sélection extrême. Alors que le biais de confusion est lié à la prise en compte ou non de certaines variables, le biais de sélection est quant à lui défini comme l'inclusion préférentielle de certains individus de la population (Bareinboim and Pearl, 2012, Bareinboim and Tian, 2015).
Deux alternatives proches ont été proposées pour contourner ce problème d'absence de témoins : l'exposition quasi-induite (Stamatiadis and Deacon, 1997) et les analyses en responsabilité (Smith and Popham, 1951, Perchonok, 1978, Terhune, 1986, Brubacher et al., 2014). Ces analyses proposent de déterminer dans un premier temps la responsabilité ou non du conducteur dans l'accident, puis de comparer les responsables de l'accident versus les non-responsables de l'accident. La différence principale entre les deux approches réside dans la sélection de l'échantillon. Dans l'approche de l'exposition quasi-induite, seuls les accidents à deux véhicules et où la responsabilité de l'accident d'un des deux conducteurs est indiscutable sont sélectionnés. Les analyses en responsabilité concernent en revanche les accidents à un ou plusieurs véhicules, et des degrés de responsabilité sont distingués.

Ces approches posent naturellement la question de la détermination de la responsabilité. La responsabilité évoquée ici est différente de la conception juridique, légale et assurantielle de la responsabilité. Elle est rattachée ici à la notion de faute de conduite (dépassement de ligne blanche, non-respect des priorités, de la vitesse réglementaire). Un taux d'alcoolémie supérieur au seuil légal, même s'il constitue une infraction au code de la route, ne doit pas intervenir dans la détermination de la responsabilité. La détermination de la responsabilité peut alors s'effectuer de plusieurs façons. Une première façon est de se référer à la méthode proposée par Robertson and Drummer (1994). D'après cette méthode, la responsabilité ou « culpabilité » d'un conducteur est déterminée avec un score variant de 8 à 25. Plus le score est élevé, plus la responsabilité du conducteur dans l'accident est engagée. Au départ, le score est à son maximum, puis selon la situation rencontrée sur 8 critères « déresponsabilisant » (« mitigating factors »), le score est abaissé. Les 8 critères concernent notamment des conditions relatives à la route, au véhicule, aux conditions de conduite, au type d'infortune rencontrée, à la difficulté de la tâche de conduite à effectuer, à l'avis de témoins, à l'infraction au code de la route, au niveau de fatigue du conducteur. Cette méthode a l'avantage d'être facile à mettre en place, d'être applicable à n'importe quel accident, et d'avoir été utilisée par plusieurs équipes dans le monde. Elle présente cependant l'inconvénient de faire appel à des informations pas toujours mentionnées dans les procédures policières, et surtout de prendre en compte les infractions commises, et donc une éventuelle al-

coolémie positive, ce qui pose problème lorsque l'objectif est d'évaluer l'effet de l'alcool sur le risque d'être responsable.
D'autres façons sont également utilisées pour déterminer la responsabilité (Salmi et al., 2014). Elle est parfois évaluée par la police, par questionnaire auprès des conducteurs ou encore par des experts accidentologues. Ces derniers analysent le contenu de chaque procédure à la recherche d'informations permettant de comprendre la genèse de l'accident. Un tel procédé permet de faire abstraction de certains facteurs liés à l'objet de l'étude ou du fait d'un potentiel rôle de confusion (âge, sexe, etc.). Il présente toutefois l'inconvénient de ne pas bénéficier d'une validation externe (et donc internationale) et d'être lourd à mettre en oeuvre.

Une fois la responsabilité déterminée, il est ensuite possible de comparer les individus responsables versus les non-responsables de l'accident, c'est-à dire d'utiliser comme groupe témoin les conducteurs non responsables de leur accident. Ce choix repose sur l'hypothèse que seules des conditions fortuites, indépendantes de la nature de leur déplacement les ont conduits à être impliqués dans un accident (Davis and Gao, 1993, Lenguerrand et al., 2008, Stamatiadis and Deacon, 1997). Considérer les non-responsables comme des témoins, c'est faire l'hypothèse que les non-responsables accidentés sont représentatifs de la population des circulants. Ainsi, le risque estimé à travers les analyses en responsabilité serait similaire à celui estimé à partir d'études cas-témoins où l'on comparerait les responsables d'accident (les cas) aux non-impliqués (les témoins) (Brubacher et al., 2014, Wahlberg and Dorn, 2007). Même si certains auteurs ont questionné la validité de cette approche (Sanghavi, 2013), les analyses en responsabilité sont communément adoptées dans le champ de la sécurité routière et rarement remises en question (Asbridge et al., 2013, Salmi et al., 2014, Wahlberg, 2009). Néanmoins, la représentativité des non-responsables est questionnable. Sélectionnés au même moment que les cas, c'est-à-dire au moment de l'accident, il est possible que certaines caractéristiques des conducteurs non-responsables accidentés soient sous ou sur réprésentées par rapport à la population des circulants. Les auteurs de l'étude SAM avaient par exemple détecté une sur représentation des conducteurs positifs à l'alcool parmi les non-responsables décédés seuls dans leur accident (Laumon et al., 2005). En voulant solutionner le biais de sélection extrême rencontré en sécurité routière, et respecter les standards d'une étude cas-témoins, on peut donc se demander si les analyses en responsabilité ne laissent pas perdurer un biais de sélection résiduel.

Objectifs des travaux de thèse

Les principales causes des accidents de la circulation routière sont connues, et les questions qui restent à éclaircir concernent davantage la mesure des effets causaux. Combinés à la prévalence des facteurs dans la population générale, la quantification des effets permettrait d'estimer des risques attribuables qui sont un moyen de quantifier les enjeux relatifs aux différentes dimensions du risque routier. Ce travail de quantification des enjeux permettrait de hiérarchiser l'importance des différents facteurs identifiés et d'aider les décideurs à mettre en place des mesures adaptées. Le principe d'action est d'intervenir sur chaque facteur pour en diminuer l'occurrence, ôtant ainsi un « maillon » de la chaîne des événements qui aurait sinon abouti à l'accident.

L'objectif principal de cette thèse est de déterminer si les données disponibles d'accidentés permettent de fournir, via les analyses en responsabilité, des estimations de quantités causales sans biais, et notamment sans un biais de sélection résiduel.

Plusieurs objectifs secondaires seront abordés. Il s'agit de

- quantifier le biais de sélection lorsque l'estimation d'une quantité causale est compromise
- donner une description formelle de la mesure d'association estimable en présence de biais de sélection
- aborder d'autres problématiques où le biais de sélection entre en compte

Dans le premier chapitre, il s'agit de préciser la notion de causalité en statistique et d'introduire les principes généraux de la théorie de l'inférence causale. En effet, fondées sur des données dites « observationnelles », les mesures d'associations estimées en sécurité routière sont souvent entachées des biais traditionnellement rencontrés en épidémiologie : biais de classement ou de mesure, biais de confusion et biais de sélection. Depuis quelques décennies, la théorie de l'inférence causale (Robins et al., 2000, Pearl, 2000, 2009) s'est largement développée afin de pouvoir identifier et mesurer des relations causales sur la base de données observationnelles, et détecter formellement les biais de causalité (biais de confusion et biais de sélection). Il s'agit donc de présenter les fondements des modèles structuraux marginaux (Pearl, 2000, 2009) en vue de les appliquer aux analyses en responsabilité.

Le second chapitre présente une formalisation de l'analyse en responsabilité. Il s'appuie sur de récents résultats issus de la littérature sur l'identifiabilité de quantités causales en présence de biais de sélection. L'objectif de ce chapitre est de déterminer quels effets causaux sont identifiables et estimables à partir des analyses en responsabilité, selon les données d'accidentés disponibles. Nous considérons comme exemple l'estimation de odds ratio causal de l'alcool sur la responsabilité de l'accident dans deux cas. Le premier, lorsque tous les types d'accidents sont disponibles (matériels, corporels et mortels). Le second, qui est le plus fréquent en pratique, lorsque les données concernent uniquement les accidents graves (corporels ou mortels par exemple). Nous montrons que, dans le premier cas, l'odds ratio causal de l'alcool sur la responsabilité peut être approximé par un autre odds ratio causal identifiable, sous certaines hypothèses. Lorsque les données disponibles dépendent de la gravité de l'accident, l'odds ratio causal de l'alcool sur la responsabilité n'est pas identifiable, et ne peut être approximé. Dans ce cas, nous montrons que les odds ratios estimés n'ont, en toute rigueur, de validité ni externe ni interne.

Le troisième chapitre présente des résultats numériques afin d'illustrer les résultats théoriques

du chapitre précédent. Il s'agit de donner un premier ordre de grandeur du biais induit par les analyses en responsabilité lorsque les données disponibles concernent uniquement les accidents graves. Lorsque l'exposition et la vitesse sont positivement corrélées, l'OR estimable à partir des analyses en responsabilité correspond à une sous-estimation de l'odds ratio causal de l'exposition sur la responsabilité. En revanche, lorsque l'exposition et la vitesse sont négativement corrélées, l'OR estimable semble correspondre à une sur-estimation de l'odds ratio causal de l'exposition sur la responsabilité.

Enfin, dans le quatrième chapitre, nous montrons que les résultats du chapitre 2, et en particulier la description formelle de la validité externe et interne, fondée sur des considérations causales, est utile dans d'autres champs que la sécurité routière. Elle permet notamment d'expliquer d'autres résultats en présence de biais de sélection, comme celui en jeu dans le cadre du paradoxe de l'obésité.

Chapitre 1

Causalité et modèles causaux structuraux

Jusqu'à présent, l'identification d'une cause en sécurité routière et la mesure de son effet causal passe par la mise en évidence d'une association statistique, où les différents types de biais rencontrés, biais de confusion et biais de sélection, sont généralement discutés. Or, il est bien connu que « association ne veut pas dire causalité », et que les mesures d'association fondées sur les études observationnelles ont une faible valeur causale, contrairement à celles obtenues dans le cadre d'expérimentation.

L'objet de ce chapitre est d'éclaircir la notion de causalité, que ce soit dans sa forme littéraire ou formalisée, afin d'introduire les modèles causaux structuraux, outils statistiques permettant de détecter les biais de causalité sur les données non expérimentales. Il s'agit tout d'abord de présenter les conceptions philosophiques de causalité, utiles aux fondements de la causalité en sciences et dans les modèles causaux structuraux. Il s'agit ensuite d'exposer les conceptions classiques de l'établissement de la causalité en biostatistique et en épidémiologie à travers le recours à l'expérimentation ou aux critères de Bradford-Hill. La deuxième partie de ce chapitre présentera une approche formalisée de la causalité en abordant les principes généraux de la théorie de l'inférence causale, et en particulier des modèles causaux structuraux. La théorie de l'inférence causale a notamment l'intérêt de présenter une vision unifiée de la causalité en intégrant sa dimension structurelle et mécaniste, sa dimension contrefactuelle, interventionniste et probabiliste. Une troisième partie sera consacrée à l'analyse causale en médiation, utile pour décomposer un effet causal total en effet direct et indirect.

1.1 De l'association à la causalité : approches philosophiques et scientifiques

1.1.1 Les grandes conceptions philosophiques de la causalité

La causalité se définit généralement comme une relation qui unit une ou plusieurs causes à un ou plusieurs effets. Plusieurs grandes conceptions de la causalité sont généralement distinguées (Drouet and Kistler, 2012) :

— Les conceptions régularistes et néo-régularistes,
— La conception « physique » ou « mécaniste »,
— La conception « inférentielle ».

1. **La conception régulariste, et néo-régulariste de la causalité**

   La première théorie de la causalité a été développée par Hume, dans le Traité de la nature humaine (Hume, 1739), et toutes les philosophies contemporaines se sont construites en réaction à cette théorie. Selon Hume, la causalité est notamment caractérisée par la régularité : les causes sont régulièrement suivies par leurs effets. En sécurité routière, on pourrait par exemple citer l'exemple de l'endormissement au volant qui est généralement suivi d'un accident. Néanmoins, la plupart des causes ne sont pas toujours suivies de leurs effets, mais le sont en présence de certains autres facteurs. Par exemple, tous les conducteurs alcoolisés n'ont pas forcément un accident. Et à l'inverse, un effet peut être produit de plusieurs façons différentes, chaque ensemble de causes étant suffisante mais non nécessaire pour produire l'effet. La cause devient un élément non suffisant mais essentiel à l'une de ces façons de produire l'effet. C'est ce que Mackie (1974) appellera une condition INUS (insufficient but non-redundant parts of a condition which is itself unnecessary but sufficient for the occurrence of the effect).
   Le modèle de cause composante suffisante (Rothman, 1976, Rothman and Greenland, 2005) s'inscrit également dans la continuité de ces approches. Il définit un ensemble minimal d'actions, d'événements ou d'états permettant la survenue de l'événement étudié. Les facteurs de risque correspondent à des composantes, qui, en se combinant, composent un mécanisme causal, appelé cause suffisante, et qui provoque l'apparition de l'événement. Dans le cadre des accidents de la circulation, il est possible d'imaginer plusieurs mécanismes causaux : un mécanisme composé des facteurs alcool, vitesse, route départementale, nuit... un autre composé des facteurs milieu urbain, vélo, distraction...

2. **La conception « physique » ou « mécaniste » de la causalité**

   La conception « mécaniste » de la causalité est une critique faite par Russell (Russell, 1948) aux approches régularistes. Il s'agit de rompre avec les notions de nécessité et de suffisance, et d'aborder la causalité en tant que processus (Aronson, 1971, Fair, 1979). Un processus causal peut être identifié par sa capacité à transmettre du changement à travers le temps et l'espace. Une cause produit un effet si et seulement s'il existe un processus physique de transmission entre la cause et l'effet (par exemple, un transfert

d'énergie). L'absorption d'alcool conduirait à augmenter la vitesse de conduite, le risque de conduire à une vitesse inappropriée et donc provoquer un accident. Les notions de mécanisme causal (Glennan, 1996, Machamer et al., 2000) et de structure causale sont également importantes : le mécanisme causal génère un changement en raison de ses éléments constitutifs et de leur organisation fonctionnelle.

3. **La conception « inférentielle » de la causalité**

Selon cette conception, les causes sont pensées comme des « facteurs de différence ». L'objectif est donc de mesurer quelles différences font les causes sur leurs effets (Drouet and Kistler, 2012). Au sein de cette théorie « inférentielle » de la causalité plusieurs courants existent : la théorie probabiliste, la théorie contrefactuelle et la théorie interventionniste.

La théorie probabiliste

L'idée fondamentale de la théorie probabiliste de la causalité est de rompre le lien entre le concept de causalité et celui de nécessité et de suffisance. Il s'agit de définir une cause en tant que telle parce qu'elle rend ses effets plus probables. Une cause produit un effet si et seulement si l'occurrence de la cause augmente la probabilité d'occurrence de l'effet.

La tradition contemporaine de la causalité se fonde sur la définition de Suppes (1970). L'événement $\{X = x\}$ cause l'événement $\{Y = y\}$ si et seulement si :
— $\mathbb{P}(Y = y|X = x) > \mathbb{P}(Y = y)$ (notion probabiliste)
— $\{X = x\}$ antérieure à $\{Y = y\}$ (notion d'antériorité présente également chez Hume)
— Il n'existe pas de cause commune $\{W = w\}$ antérieure à $\{X = x\}$ et $\{Y = y\}$ telle que $\mathbb{P}(Y = y|X = x, W = w) = \mathbb{P}(Y = y|W = w)$

Néanmoins, dans certains cas particuliers, la théorie de Suppes peut conduire à considérer comme causales des relations qui ne le sont pas, ou inversement (Drouet and Kistler, 2012). Le plus célèbre de ces cas particuliers, et le plus fréquent en épidémiologie, correspond au « paradoxe de Simpson » (Simpson, 1951, Blyth, 1972). Le paradoxe de Simpson se manifeste quand une corrélation est positive (resp. négative) dans la population globale, et négative (resp. positive) dans toutes les sous-populations de cette population. Un des exemples connus correspond à l'admission en troisième cycle à l'université de Berkeley (Bickel et al., 1975). De façon générale, les statistiques montrent que le taux d'admission est plus élevé pour les garçons que pour les filles (44 % vs 35 %). Mais en détaillant les statistiques par départements (cf. tableau 1.1, le taux de filles admises est plus élevé que le taux de garçons admis dans presque tous les départements. Ainsi, la théorie de Suppes conduit à conclure que le fait d'être un garçon cause l'admission. En réalité, le taux d'admission global des filles est plus faible, car celles-ci s'orientent davantage vers les départements les plus sélectifs, là où les taux d'admission sont les plus faibles. Pour pallier au « paradoxe de Simpson », l'idée fondatrice $\mathbb{P}(Y = y|X = x) > \mathbb{P}(Y = y)$ devient $\mathbb{P}(Y = y|X = x, W = w) > \mathbb{P}(Y = y|W = w)$ où $W = w$ caractérise une strate pertinente, ici le département. C'est le principe de la stratification. L'idée qui émerge ensuite est d'analyser la causalité comme une augmentation de la probabilité toutes choses égales par ailleurs, *ceteris paribus* (Cartwright, 1979, Skyrms, 1980, Cartwright, 1989). C'est sur cette théorie de la causalité que se fonde l'établissement de la causalité en épidémio-

logie et en biostatistique mais tout l'enjeu est de déterminer sur quelle strate pertinente ajustée. Or, certains outils de l'inférence causale permettent de déterminer les variables W nécessaires pour interpréter un effet d'association en effet causal (cf. section 1.2).

TABLE 1.1 – Paradoxe de Simpson - Admission en troisième cycle à l'université de Berkeley (Bickel et al., 1975)

	Garçons		Filles	
Département	Candidats	Taux d'admis	Candidats	Taux d'admis
A	825	62 %	108	82 %
B	560	63 %	25	68 %
C	325	37 %	593	34 %
D	417	33 %	375	35 %
E	191	28 %	393	24 %
F	373	6 %	341	7 %
Total	2691	44 %	1835	35 %

La théorie contrefactuelle

Selon la théorie contrefactuelle de la causalité, initiée par Lewis (Lewis, 1973, 2000), la causalité ne se déduit pas uniquement de ce qui est observable, mais de ce qu'il se passe dans un autre monde possible (Chambaz et al., 2013). Une cause produit un effet s'il est vrai que « si la cause ne s'était pas produite, alors l'effet ne se serait pas produit non plus ». L'idée principale est de se poser la question suivante : « Que se serait-il passé si, contrairement au fait, ... le conducteur n'avait pas bu d'alcool ? ».

En recherche clinique, l'approche contrefactelle a conduit au développement d'un modèle connu sous le nom de modèle des « résultats potentiels », i.e. « potential outcome framework » (Rubin, 1974, Holland, 1986, Splawa-Neyman, 1990). Dans le cas simple où la variable potentiellement causale et l'effet sont binaires, le modèle se définit de la façon suivante :

Soient X une variable causale binaire qui prend pour valeur 1 ou 0, et Y la variable dépendante qui mesure l'effet de X sur une unité d'observation, l'individu i. Chaque individu peut alors potentiellement donner deux réponses Y_i :

$$\begin{cases} Y_{i1} & \text{la variable réponse que l'on observerait si on fixait } X_i = 1 \\ Y_{i0} & \text{la variable réponse que l'on observerait si on fixait } X_i = 0 \end{cases}$$

En observant (X_i, Y_i), une hypothèse classique est de considérer que la variable Y_i finalement observée est telle que :

$$Y_i = X_i Y_{i1} + (1 - X_i) Y_{i0},$$

c'est-à-dire

$$\begin{cases} Y_i = Y_{i1} & \text{si } X_i = 1 \\ Y_i = Y_{i0} & \text{si } X_i = 0 \end{cases}$$

Cette égalité s'interprète comme « la coïncidence de l'issue dans le monde actuel avec l'issue dans le monde contrefactuel exploré » (Chambaz et al., 2013).

Au niveau individuel, l'effet causal de X sur Y peut alors être défini comme la différence des « résultats potentiels » : $Y_{i1} - Y_{i0}$.

Au niveau de la population, on note alors

$$\begin{cases} Y_1 & \text{la variable réponse que l'on observerait si on fixait } X = 1 \\ Y_0 & \text{la variable réponse que l'on observerait si on fixait } X = 0 \end{cases}$$

L'effet causal moyen ou « Average Causal Effect » (ACE) est défini tel que :

$$ACE \quad = \quad \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1)$$

Ici, l'effet causal est défini sur l'échelle additive, mais il peut également se définir sur l'échelle multiplicative via le risque relatif causal $CRR = \mathbb{P}(Y_1 = 1)/\mathbb{P}(Y_0 = 1)$ ou l'odds ratio causal $COR = (\mathbb{P}(Y_1 = 1)/\mathbb{P}(Y_1 = 0))/(\mathbb{P}(Y_0 = 1)/\mathbb{P}(Y_0 = 0)$. A ce stade, l'effet causal est défini mais pas forcément estimable, puisqu'un seul des deux résultats potentiels est connu pour chaque individu : celui correspondant au niveau de X effectivement reçu. L'autre résultat potentiel correspond à une donnée manquante.

La théorie interventionniste

L'idée principale de la théorie interventionniste (Price, 1991, Woodward, 2003) est de définir X comme une cause de Y si et seulement si il est possible de changer la valeur de Y en changeant la valeur de X. Derrière cette théorie se profile l'idée d'une manipulation expérimentale idéalisée puisque la modification de X se réalise de manière fictive. Cela revient à tester une hypothèse du type : « X est la cause de Y si et seulement si il est vrai que si X variait, Y varierait également, toutes choses égales par ailleurs ».

C'est sur les conceptions inférentielles de la causalité que repose l'établissement de la causalité en épidémiologie et en biostatistique.

1.1.2 Etablir la causalité en épidémiologie et biostatistique, les approches classiques

Afin d'identifier, et surtout mesurer les effets des « causes » sur un événement, les études épidémiologiques et biostatistiques s'attachent à établir l'existence d'une association statistique entre une exposition (un facteur de risque ou un facteur protecteur) et un événement (une maladie, un accident ou une guérison). Dans le cas où X et Y sont par exemple binaires, une mesure d'association possible est le risque relatif $RR = \mathbb{P}(Y = 1|X = 1)/\mathbb{P}(Y = 1|X = 0)$. Il compare la probabilité d'occurrence de Y chez les exposés à la probabilité d'occurence de Y chez les non-exposés.

Lorsque l'expérimentation est possible, les associations statistiques qui en résultent sont considérées comme ayant une forte valeur causale. Par exemple, la mise en place d'un essai randomisé permet de contrôler l'attribution de l'exposition (avec ou sans tirage au sort), et de rendre les

sujets « exposés » et « non exposés » comparables, échangeables, et d'éviter ainsi les biais de confusion. L'essai randomisé permet en fait d'estimer la quantité $CRR = \mathbb{P}(Y_1 = 1)/\mathbb{P}(Y_0 = 1)$ malgré le fait que, pour chaque individu, une des variables contrefactuelles est manquante.
En effet, dans la majorité des cas, la probabilité $\mathbb{P}(Y_x = 1)$ diffère de celle traditionnellement estimée $\mathbb{P}(Y = 1 \mid X = x)$ pour $x \in (0, 1)$ (Meek and Glymour, 1994). Dans le second cas, il s'agit d'un raisonnement factuel. Le conditionnement probabiliste par X limite l'analyse à la sous-population des individus pour lesquels $X = x$. Dans le premier cas, le raisonnement est de type contrefactuel. Il s'agit de déterminer quelle serait la proportion d'individus où $Y = 1$ parmi l'ensemble de la population dans la situation où les variables qui impactent X seraient contrôlées, et si X était fixée à x (cf. figure 1.1).

FIGURE 1.1 – Causalité vs Association : différentes populations de références

Source :Causal Inference, Miguel A. Hernan, James M. Robins, 2011

Supposons que la population de notre essai randomisé soit représentée par la figure 1.1, et que les sous-populations des exposés et non-exposés soient assignées aléatoirement par une méthode pile ou face. Les sujets pour lesquels la pièce tombe sur « face » composent le groupe blanc, ceux pour lesquels la pièce tombe sur « pile » composent le groupe gris. On demande ensuite d'attribuer l'exposition ou le traitement au groupe blanc $X = 1$, et le groupe gris compose le groupe placebo où $X = 0$. On observe ensuite les probabilités suivantes, par exemple, $\mathbb{P}(Y = 1|X = 1) = 0.3$ et $\mathbb{P}(Y = 1|X = 0) = 0.6$. Le risque relatif est donc égal à $RR = 0.5$. Supposons maintenant que par erreur, ce soit le groupe gris qui ait reçu l'exposition ou le traitement, et le groupe blanc, le placebo. Comme l'exposition est assignée aléatoirement, cela ne change en fait pas les résultats. On observera à nouveau $\mathbb{P}(Y = 1|X = 1) = 0.3$ et $\mathbb{P}(Y = 1|X = 0) = 0.6$, et donc $RR = 0.5$. Les groupes d'exposés et non-exposés sont en fait « échangeables », « comparables ». Cela signifie notamment que la probabilité d'occurrence de Y dans le groupe blanc aurait été la même que celle observée dans le groupe gris, si les individus du groupe blanc avait reçu le traitement placebo du groupe gris. Formellement, cela équivaut à écrire l'égalité suivante $\mathbb{P}(Y_x = 1|X = 1) = \mathbb{P}(Y_x = 1|X = 0)$ pour $x \in (0, 1)$, et signifie que le traitement ou l'exposition effectivement reçu n'a pas d'impact sur la variable contrefactuelle. On note cela $Y_x \perp\!\!\!\perp X$. Et donc que les valeurs manquantes sur les issues contrefactuelles sont là « par hasard ». En présence de facteurs de confusion $\boldsymbol{W}$ influençant l'exposition X et l'événement Y, la randomisation rend les groupes « échangeables » conditionnellement à $\boldsymbol{W}$. On note cela $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$. Le recours à l'expérimentation via le processus de randomisation rend ainsi l'exposition indépendante de tous facteurs de confusion, et l'estimation de l'association statistique entre l'exposition et l'événement sans biais.

Chez les épidémiologistes, l'établissement de cette association statistique réside le plus souvent sur des études dites « observationnelles » en raison de l'impossibilité de la mise en place d'expérimentation, soit pour des raisons éthiques ou pratiques. Dans le cadre de la sécurité routière, il est par exemple difficile d'imaginer définir aléatoirement des groupes de sujets soumis ou non à une dose d'alcool, et d'observer ensuite si l'accident a lieu sur un itinéraire donné. Les études en sécurité routière fondées uniquement sur les cas, les études cas-témoins ou les enquêtes de cohorte font ainsi partie des études « observationnelles ». L'exposition est constatée, observée,

et ne peut être contrôlée par l'investigateur avant ou durant la collecte des données. Même si les chercheurs estiment ensuite des mesures d'association « ajustées » sur un ensemble de facteurs de confusion pour contrôler les biais, les arguments en faveur de la causalité restent faible.
L'établissement de la causalité repose donc sur un faisceau de critères plus larges, et s'appuie également sur des résultats externes à l'étude menée. Les critères de Bradford-Hill (Hill, 1965, Morabia, 1991) ont ainsi permis d'établir des relations causales sur le respect d'un maximum de critères parmi les huit présentés :

- la force de l'association (mesurée par l'odds ratio ou le risque relatif) : plus l'association entre le facteur de risque et l'événement est grande, plus il y a de chance que le facteur de risque soit causal.
- la cohérence : résultats similaires sur d'autres populations
- la spécificité : une cause produit un effet
- la temporalité : la cause doit précéder l'effet
- la relation dose-effet : plus la dose d'exposition est forte, plus l'effet est prononcé
- la plausibilité biologique
- la preuve expérimentale
- l'analogie avec d'autres événements

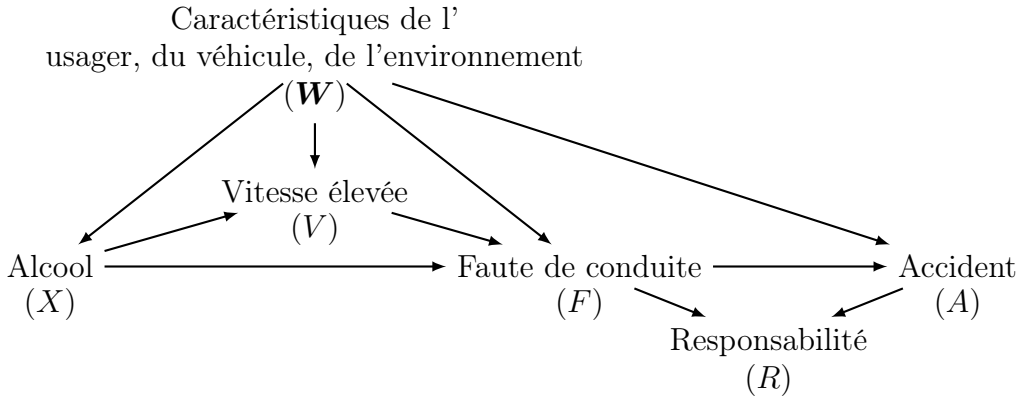
Néanmoins, dans le cas où l'ensemble des facteurs de confusion influençant l'exposition et l'événement seraient connus, récoltés et insérés dans les modèles d'estimation des mesures d'association (excès de risque, risque relatif ou odds ratio), l'approche observationnelle équivaut à l'approche expérimentale. Et la théorie de l'inférence causale permet notamment de montrer cette équivalence (cf. section 1.2.3).

1.2 Approche formalisée de la causalité : introduction aux modèles causaux structuraux

L'inférence causale, et en particulier la théorie des modèles structuraux (Pearl, 2000, 2009), établit des conditions suffisantes pour identifier et estimer des effets causaux définis à partir de variables contrefactuelles sur données observationnelles. La théorie de l'inférence causale a également l'avantage de présenter une vision unifiée de la causalité en intégrant sa dimension structurelle et mécaniste, sa dimension contrefactuelle, interventionniste et probabiliste.

Pour cela, la théorie de l'inférence causale peut s'appuyer sur un modèle graphique, le DAG (directed acyclic graph), appelé également structure causale. Le DAG relie les différents facteurs de risques conduisant à la survenue de l'événement. Un exemple de DAG représentant le mécanisme causal entre la consommation d'alcool et la survenue de l'accident est fourni par la figure 1.2. C'est ce modèle graphique qui contient l'information sur les potentielles relations causales entre les variables. Il est généralement déterminé par les experts, même si des travaux récents proposent d'estimer le DAG à partir des données (Kalisch and Bühlmann, 2013). Dans ce chapitre, on supposera que le DAG est connu et donc que l'ensemble des facteurs de confusion entre X et Y est également connu. C'est également la structure du DAG qui permet de déterminer si l'effet causal d'une variable sur un événement est identifiable à partir des données observées, et donc estimable.

FIGURE 1.2 – DAG représentant le mécanisme de survenue de l'accident



La démarche scientifique afin d'établir des relations causales d'après cette théorie consiste à :

1. Formuler les hypothèses causales et représenter leurs formes structurelles par le DAG
2. Définir la quantité causale d'intérêt : l'excès de risque, le risque relatif ou l'odds ratio
3. Déterminer si l'effet causal est identifiable selon les variables disponibles
4. Estimer l'effet causal par les méthodes statistiques appropriées

Dans un premier temps, nous présenterons les propriétés du DAG. Nous aborderons leur équivalence avec les systèmes d'équations structurelles qui permettent notamment la définition formelle

des variables contrefactuelles Y_x et donc des effets causaux. Les conditions d'identifiabilité de l'effet causal selon la structure du DAG dans un modèle markovien seront également abordées. Il s'agira ensuite de présenter les conditions d'identifiabilité de l'effet causal dans des modèles non markoviens, en particulier quand certaines variables du DAG ne sont pas observées. Ensuite, des exemples d'estimation de l'effet causal seront exposés.

1.2.1 DAG, réseau d'indépendances marginales et conditionnelles, loi jointe

Le DAG constitue la base de la modélisation causale, et correspond à une structure causale (cf. définition 1) (Pearl, 1995, Greenland et al., 1999, Glymour and Greenland, 2008).

Nous noterons :

- G, un DAG.
- X, la variable binaire, correspondant à l'exposition considérée.
- Y, la variable binaire, correspondant à l'événement d'intérêt.
- $\boldsymbol{W}$, l'ensemble des facteurs de confusion.
- $\boldsymbol{V}$, l'ensemble des variables composant le DAG G.
- PA, l'ensemble des parents d'une variable composant le DAG.
- ND, l'ensemble des non-descendants d'une variable composant le DAG.

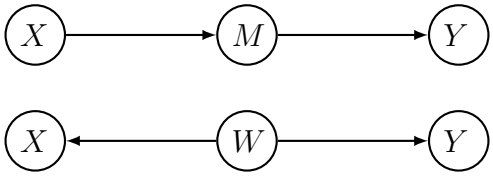
Définition 1. ***Structure causale*** *(Pearl, 2009)*
Une structure causale composée d'un ensemble de variables $\boldsymbol{V} = \{V_1, ..., V_n\}$ *est un diagramme acyclique orienté (DAG) dans lequel chaque noeud correspond à un élément distinct de* $\boldsymbol{V}$*, et chaque lien représente une relation fonctionnelle directe parmi* $\boldsymbol{V}$*.*

Le DAG est constitué d'un ensemble de chemins, composés de flèches et de noeuds. Ces derniers correspondant aux variables. Le DAG s'incrit ainsi dans la lignée du « path diagram » (Wright, 1921), première structure causale apparue dans le domaine de la biométrie, et repris ensuite en économétrie (Haavelmo, 1943) et en sociologie (Duncan, 1975). Le DAG représente les hypothèses relatives aux relations de cause à effet qui existent entre l'ensemble des variables $\boldsymbol{V}$. Le DAG décrit également le sens des relations entre les variables par le le sens des flèches. Les hypothèses représentées par le DAG sont qualitatives, elles portent sur l'existence d'une relation de cause à effet entre les variables, avant de porter sur l'intensité ou la forme de ces relations. La principale information du DAG réside alors dans l'absence de flèches. Les hypothèses du DAG prennent également en compte le caractère multivarié de la causalité, c'est-à-dire le fait qu'une même cause peut avoir plusieurs effets et qu'un même effet peut avoir plusieurs causes (Drouet and Kistler, 2012). Le DAG est toutefois un modèle graphique « dirigé » car les flèches représentent des directions, « acyclique » car une variable ne peut avoir d'impact sur elle-même, de manière directe ou indirecte. Les relations entre les variables peuvent être décrites comme des relations de parenté : lorsque qu'une variable a un impact direct ou indirect sur une autre, celle-ci correspond à un « ancêtre », tandis que l'autre variable correspond à un « descendant ». Si la relation est directe, c'est-à-dire non médiée par une troisième variable, on parle alors de « parent » et « d'enfant ». Le DAG représente ainsi un ensemble de chemins « ouverts » ou « fermés » correspondant à un réseau de dépendances/indépendances marginales. L'ajustement sur une variable « ferme » ou «

ouvre » les chemins créant ainsi des dépendances/indépendances conditionnelles.

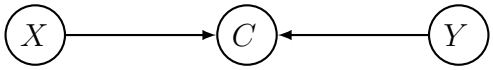
Il existe en effet deux types de chemins ouverts (cf. figure 1.3), correspondant à la dépendance marginale $X \not\perp\!\!\!\perp Y$.

FIGURE 1.3 – Exemples simples de chemins ouverts



Lorsque C est un « collider », c'est-à-dire lorsque deux flèches convergent vers une variable, le chemin est dit « fermé » ou « bloqué » (cf. figure 1.4), et nous avons une indépendance marginale $X \perp\!\!\!\perp Y$.

FIGURE 1.4 – Exemple simple de chemins fermés

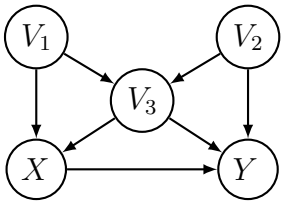


Ajuster sur une variable d'un chemin ouvert « bloque » le chemin, et $X \perp\!\!\!\perp Y|M$ ou $X \perp\!\!\!\perp Y|W$. A l'inverse, ajuster sur un collider ouvre un chemin initialement fermé, et nous avons $X \not\perp\!\!\!\perp Y|C$.

Le DAG représente ainsi un réseau de dépendances et d'indépendances conditionnelles défini par la condition de Markov causale (cf. définition 2). Le DAG correspond à un processus de Markov dans le sens où « le futur est indépendant du passé, conditionnellement au présent ».

Définition 2. ***Condition de Markov causale*** *(Pearl, 2009)*
Soit un DAG G représenté par un ensemble de noeuds, de covariables $V_1, ..., V_n$.
Soit PA_i *l'ensemble des parents de* V_i. *Soit* ND_i, *l'ensemble des non-descendants de* V_i *,tel que* $ND_i \not\subset PA_i$.
Si un DAG formalise une structure causale alors $V_i \perp\!\!\!\perp ND_i \mid PA_i$

FIGURE 1.5 – Exemple de DAG et d'indépendances associées



$Y \perp\!\!\!\perp V_1 \mid (V_2, V_3, X)$
$X \perp\!\!\!\perp V_2 \mid (V_3, V_1)$
$V_1 \perp\!\!\!\perp V_2$

Le DAG permet de définir également une fonction de probabilité, exprimant la loi jointe des variables incluses dans le DAG en fonction de leurs lois marginales et conditionnelles. Le lien

entre le DAG et la fonction de probabilité repose sur la condition de Markov causale factorisée.

Définition 3. ***Condition de Markov causale factorisée*** *(Pearl, 2009)*
Soit $\boldsymbol{V}$ un ensemble de variables $V_1, ..., V_n$, G un DAG sur $\boldsymbol{V}$. Si un DAG G formalise une structure causale, alors la probabilité jointe de l'ensemble $\boldsymbol{V}$ peut être factorisée sous la forme suivante :

$$\mathbb{P}(V_1, ..., V_n) = \prod_{i=1}^{n} \mathbb{P}(V_i \mid PA_i)$$

où PA_i représente l'ensemble des parents de V_i .

Par exemple, la loi jointe du DAG représenté par la figure 1.5 est telle que $\mathbb{P}(V_1, V_2, V_3, X, Y) = \mathbb{P}(V_1)\mathbb{P}(V_2)\mathbb{P}(V_3|V_1, V_2)\mathbb{P}(X|V_1, V_3)\mathbb{P}(Y|X, V_2, V_3)$.

1.2.2 Equations structurelles, et définition de quantités causales

A chaque DAG correspond un modèle d'équations structurelles où chaque variable du DAG est associée à une fonction autonome faisant intervenir les parents de cette variable, et une perturbation ε, appelée également variable exogène. Le DAG et le système d'équations structurelles constituent ainsi un modèle causal (cf. définition 4), et permettent de définir formellement les variables contrefactuelles Y_x, variables que l'on aurait observées si l'on avait fixé X à x.

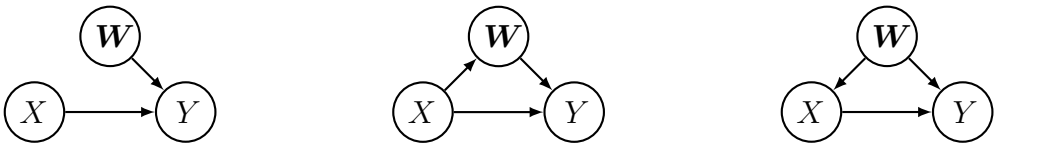
Définition 4. ***Modèle causal*** *(Pearl, 2009)*
Un modèle causal est un couple $\mathcal{M} = (G, \Theta_G)$ composé d'une structure causale G et un ensemble de paramètres Θ_G. Les paramètres Θ_G assignent une fonction $v_i = f_i(PA_i, \varepsilon_i)$ à chaque $V_i \in \boldsymbol{V}$ et une mesure de probabilité $\mathbb{P}(\varepsilon_i)$ à chaque ε_i, où PA_i sont les parents de V_i dans G et ε_i représentent des perturbations aléatoires distribuées selon $\mathbb{P}(\varepsilon_i)$, indépendamment des autres ε_i. L'indépendance des ε_i fait que le modèle causal est dit markovien.

Trois exemples de DAG et système d'équations structurelles correspondant sont fournis à travers la figure 1.6 où X, $\boldsymbol{W}$ et Y sont considérées comme binaires. Ces système d'équations structurelles gouvernent les distributions des variables $\boldsymbol{V} = (X, \boldsymbol{W}, Y)$ observées dans le monde réel et disponibles dans les données observationnelles. Mais elles permettent également la description de ce que l'on aurait observé dans le monde contrefactuel où l'on serait intervenu pour fixer $X = x$. Cette intervention, pouvant être modélisée par l'opérateur-*do* introduit par Pearl (2009), a pour effet de supprimer toutes flèches pointant sur X dans le DAG. Ainsi, d'après la figure 1.6, les DAGs a' et c' sont identiques alors que les DAGs a et c sont différents. Les deux modèles causaux sont identiques dans le monde contrefactuel où l'on aurait fixé X à x, mais différents dans le monde réel.
Dans la lignée de la vision mécaniste de la causalité, les quantités causales définies via les modèles causaux structuraux sous-entendent une capacité à transférer du changement entre les variables (Pearl, 2009). Même si l'intervention ne modifie pas les fonctions f_i ni les pertubations ε_i, elle affecte la distribution des descendants de X dans le DAG. Plus précisément, dans le monde contrefactuel suivant l'intervention $X = x$, le modèle causal $\mathcal{M}$ devient $\mathcal{M}_x$ et décrit les observations $\boldsymbol{V}_x = (x, \boldsymbol{W}_x, Y_x)$ au lieu de $\boldsymbol{V} = (X, \boldsymbol{W}, Y)$. Dans le DAG a par exemple, nous avons $\boldsymbol{W}_x = f_{\boldsymbol{W}}(\varepsilon_{\boldsymbol{W}}) = \boldsymbol{W}$ mais $Y_x = f_Y(x, \boldsymbol{W}_x, \varepsilon_Y) = f_Y(x, \boldsymbol{W}, \varepsilon_Y)$ qui est différent de $Y = f_Y(X, \boldsymbol{W}, \varepsilon_Y)$. Les équations structurelles permettent ainsi de définir formellement les

FIGURE 1.6 – Exemples de DAG et SEM associés dans le monde réel et le monde contrefactuel

Monde réel

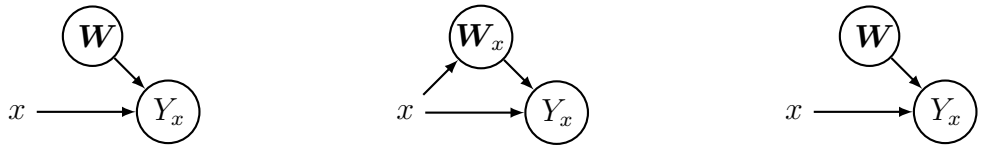
DAG a DAG b DAG c



$$\begin{cases} X &= f_X(\varepsilon_X) \\ \boldsymbol{W} &= f_{\boldsymbol{W}}(\varepsilon_{\boldsymbol{W}}) \\ Y &= f_Y(X, \boldsymbol{W}, \varepsilon_Y) \end{cases} \quad \begin{cases} X &= f_X(\varepsilon_X) \\ \boldsymbol{W} &= f_{\boldsymbol{W}}(X, \varepsilon_{\boldsymbol{W}}) \\ Y &= f_Y(X, \boldsymbol{W}, \varepsilon_Y) \end{cases} \quad \begin{cases} \boldsymbol{W} &= f_{\boldsymbol{W}}(\varepsilon_{\boldsymbol{W}}) \\ X &= f_X(\boldsymbol{W}, \varepsilon_X) \\ Y &= f_Y(X, \boldsymbol{W}, \varepsilon_Y) \end{cases}$$

Mondes contrefactuels correspondant, suivant l'intervention $do(X = x)$

DAG a' DAG b' DAG c'



$$\begin{cases} x & \\ \boldsymbol{W} &= f_{\boldsymbol{W}}(\varepsilon_{\boldsymbol{W}}) \\ Y_x &= f_Y(x, \boldsymbol{W}, \varepsilon_Y) \end{cases} \quad \begin{cases} x & \\ \boldsymbol{W}_x &= f_{\boldsymbol{W}}(x, \varepsilon_{\boldsymbol{W}}) \\ Y_x &= f_Y(x, \boldsymbol{W}_x, \varepsilon_Y) \end{cases} \quad \begin{cases} x & \\ \boldsymbol{W} &= f_{\boldsymbol{W}}(\varepsilon_{\boldsymbol{W}}) \\ Y_x &= f_Y(x, \boldsymbol{W}, \varepsilon_Y) \end{cases}$$

variables contrefactuelles, là où le modèle des résultats potentiels fournissait uniquement une définition verbale (Rubin, 1974, Holland, 1986, Splawa-Neyman, 1990) (cf. définition).
Ces variables contrefactuelles sont reliées au monde réel à travers l'hypothèse de consistence, aussi appelée hypothèse de cohérence :

$$\text{(Coh)} \qquad (X = x) \Rightarrow (Y = Y_x), \quad \text{pour toutes les valeurs potentielles } x \text{ de } X.$$

Cette hypothèse établit que l'événement observé dans le monde réel pour un individu ayant comme niveau d'exposition le niveau x, correspond à l'événement qui aurait été observé dans le monde contrefactuel pour le même individu si une intervention avait fixé le niveau de l'exposition au niveau x. Cette hypothèse est toujours vérifiée dans les exemples présentés ci-dessus puisque Y est toujours exprimée en fonction de X et Y_x en fonction de x. Cette hypothèse implique également la relation suivante : $Y = XY_1 + (1 - X)Y_0$. Les variables contrefactuelles Y_x pour $x \in (0, 1)$ sont donc des variables latentes partiellement observées. Ainsi, la problématique essentielle de l'inférence causale peut être vue comme un problème de données manquantes où l'on cherche à déterminer si $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y = 1 \mid do(X = 1))$ est estimable à partir des données recueillies et observées. $\mathbb{P}(Y_x = 1)$ représente la probabilité que Y soit égale à 1 dans le monde

contrefactuel où X aurait été fixée à x.

L'effet causal de X et Y peut ensuite être estimé sur différentes échelles : l'excès de risque ACE, le risque relatif CRR ou encore l'odds ratio COR. Dans le cas où X et Y sont binaires, ces trois quantités causales se définissent de la façon suivante :

$$\begin{aligned}
ACE &= \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1) \\
&= \mathbb{P}(Y = 1 \mid do(X = 1)) - \mathbb{P}(Y = 1 \mid do(X = 0)) \\
CRR &= \frac{\mathbb{P}(Y_1 = 1)}{\mathbb{P}(Y_0 = 1)} \\
&= \frac{\mathbb{P}(Y = 1 \mid do(X = 1))}{\mathbb{P}(Y = 1 \mid do(X = 0))} \\
COR &= \frac{\mathbb{P}(Y_1 = 1)/(1 - \mathbb{P}(Y_1 = 1)}{\mathbb{P}(Y_0 = 1)/(1 - \mathbb{P}(Y_0 = 1)} \\
&= \frac{\mathbb{P}(Y = 1 \mid do(X = 1))/(1 - \mathbb{P}(Y = 1 \mid do(X = 1))}{\mathbb{P}(Y = 1 \mid do(X = 0))/(1 - \mathbb{P}(Y = 1 \mid do(X = 0))}
\end{aligned}$$

Toute la question de l'inférence causale est de déterminer si la quantité causale d'intérêt (ACE, CRR ou COR) peut s'estimer à partir des données observées. Et c'est la structure du DAG qui permet d'étudier l'identifiabilité et donc l'estimation ou non d'une quantité causale.

1.2.3 Identifiabilité d'une quantité causale dans les modèles markoviens

L'identifiabilité de $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y = 1 \mid do(X = 1))$ et donc d'une quantité causale est définie de la façon suivante.

Définition 5. ***Identifiabilité de*** $\mathbb{P}(Y_x = 1)$ *(Bareinboim and Tian, 2015)*
$\mathbb{P}(Y_x = 1) = \mathbb{P}((Y = 1)|do(X = x))$ *est dite identifiable si les hypothèses induites par la structure du DAG assurent que cette quantité peut être exprimée à partir de la distribution des variables observées* $\boldsymbol{V}$ *qui composent le DAG.*

Nous précisons maintenant des conditions suffisantes pour l'identifiabilité de $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y = 1|do(X = x))$, à savoir l'hypothèse d'ignorabilité (Greenland et al., 1999) (aussi appelée hypothèse d'échangeabilité), et l'hypothèse d'ignorabilité conditionnelle (Rosenbaum and Rubin, 1983). Rappelons que ces conditions garantissent la comparabilité des groupes « exposés » vs « non-exposés » dans le monde réel. En particulier, nous avons :

$$\text{(Ign)} \qquad Y_x \perp\!\!\!\perp X \text{ pour toutes les valeurs potentielles } x \text{ de } X,$$

Sous l'hypothèse de cohérence et d'ignorabilité, et dès lors que $0 < \mathbb{P}(X = 1) < 1$ (hypothèse de

positivité), alors

$$\begin{aligned}\text{ACE} &= \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1) \\ &\overset{\text{Ign}}{=} \mathbb{P}(Y_1 = 1|X = 1) - \mathbb{P}(Y_0 = 1|X = 0) \\ &\overset{\text{Coh}}{=} \mathbb{P}(Y = 1|X = 1) - \mathbb{P}(Y = 1|X = 0)\end{aligned}$$

L'ACE est identifiable, c'est-à-dire estimable à partir des données observées, dès lors que la variable contrefactuelle est indépendante de la variable de l'exposition X, ie $Y_x \perp\!\!\!\perp X$. Comme mentionné dans la partie précédente, cette hypothèse est en particulier vérifiée lorsque l'exposition est attribuée aléatoirement, donc dans l'essai thérapeutique où l'expérimentateur intervient directement sur la variable X de manière aléatoire.

Sur les études observationnelles, l'hypothèse d'ignorabilité n'est pas testable. C'est la structure du DAG qui permet de déterminer la validité ou non de cette hypothèse. Plusieurs approches sont possibles, notamment les équations structurelles, et l'approche SWIGs (Single-World Intervention Graphs) (Richardson and Robins, 2013).

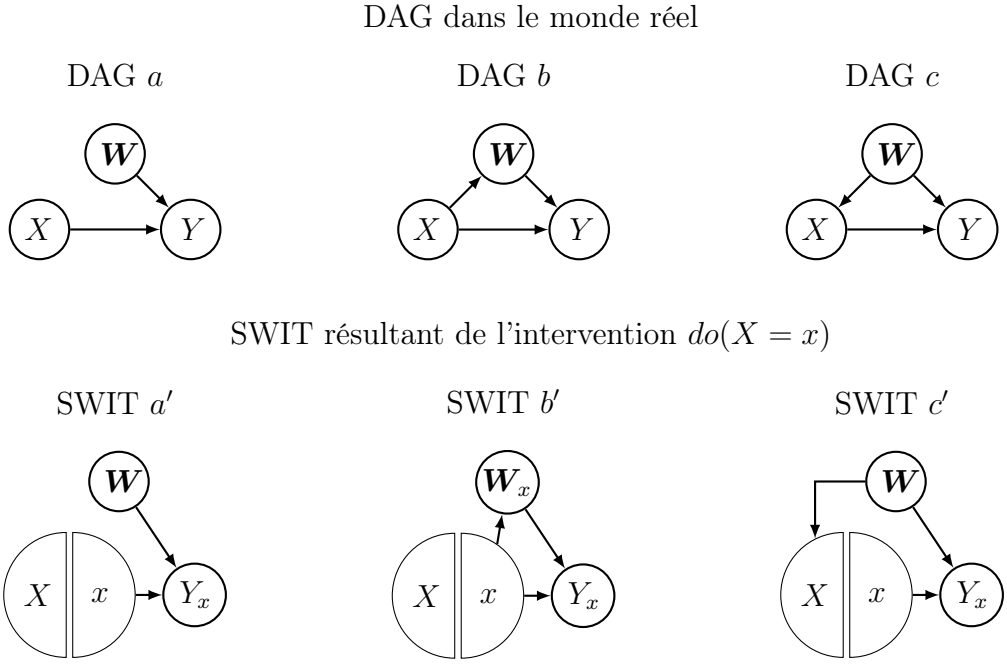
Les équations structurelles permettent de déterminer si l'hypothèse d'ignorabilité $Y_x \perp\!\!\!\perp X$ est vérifiée à travers la structure du DAG. Par exemple, dans le DAG a' de la figure 1.6, $Y_x = f_Y(x, \boldsymbol{W}, \varepsilon_Y)$. Or $\boldsymbol{W}$ est indépendant de X d'après le DAG a, et ε_Y est également indépendant de X. Par conséquent, $Y_x \perp\!\!\!\perp X$. Dans le DAG b' où $Y_x = f_Y(x, \boldsymbol{W}_x, \varepsilon_Y)$, W_x est indépendant de X (car $\varepsilon_{\boldsymbol{W}}$ est indépendant de X dans le DAG b), et ε_Y est toujours indépendant de X. Par conséquent, $Y_x \perp\!\!\!\perp X$ est vérifiée également. Dans le DAG c' où $Y_x = f_Y(x, \boldsymbol{W}, \varepsilon_Y)$, $\boldsymbol{W}$ et X ne sont pas indépendants d'après le DAG c et par conséquent $Y_x \not\perp\!\!\!\perp X$. En revanche, l'hypothèse d'ignorabilité conditionnelle $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$ est vérifiée (cf. (C.Ign) ci après)).

D'autre part, l'indépendance entre Y_x et X peut facilement être confirmée ou infirmée en utilisant l'approche SWIG (Richardson and Robins, 2013). L'approche SWIG a l'intérêt d'unifier la théorie graphique des DAG et ses chemins, avec les variables contrefactuelles. Cela en représentant graphiquement l'intervention par un noeud « partagé ». A chaque DAG de la figure 1.7 est donc associé un SWIT (Single-World Intervention Template), permettant de visualiser les chemins entre Y_x et X après l'intervention $do(X = x)$. En utilisant les critères graphiques définis dans la sous-section 1.2.1, il est alors possible de déterminer si $Y_x \perp\!\!\!\perp X$ est vérifiée. Dans le SWIT a' et b' de la figure 1.7, il n'existe par exemple pas de chemin entre Y_x et X, et donc $Y_x \perp\!\!\!\perp X$. Dans le SWIT c' associé au DAG c représentant le traditionnel phénomène de confusion, il existe un chemin ouvert entre les deux variables : $X \longleftarrow \boldsymbol{W} \longrightarrow Y_x$. Par conséquent, $Y_x \not\perp\!\!\!\perp X$. En revanche, nous avons $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$. Nous retrouvons l'hypothèse d'ignorabilité conditionnelle (Rosenbaum and Rubin, 1983) :

$$\text{(C.Ign)} \qquad Y_x \perp\!\!\!\perp X|\boldsymbol{W} \text{ pour toutes les valeurs potentielles } x \text{ de } X,$$

Sous l'hypothèse de cohérence et d'ignorabilité conditionnelle, et dès lors que $0 < \mathbb{P}(X = 1|W =$

FIGURE 1.7 – DAG et SWIT



$w) < 1$ (hypothèse de positivité), alors

$$\begin{aligned}
\text{ACE} &= \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1) \\
&= \sum_{\boldsymbol{w}} \{\mathbb{P}(Y_1 = 1|\boldsymbol{W} = \boldsymbol{w}) - \mathbb{P}(Y_0 = 1|\boldsymbol{W} = \boldsymbol{w})\}\mathbb{P}(\boldsymbol{W} = \boldsymbol{w}) \\
&\overset{\text{C.Ign}}{=} \sum_{\boldsymbol{w}} \{\mathbb{P}(Y_1 = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}) - \mathbb{P}(Y_0 = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})\}\mathbb{P}(\boldsymbol{W} = \boldsymbol{w}) \\
&\overset{\text{Coh}}{=} \sum_{\boldsymbol{w}} \{\mathbb{P}(Y = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}) - \mathbb{P}(Y = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})\}\mathbb{P}(\boldsymbol{W} = \boldsymbol{w})
\end{aligned}$$

Par conséquent, sous l'hypothèse de cohérence et d'ignorabilité conditionnelle, nous avons

$$\text{ACE} = \sum_{\boldsymbol{w}} \{\mathbb{P}(Y = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}) - \mathbb{P}(Y = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})\}\mathbb{P}(\boldsymbol{W} = \boldsymbol{w}), \tag{1.1}$$

où l'ACE correspond à une version « moyennée » de l'excès de risque ajusté.

Plus généralement, lorsque toutes les variables du DAG sont observées, que l'hypothèse de positivité est vérifiée et que les perturbations sont indépendantes, une quantité causale est toujours identifiable et donc estimable (Pearl, 2009). La question suivante est donc de se demander sous quelles conditions la quantité causale est tout de même identifiable lorsque certains facteurs ne sont pas observés.

1.2.4 Identifiabilité d'une quantité causale dans les modèles non markoviens

En présence de facteurs de confusion inobservés, les perturbations peuvent être dépendantes, et rendre le modèle causal non markovien. Dans cette situation, certains critères graphiques propres à l'inférence causale permettent de déterminer si une quantité causale est tout de même identifiable en fonction d'un ensemble suffisant de variables d'ajustement. Les plus connus de ces critères graphiques définis par Pearl correspondent au back-door et front-door criterion. Ces deux critères se fondent notamment sur un autre critère graphique appelé « d-séparation » (cf. figure 1.8). Ce critère permet notamment de résumer le réseau d'indépendances et de dépendances conditionnelles qui découle d'un DAG.

Définition 6. ***d-separation***

Un ensemble de noeuds W_0 bloque, ou d-sépare, un chemin p si

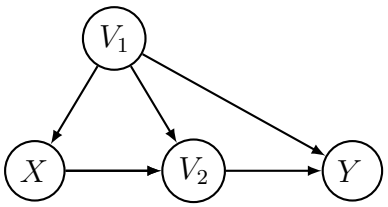
(*i*) *p contient au moins un noeud w émettant une flèche et tel que $w \in W_0$, ie p contient une séquence $i \to w \to j$ ou $i \leftarrow w \to j$, avec $w \in W_0$*

ou

(*ii*) *p contient au moins un collider w ($i \to w \leftarrow j$) tel que $w \notin W_0$ et qui n'a pas de descendant dans W_0.*

L'ensemble W_0 d-sépare X et Y dans G (ce qui implique $X \perp\!\!\!\perp Y | W_0$), si et seulement si W_0 d-sépare tous les chemins entre X et Y dans G.

FIGURE 1.8 – Illustration du critère de d-séparation



Soit $p_1 : X \leftarrow V_1 \to Y$. p_1 d-séparé par $\{V_1\}$
Soit $p_2 : X \leftarrow V_1 \to V_3 \to Y$. p_2 d-séparé par $\{V_1\}, \{V_2\}, \{V_1, V_2\}$
Soit $p_3 : X \leftarrow V_2 \to Y$. p_3 d-séparé par $\{V_2\}$.
$\Longrightarrow$ L'ensemble $\{V_1, V_2\}$ d-sépare tous les chemins entre X et Y.

Le critère back-door

Le back-door criterion permet d'identifier un ensemble de variables minimal W_0, suffisant pour identifier puis estimer $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y \mid do(X = x))$. Il a l'avantage de déterminer s'il est possible de se passer de certaines variables inobservées, selon leur localisation dans le DAG, mais également de déterminer les facteurs d'ajustement suffisants pour le calcul d'une quantité causale.

Définition 7. ***Le critère back-door***

Un ensemble de noeuds W_0 est admissible (suffisant pour l'ajustement) si les deux conditions suivantes sont vérifiées

(*i*) *Aucun élément de W_0 n'est un descendant de X*

(*ii*) *Les éléments de W_0 bloquent tous les chemins « back-door » entre X et Y, c'est-à-dire les chemins qui se terminent par une flèche pointant vers X.*

Dans le DAG de la figure 1.8, $\{V_1\}$ satisfait le critère back-door car il bloque le chemin back-door $X \leftarrow V_1 \rightarrow Y$. L'ensemble $\{V_3\}$ ne le satisfait pas car c'est un descendant de X.

Si W_0 satisfait le critère back-door, alors $Y_x \perp\!\!\!\perp X|W_0$, et donc

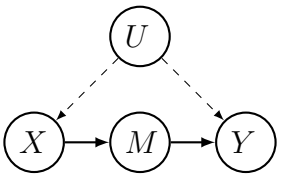
$$\begin{aligned}\mathbb{P}(Y_x = 1) &= \sum_{w_0} \mathbb{P}(Y_x = 1|W_0 = w_0)\mathbb{P}(W_0 = w_0) \\ &= \sum_{w_0} \{\mathbb{P}(Y_x = 1|X = x, W_0 = w_0)\mathbb{P}(W_0 = w_0) \quad \text{puisque} Y_x \perp\!\!\!\perp X|W_0 \\ &\overset{\text{Coh}}{=} \sum_{w_0} \{\mathbb{P}(Y = 1|X = x, W_0 = w_0)\mathbb{P}(W_0 = w_0)\end{aligned}$$

Dans le DAG c de la figure 1.6 où $W_0 = \boldsymbol{W}$, $\boldsymbol{W}$ satisfait le critère back-door puisqu'il bloque la piste back-door $X \leftarrow \boldsymbol{W} \rightarrow Y$ et n'est pas un descendant de X. Nous retrouvons ici la condition d'ignorabilité conditionnelle $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$. Si $\emptyset$ satisfait le critère back-door, en particulier s'il n'existe pas de chemin back-door entre X et Y, comme par exemple dans le DAG a et b de la figure 1.6, nous retrouvons la condition d'ignorabilité $Y_x \perp\!\!\!\perp X$ et donc $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y = 1|do(X = x)) = \mathbb{P}(Y = 1|X = x)$.

Le critère front-door

En l'absence d'un ensemble suffisant pour l'identifiabilité d'une quantité causale, cette dernière peut parfois être identifiée grâce au front-door criterion. Prenons comme exemple, le DAG de la figure 1.9 où U représente une variable inobservée.

FIGURE 1.9 – Illustration du Front-door criterion



Définition 8. ***Le critère front-door***

M satisfait le front-door criterion par rapport à (X, Y) si :

- *M bloque tous les chemins dirigés entre X et Y*
- *Il n'y a pas de chemins back-door non bloqués entre X et M*
- *Tous les chemins back-door entre M et Y sont bloqués par X*

Ces trois conditions sont vérifiées dans la figure 1.9. M d-sépare le chemin $X \longleftarrow M \longrightarrow Y$. Il n'y a pas de chemins back-door non bloqués entre X et M : Y étant un collider, le chemin back-door $X \longleftarrow U \longrightarrow Y \longleftarrow M$ est un chemin bloqué. Enfin, le chemin back-door $M \longleftarrow X \longleftarrow U \longrightarrow Y$ est bloqué par X.

Si de plus $\mathbb{P}(x, m) > 0$, alors

$$\begin{aligned}
\mathbb{P}(Y_x = 1) &= \sum_m \mathbb{P}(M_x = m)\ \mathbb{P}(Y_m = 1) \\
&= \sum_m \mathbb{P}(M_x = m)\ \sum_u \mathbb{P}(Y_m = 1|U = u)\mathbb{P}(U = u) \\
&= \sum_m \mathbb{P}(M_x = m|X = x)\ \sum_u \mathbb{P}(Y_m = 1|U = u, M = m)\mathbb{P}(U = u) \\
&\qquad \text{car } M_x \perp\!\!\!\perp X \text{ et } Y_m \perp\!\!\!\perp M|U \\
&= \sum_m \mathbb{P}(M = m|X = x)\ \sum_u \mathbb{P}(Y = 1|U = u, M = m)\mathbb{P}(U = u) \text{ par cohérence} \\
&= \sum_m \mathbb{P}(M = m|X = x)\ \sum_{u,x'} \mathbb{P}(Y = 1|U = u, M = m)\mathbb{P}(U = u|X = x')\mathbb{P}(X = x') \\
&= \sum_m \mathbb{P}(M = m|X = x)\ \sum_{u,x'} \mathbb{P}(Y = 1|U = u, M = m, X = x')\mathbb{P}(U = u|X = x', M = m)\mathbb{P}(X = x') \\
&\qquad \text{car } Y \perp\!\!\!\perp X|(M, U) \text{ et } U \perp\!\!\!\perp M|X \\
&= \sum_m \mathbb{P}(M = m|X = x)\ \sum_{x'} \mathbb{P}(Y = 1|M = m, X = x')\mathbb{P}(X = x')
\end{aligned}$$

Cette égalité peut être interprétée comme une application en deux étapes du critère back-door. D'une part, $\mathbb{P}(M = m \mid do(X = x)) = \mathbb{P}(M = m \mid X = x)$ car il n'y a pas de chemin back-door entre X et M. D'autre part, pour $\mathbb{P}(Y = 1 \mid do(M = m))$, il y a un chemin back-door $M \longleftarrow X \longleftarrow U \longrightarrow Y$ où X d-sépare le chemin entre M et Y.

Selon l'emplacement des facteurs non observés, une quantité causale peut donc être identifiable. Des situations plus complexes que le cas des critères back-door ou front-door peuvent également considérés (Tian and Pearl, 2002).

1.2.5 Estimation d'une quantité causale

Si une quantité causale est identifiable d'après la structure du graphe, il s'agit ensuite de l'estimer. Plusieurs options plus ou moins complexes sont possibles : des estimations non paramétriques, des estimations de type régression logistique ou linéaire, ou des méthodes plus avancées comme le score de propension ou la pondération par l'inverse du score de propension.

Si aucun ajustement n'est nécessaire pour identifier et estimer une des quantités causales, comme par exemple dans le cas du DAG b de la figure 1.6, une estimation non paramétrique suffit.
En effet, dans le DAG b, $Y_x \perp\!\!\!\perp X$, donc $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y = 1|do(X = x)) = \mathbb{P}(Y = 1|X = x)$. Il s'agit d'estimer $\pi_1 = \mathbb{P}(Y = 1|X = 1)$ et $\pi_0 = \mathbb{P}(Y = 1|X = 0)$. Soit un échantillon de taille n : $(X_1, Y_1), \ldots, (X_n, Y_n)$

$$\hat{\pi}_1 = \frac{\sum_i X_i Y_i}{\sum_i X_i} \quad \text{et} \quad \hat{\pi}_0 = \frac{\sum_i (1 - X_i) Y_i}{\sum_i (1 - X_i)}$$

où $\hat{\pi}_1$ correspond à la proportion de malades ($Y_i = 1$) chez les exposés ($X_i = 1$) et $\hat{\pi}_0$ correspond à la proportion de malades ($Y_i = 1$) chez les non-exposés ($X_i = 0$).
La quantité causale ACE est alors estimable de telle sorte que $\hat{\text{ACE}} = \hat{\pi}_1 - \hat{\pi}_0$.

Si un ajustement est nécessaire pour estimer une quantité causale, comme dans le DAG c, plusieurs options sont possibles.
Dans le DAG c, $Y_x \perp\!\!\!\perp X|W$, donc $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y = 1|do(X = x)) = \sum_{\boldsymbol{w}} \mathbb{P}(Y = 1|X = x, \boldsymbol{W} = \boldsymbol{w})\}\mathbb{P}(\boldsymbol{W} = \boldsymbol{w})$. La quantité causale d'intérêt est estimable à partir de l'estimation de $\pi(x, \boldsymbol{w}) = \mathbb{P}(Y = 1|X = x, \boldsymbol{W} = \boldsymbol{w})$. Il s'agit donc d'estimer $\pi(x, w) = \mathbb{P}(Y = 1|X = x, W = w)$ pour tout x, w, par exemple à l'aide d'un modèle logistique présenté ici sans interaction. Etant donné un échantillon de taille n : $(\boldsymbol{W}_1, X_1, Y_1), \ldots, (\boldsymbol{W}_n, X_n, Y_n)$

$$\hat{\pi}(x, w) = \frac{\exp(\hat{b}_0 + \hat{b}_X x + \hat{b}_{\boldsymbol{W}} \boldsymbol{w})}{1 + \exp(\hat{b}_0 + \hat{b}_X x + \hat{b}_{\boldsymbol{W}} \boldsymbol{w})}$$

Et dans ce cas, la quantité causale ACE est par exemple estimable de la façon suivante :

$$\hat{\text{ACE}} = \frac{1}{n} \sum_i \left\{\hat{\pi}(1, \boldsymbol{W}_i) - \hat{\pi}(0, \boldsymbol{W}_i)\right\}$$

Sous l'hypothèse où Y est rare, $\pi(x, w) << 1$, nous obtenons $\hat{\text{CRR}} \simeq \exp(\hat{b}_X)$. En effet, si Y est rare, $\hat{\pi}(x, w) \simeq \exp(\hat{b}_0 + \hat{b}_X x + \hat{b}_{\boldsymbol{W}} \boldsymbol{w})$. Par conséquent, $\hat{\text{CRR}} \simeq \frac{\exp(\hat{b}_0 + \hat{b}_X + \hat{b}_{\boldsymbol{W}} \boldsymbol{w})}{\exp(\hat{b}_0 + \hat{b}_{\boldsymbol{W}} \boldsymbol{w})} = \exp(\hat{b}_X)$. La quantité causale correspond à l'association estimé par un modèle de régression logistique standard.
En considérant le même DAG c mais avec Y continue et où il existerait une interaction entre X et $\boldsymbol{W}$, l'effet causal se définit tel que $ACE = \mathbb{E}(Y_1) - \mathbb{E}(Y_0)$. Comme $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$, alors $\mathbb{E}(Y_x) = \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}[Y|X = x, \boldsymbol{W}]\}$. Pour estimer l'ACE, il s'agit d'estimer $e(x, \boldsymbol{w}) = \mathbb{E}(Y|X = x, \boldsymbol{W} = \boldsymbol{w})$ pour tout $x, \boldsymbol{w}$. En choisissant un modèle de régression linéaire avec interaction entre X et $\boldsymbol{W}$, nous obtenons le résultat suivant. Etant donné un échantillon de taille n : $(\boldsymbol{W}_1, X_1, Y_1), \ldots, (\boldsymbol{W}_n, X_n, Y_n)$

$$\hat{e}(x, w) = \hat{\phi}_0 + \hat{\phi}_X x + \hat{\phi}_{\boldsymbol{W}} w + \hat{\gamma} x \boldsymbol{w}$$

Et dans ce cas, la quantité causale ACE est estimable de la manière suivante :

$$\begin{aligned}\hat{\text{ACE}} &= \frac{1}{n} \sum_i \left\{\hat{e}(1, \boldsymbol{W}_i) - \hat{e}(0, \boldsymbol{W}_i)\right\} \\ &= \hat{\phi}_X + \hat{\gamma} \bar{\boldsymbol{W}}_n.\end{aligned}$$

où $\bar{\boldsymbol{W}}_n = \frac{1}{n} \sum_i \boldsymbol{W}_i$.
La quantité causale considérée est cette fois-ci différent de l'effet d'association estimé par un modèle de régression linéaire standard avec interaction, puisqu'il fait intervenir la moyenne des $\boldsymbol{W}_i$.

Si les facteurs de confusion sont nombreux, l'estimation des quantités causales est plus complexe.

D'autres méthodes, comme par exemple le score de propension ou la pondération par l'inverse du score de propension (Rosenbaum and Rubin, 1983), peuvent être utilisées.

Le score de propension

Le score de propension $e(\boldsymbol{W})$ est défini comme la probabilité d'être exposé à X conditionnellement à un sous-ensemble de variables explicatives compris dans $\boldsymbol{W}$. Dans le cas où X est binaire, le score de propension se définit de la façon suivante :

$$SP = e(\boldsymbol{W}) = \mathbb{P}(X = 1|\boldsymbol{W} = \boldsymbol{w})$$

L'ajustement par le score de propension est justifiée par la propriété d'équivalence entre les indépendances conditionnelles. L'hypothèse d'ignorabilité peut être reformulée en remplaçant le vecteur de covariables $\boldsymbol{W}$ par le scalaire $SP = e(\boldsymbol{W})$ (Rosenbaum and Rubin, 1983, Imbens, 2000, Hirano and Imbens, 2004) :

$$Y_x \perp\!\!\!\perp X|\boldsymbol{W} \iff Y_x \perp\!\!\!\perp X|e(\boldsymbol{W})$$

L'idée est de créer un indicateur synthétique résumant les caractéristiques initiales des individus pour constituer ensuite des groupes comparables et différant uniquement sur l'exposition. Le calcul d'une quantité causale peut s'effectuer ensuite selon différentes stratégies (Williamson et al., 2012) :

- l'ajustement avec le score de propension comme covariable
  Cet ajustement se fait en intégrant le score de propension dans un modèle multivarié. Cette méthode est peu utilisée car sensible aux erreurs d'élaboration du score de propension. Elle nécessite également d'évaluer la relation entre Y et le score de propension si nécessaire (Rubin, 2004).
- la stratification des sujets sur le score de propension
  La stratification des sujets se réalise le plus souvent sur les quantiles du score de propension. Il s'agit ensuite d'estimer l'effet de l'exposition dans chaque strate, avant de calculer l'effet global à partir des estimations spécifiques de chaque strate.
- l'appariement
  L'appariement consiste à créer des paires de sujets « traités-non traités » à partir de la valeur du score de propension, et ensuite estimer l'effet de l'exposition sur la base appariée. Différentes techniques d'appariement peuvent être utilisées : la méthode du plus proche voisin, le caliper range (distance maximale entre 2 paires)...
- la pondération inverse
  Il s'agit d'estimer une quantité causale en pondérant le modèle multivarié par un indicateur construit à partir du score de propension.
  Dans le cas où X et Y sont binaires, on a par exemple :

$$
\begin{aligned}
ACE &= \mathbb{P}(Y=1|do(X=1)) - \mathbb{P}(Y=1|do(X=0)) \\
&= \sum_{\boldsymbol{w}} \{\mathbb{P}(Y=1|do(X=1), \boldsymbol{W}=\boldsymbol{w}) - \mathbb{P}(Y=1|do(X=0), \boldsymbol{W}=\boldsymbol{w})\}\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&= \sum_{\boldsymbol{w}} \{\mathbb{P}(Y=1|X=1, \boldsymbol{W}=\boldsymbol{w}) - \mathbb{P}(Y=1|X=0, \boldsymbol{W}=\boldsymbol{w})\}\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&= \sum_{\boldsymbol{w}} \{\frac{\mathbb{P}(Y=1, X=1, \boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(X=1, \boldsymbol{W}=\boldsymbol{w})} - \frac{\mathbb{P}(Y=1, X=0, \boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(X=0, \boldsymbol{W}=\boldsymbol{w})}\}\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&= \sum_{\boldsymbol{w}} \{\frac{\mathbb{P}(Y=1, X=1, \boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(X=1|\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w})} - \frac{\mathbb{P}(Y=1, X=0, \boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(X=0|\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w})}\}\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&= \sum_{\boldsymbol{w}} \{\frac{\mathbb{P}(Y=1, X=1, \boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(X=1|\boldsymbol{W}=\boldsymbol{w})} - \frac{\mathbb{P}(Y=1, X=0, \boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(X=0|\boldsymbol{W}=\boldsymbol{w})}\}
\end{aligned}
$$

Un estimateur de l'effet causal sur l'échelle de l'excès de risque ou ACE, peut ensuite se définir de la façon suivante :

$$
\begin{aligned}
\widehat{ACE} &= \frac{1}{n}\sum_{i=1}^{n} \{\frac{\mathbb{1}(Y_i=1, X_i=1, \boldsymbol{W}_i=w_i)}{e(\boldsymbol{w}_i)} - \frac{\mathbb{1}(Y_i=1, X_i=0, \boldsymbol{W}_i=\boldsymbol{w}_i)}{1-e(\boldsymbol{w}_i)}\} \\
&= \frac{1}{n}\sum_{i=1}^{n} \{\frac{Y_i X_i}{e(\boldsymbol{w}_i)} - \frac{Y_i(1-X_i)}{1-e(\boldsymbol{w}_i)}\}
\end{aligned}
$$

1.3 Analyse causale en médiation : décomposition d'un effet total

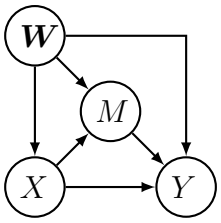
Jusqu'à présent, la probabilité $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y = 1|do(X = x))$ correspondait à l'effet *total* de la variable X sur Y. En présence de facteur intermédiaire ou médiateur (figure 1.10), l'effet total s'obtient sans ajustement sur le médiateur M. Il est maintenant utile de s'interroger à ce que l'on obtient si l'on ajuste sur M, car selon la question posée, il est plus intéressant de considérer cet effet ajusté, appelé effet *direct*. L'effet *direct* peut être défini comme un effet qui n'est pas « médié » par d'autres variables du modèle (Pearl, 2009), contrairement à l'effet *total*.
L'analyse en médiation permet ainsi de décomposer l'effet causal d'un facteur, c'est-à-dire l'effet total, en effet direct et indirect (Robins and Greenland, 1992, Pearl, 2001). Plusieurs approches classiques co-existent afin d'identifier les différents effets à partir de modèle de régression classiques. La théorie de l'inférence causale permet néanmoins de définir formellement ces différents effets et leur identifiabilité selon la structure du DAG (VanderWeele, 2009, VanderWeele and Vansteelandt, 2009, 2010, Imai et al., 2010).

1.3.1 Approches classiques de la médiation

Deux approches classiques existent pour aborder la médiation : l'approche par la différence et l'approche par produit (Baron and Kenny, 1986).

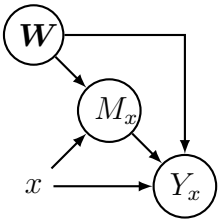
FIGURE 1.10 – DAG avec mediateur et SEM associé

DAG 1.a



$$\begin{cases} \boldsymbol{W} = f_{\boldsymbol{W}}(\varepsilon_{\boldsymbol{W}}) \\ X = f_X(\boldsymbol{W}, \varepsilon_X) \\ M = f_M(X, \boldsymbol{W}, \varepsilon_M) \\ Y = f_Y(X, M, \boldsymbol{W}, \varepsilon_Y) \end{cases}$$

DAG 1.b



$$\begin{cases} \boldsymbol{W} = f_{\boldsymbol{W}}(\varepsilon_{\boldsymbol{W}}) \\ x \\ M_x = f_M(x, \boldsymbol{W}, \varepsilon_M) \\ Y_x = f_Y(x, M_x, \boldsymbol{W}, \varepsilon_Y) \end{cases}$$

L'approche par la différence

L'approche par différence considère deux modèles de régression sur Y, un modèle non ajusté sur le médiateur M, et un modèle ajusté sur le médiateur.

$$\mathbb{E}[Y|X = x, \boldsymbol{W} = \boldsymbol{w}] = \phi_0 + \phi_X x + \phi_{\boldsymbol{W}} \boldsymbol{w}$$
$$\mathbb{E}[Y|X = x, M = m, \boldsymbol{W} = \boldsymbol{w}] = \theta_0 + \theta_X x + \theta_M m + \theta_{\boldsymbol{W}} \boldsymbol{w}$$

D'après cette approche, une quantité causale correspondant à l'effet total de X sur Y, l'effet direct et indirect se défini comme suit :

$$\text{ACE} = \phi_X$$
$$\text{Effet direct} = \theta_X$$
$$\text{Effet indirect} = \phi_X - \theta_X$$

L'approche par produit

L'approche par produit, ou l'approche Baron-Kenny (Baron and Kenny, 1986), considère deux modèles de régression : un modèle de régression sur Y et ajusté sur le médiateur M et un modèle de régression sur ce médiateur M.

$$\mathbb{E}[Y|X = x, M = m, \boldsymbol{W} = \boldsymbol{w}] = \theta_0 + \theta_X x + \theta_M m + \theta_{\boldsymbol{W}} \boldsymbol{w}$$
$$\mathbb{E}[M|X = x, \boldsymbol{W} = \boldsymbol{w}] = \beta_0 + \beta_X x + \beta_{\boldsymbol{W}} \boldsymbol{w}$$

D'après le DAG de la figure 1.10, on peut alors montrer que :

$$\begin{aligned} Y &= \theta_0 + \theta_X X + \theta_M M + \theta_{\boldsymbol{W}} \boldsymbol{w} + \varepsilon_Y \\ &= \theta_0 + \theta_X X + \theta_M(\beta_0 + \beta_X X + \beta_{\boldsymbol{W}} \boldsymbol{w} + \varepsilon_M) + \theta_{\boldsymbol{W}} \boldsymbol{w} + \varepsilon_Y \\ &= \theta_0 + \theta_M \beta_0 + (\theta_X + \theta_M \beta_X) X + (\theta_{\boldsymbol{W}} + \theta_M \beta_{\boldsymbol{W}}) \boldsymbol{W} + \varepsilon_Y + \theta_M \varepsilon_M \\ \textit{et donc } \mathbb{E}(Y|X, \boldsymbol{W}) &= \theta_0 + \theta_M \beta_0 + (\theta_X + \theta_M \beta_X) X + (\theta_{\boldsymbol{W}} + \theta_M \beta_{\boldsymbol{W}}) \boldsymbol{W} \end{aligned}$$

D'après l'approche produit, une quantité causale correspondant à l'effet total, direct et indirect se définit comme suit :

$$\begin{aligned} \text{ACE} &= \theta_X + \theta_M \beta_X \\ \text{Effet direct} &= \theta_X \\ \text{Effet indirect} &= \theta_M \beta_X \end{aligned}$$

1.3.2 Définition des effets totaux, controlés, directs et indirects

La théorie de l'inférence causale, et notamment les notations contrefactuelles permettent, contrairement aux approches standard, de définir formellement les effets direct et indirect.

Soient Y, X, et M des variables binaires. Introduisons les variables contrefactuelles suivantes :

- Y_x, variable Y suivant l'intervention $do(X = x)$
- M_x, variable M suivant l'intervention $do(X = x)$
- $Y_{x,m}$, variable Y suivant la double l'intervention $do(X = x, M = m)$
- Y_{x,M_x}, variable Y suivant l'intervention $do(X = x, M = M_x)$

Par conséquent, pour $x \in (0, 1)$, il existe 4 variables du type Y_{x,M_x} :

- Y_{0,M_0}, variable Y suivant l'intervention $do(X = 0, M = M_0)$
- Y_{1,M_1}, variable Y suivant l'intervention $do(X = 1, M = M_1)$
- Y_{0,M_1}, variable Y suivant l'intervention $do(X = 0, M = M_1)$
- Y_{1,M_0}, variable Y suivant l'intervention $do(X = 1, M = M_0)$

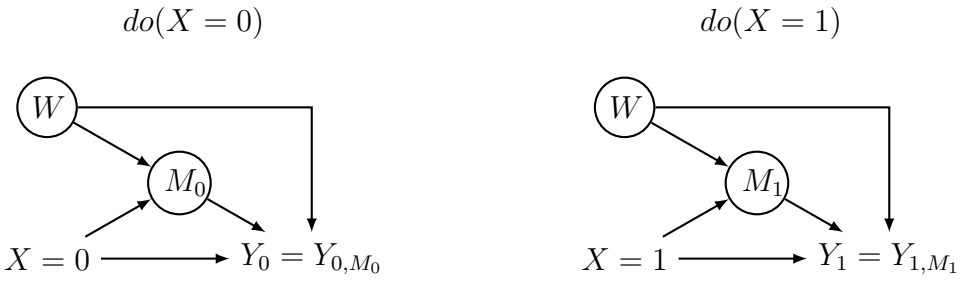
En d'autres termes, si X représente l'alcool, M la vitesse et Y la responsabilité de l'accident, alors :

- M_1 est la vitesse d'un conducteur sur lequel on serait intervenu pour qu'il soit alcoolisé
- M_0 est la vitesse d'un conducteur sur lequel on serait intervenu pour qu'il ne soit pas alcoolisé
- $Y_{1,M_1} = Y_1$ est la responsabilité d'un conducteur sur lequel on serait intervenu pour qu'il soit alcoolisé, et qu'il roule à la vitesse à laquelle il aurait roulé si on était intervenu pour qu'il soit alcoolisé.
- Y_{0,M_0} est la responsabilité d'un conducteur sur lequel on serait intervenu pour qu'il ne soit pas alcoolisé, et qu'il roule à la vitesse à laquelle il aurait roulé si on était intervenu pour qu'il ne soit pas alcoolisé.
- Y_{0,M_1}, est la responsabilité d'un conducteur sur lequel on serait intervenu pour qu'il ne soit pas alcoolisé, et qu'il roule à la vitesse à laquelle il aurait roulé si on était intervenu pour qu'il soit alcoolisé.
- Y_{1,M_0}, est la responsabilité d'un conducteur sur lequel on serait intervenu pour qu'il soit alcoolisé, et qu'il roule à la vitesse à laquelle il aurait roulé si on était intervenu pour qu'il ne soit pas alcoolisé.

Notons qu'aucune des quantités Y_{0,M_1} et Y_{1,M_0} ne sont pas observables, même en situation expérimentale.

Effet total (TE)
L'effet total compare l'effet de l'intervention où $x = 1$ à l'intervention où $x = 0$. Chaque intervention modifie également la valeur de M_x à M_1 ou M_0 respectivement (cf. figure 1.11).

FIGURE 1.11 – DAG avec mediateur dans chaque monde contrefactuel



Par conséquent, sur l'échelle des excès de risque, l'effet total se définit tel que :

$$\begin{aligned} ACE^{TE} &= \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1) \\ &= \mathbb{P}(Y_{1,M_1} = 1) - \mathbb{P}(Y_{0,M_0} = 1) \end{aligned}$$

Effet direct controlé (CDE)
L'effet contrôlé compare l'effet de l'intervention $X = 1$ à l'intervention $X = 0$ en fixant $M = m$. Une quantité causale peut alors être :

$$ACE^{CDE}(m) = \mathbb{P}(Y_{1,m} = 1) - \mathbb{P}(Y_{0,m} = 1)$$

Effet direct « naturel » (NDE)
L'effet direct compare l'effet l'intervention $X = 1$ à l'intervention $X = 0$ en laissant fixe la valeur de M soit à $M = M_0$ soit à $M = M_1$. Une quantité causale peut alors être :

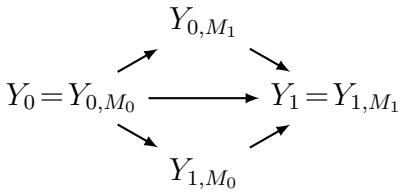
$$\begin{aligned} & & ACE^{NDE}(0) &= \mathbb{P}(Y_{1,M_0} = 1) - \mathbb{P}(Y_{0,M_0} = 1) \\ & \text{ou respectivement} & ACE^{NDE}(1) &= \mathbb{P}(Y_{1,M_1} = 1) - \mathbb{P}(Y_{0,M_1} = 1) \end{aligned}$$

Effet indirect « naturel » (NIE)
L'effet indirect compare l'effet de l'intervention sur le médiateur $M = M_1$ vs $M = M_0$ en laissant fixe X soit à la valeur $X = 1$ soit à $X = 0$. Une quantité causale peut alors être :

$$\begin{aligned} & & ACE^{NIE}(1) &= \mathbb{P}(Y_{1,M_1} = 1) - \mathbb{P}(Y_{1,M_0} = 1) \\ & \text{ou respectivement} & ACE^{NIE}(0) &= \mathbb{P}(Y_{0,M_1} = 1) - \mathbb{P}(Y_{0,M_1} = 1) \end{aligned}$$

Décomposition de l'effet total
Il existe deux possibilités de décomposition de l'effet total, qui peuvent être schématiquement représentées par la figure 1.12.
Sur l'échelle des excès de risque, la décomposition de l'effet total correspond à la somme des

FIGURE 1.12 – Décomposition effet total



effets directs et indirects :

$$\begin{aligned}
ACE^{TE} &= \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1) \\
&= \mathbb{P}(Y_{1,M_1} = 1) - \mathbb{P}(Y_{0,M_0} = 1) \\
&= \mathbb{P}(Y_{1,M_1} = 1) - \mathbb{P}(Y_{1,M_0} = 1) + \mathbb{P}(Y_{1,M_0} = 1) - \mathbb{P}(Y_{0,M_0} = 1) \\
&= ACE^{NIE}(1) + ACE^{NDE}(0) \\
&= \mathbb{P}(Y_{1,M_1} = 1) - \mathbb{P}(Y_{0,M_1} = 1) + \mathbb{P}(Y_{0,M_1} = 1) - \mathbb{P}(Y_{0,M_0} = 1) \\
&= ACE^{NDE}(1) + ACE^{NIE}(0)
\end{aligned}$$

Sur l'échelle des odds ratios, la décomposition de l'effet total correspond au produit des effets directs et indirects (VanderWeele and Vansteelandt, 2010) :

$$\begin{aligned}
OR^{TE} &= \frac{\mathbb{P}(Y_1 = 1)/(1 - \mathbb{P}(Y_0 = 1))}{\mathbb{P}(Y_0 = 1)/(1 - \mathbb{P}(Y_0 = 1))} \\
&= \frac{\mathbb{P}(Y_{1,M_1} = 1)/(1 - \mathbb{P}(Y_{1,M_1} = 1))}{\mathbb{P}(Y_{0,M_0} = 1)/(1 - \mathbb{P}(Y_{0,M_0} = 1))} \\
&= \frac{\mathbb{P}(Y_{1,M_1} = 1)/(1 - \mathbb{P}(Y_{1,M_1} = 1))}{\mathbb{P}(Y_{1,M_0} = 1)/(1 - \mathbb{P}(Y_{1,M_0} = 1))} \times \frac{\mathbb{P}(Y_{1,M_0} = 1)/(1 - \mathbb{P}(Y_{1,M_0} = 1))}{\mathbb{P}(Y_{0,M_0} = 1)/(1 - \mathbb{P}(Y_{0,M_0} = 1))} \\
&= OR^{NIE}(1) \times OR^{NDE}(0) \\
&= \frac{\mathbb{P}(Y_{1,M_1} = 1)/(1 - \mathbb{P}(Y_{1,M_1} = 1))}{\mathbb{P}(Y_{0,M_1} = 1)/(1 - \mathbb{P}(Y_{0,M_1} = 1))} \times \frac{\mathbb{P}(Y_{0,M_1} = 1)/(1 - \mathbb{P}(Y_{0,M_1} = 1))}{\mathbb{P}(Y_{0,M_0} = 1)/(1 - \mathbb{P}(Y_{0,M_0} = 1))} \\
&= OR^{NDE}(1) \times OR^{NIE}(0)
\end{aligned}$$

1.3.3 Identifiabilité et estimation des effets directs et indirects

Afin d'estimer les effets directs et indirects, plusieurs conditions d'identifiabilité sont requises (VanderWeele, 2009, VanderWeele and Vansteelandt, 2009, 2010, Imai et al., 2010) :

- $Y_{x,m} \perp\!\!\!\perp X | \boldsymbol{W}$, i.e. l'absence de facteurs inobservés entre l'exposition et l'outcome.
- $Y_{x,m} \perp\!\!\!\perp M | \boldsymbol{W}, X$, i.e. l'absence de facteurs inobservés entre le médiateur et l'outcome.
- $M_x \perp\!\!\!\perp X | \boldsymbol{W}$, i.e. l'absence de facteurs inobservés entre l'exposition et le médiateur.
- $Y_{x,m} \perp\!\!\!\perp M_{x^*} | \boldsymbol{W}$, i.e. l'absence de facteurs de confusion entre le médiateur et l'outcome, affecté par l'exposition.

Si toutes ces hypothèses sont vérifiées, on obtient, par exemple :

$$\begin{aligned}
\mathbb{P}(Y_{0,M_1}=1) &= \sum_{w} \mathbb{P}(Y_{0,M_1}=1|\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&= \sum_{w,m} \mathbb{P}(Y_{0,M_1}=1|M_1=m,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(M_1=m|\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&= \sum_{w,m} \mathbb{P}(Y_{0,m}=1|M_1=m,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(M_1=m|\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&\quad \text{car si } M_1=m \text{ alors } Y_{0,M_1}=Y_{0,m} \\
&= \sum_{w,m} \mathbb{P}(Y_{0,m}=1|\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(M_1=m|\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&\quad \text{puisque } Y_{0,m} \perp\!\!\!\perp M_1|\boldsymbol{W} \\
&= \sum_{w,m} \mathbb{P}(Y_{0,m}=1|\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(M=m|X=1,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&\quad \text{puisque } M_1 \perp\!\!\!\perp X|\boldsymbol{W} \text{ et par cohérence} \\
&= \sum_{w,m} \mathbb{P}(Y_{0,m}=1|X=0,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(M=m|X=1,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&\quad \text{puisque } Y_{x,m} \perp\!\!\!\perp X|\boldsymbol{W} \\
&= \sum_{w,m} \mathbb{P}(Y_{0,m}=1|X=0,M=m,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(M=m|X=1,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&\quad \text{puisque } Y_{x,m} \perp\!\!\!\perp M|\boldsymbol{W},X \\
&= \sum_{w,m} \mathbb{P}(Y_{0,m}=1|X=0,M=m,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(M=m|X=1,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
&\quad \text{car si } X=0,M=m \text{ alors } Y=Y_{0,m}
\end{aligned}$$

Les quantités $\mathbb{P}(Y_{1,M_0}=1)$, $\mathbb{P}(Y_{1,M_0}=1)$ et $\mathbb{P}(Y_{1,M_0}=1)$ sont exprimables de la même manière en suivant un raisonnement analogue. Par conséquent,

$$\begin{aligned}
ACE^{NDE}(0) &= \sum_{m,w}\{\mathbb{P}(Y=1|X=1,M=m,\boldsymbol{W}=\boldsymbol{w})-\mathbb{P}(Y=1|X=0,M=m,\boldsymbol{W}=\boldsymbol{w})\} \\
&\qquad \mathbb{P}(M=m|X=0,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
ACE^{NDE}(1) &= \sum_{m,w}\{\mathbb{P}(Y=1|X=1,M=m,\boldsymbol{W}=\boldsymbol{w})-\mathbb{P}(Y=1|X=0,M=m,\boldsymbol{W}=\boldsymbol{w})\} \\
&\qquad \mathbb{P}(M=m|X=1,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
ACE^{NIE}(0) &= \sum_{m,w}\{\mathbb{P}(M=m|X=1,\boldsymbol{W}=\boldsymbol{w})-\mathbb{P}(M=m|X=0,\boldsymbol{W}=\boldsymbol{w})\} \\
&\qquad \mathbb{P}(Y=1|X=0,M=m,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w}) \\
ACE^{NIE}(1) &= \sum_{m,w}\{\mathbb{P}(M=m|X=1,\boldsymbol{W}=\boldsymbol{w})-\mathbb{P}(M=m|X=0,\boldsymbol{W}=\boldsymbol{w})\} \\
&\qquad \mathbb{P}(Y=1|X=1,M=m,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w})
\end{aligned}$$

Ces formules généralisent les approches classiques de la médiation par différence ou par produit.

Si l'on suppose par exemple que le modèle causal repose sur les modèles linéaires introduit dans l'approche Baron-Kenny, on retrouve alors $ACE^{NDE}(0) = ACE^{NDE}(1) = \theta_X$ et $ACE^{NIE}(0) = ACE^{NIE}(1) = \theta_M\beta_X$. En effet, en reprenant les notations des modèles impliqués dans l'approche produit de Baron-Kenny, il vient, en supposant $\boldsymbol{W}$ univarié, d'une part :

$$\begin{aligned}
ACE^{NDE}(0) &= \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}_{M|(X=0,\boldsymbol{W})}[\mathbb{E}(Y|X=1,M,\boldsymbol{W}) - \mathbb{E}(Y|X=0,M,\boldsymbol{W})]\} \\
&= \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}_{M|(X=0,\boldsymbol{W})}(\theta_0 + \theta_X + \theta_M M + \theta_{\boldsymbol{W}}\boldsymbol{W} - \theta_0 - \theta_M M - \theta_{\boldsymbol{W}}\boldsymbol{W})\} \\
&= \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}_{M|(X=0,\boldsymbol{W})}(\theta_X)\} \\
&= \theta_X
\end{aligned}$$

En procédant de même, il vient facilement $ACE^{NDE}(1) = \theta_X$. D'autre part,

$$\begin{aligned}
ACE^{NIE}(1) &= \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}_{M|(X=1,\boldsymbol{W})}\mathbb{E}(Y|X=1,M=m,\boldsymbol{W})\} - \mathbb{E}_{W}\{\mathbb{E}_{M|(X=0,\boldsymbol{W})}\mathbb{E}(Y|X=1,M=m,\boldsymbol{W})\} \\
&= \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}_{M|(X=1,\boldsymbol{W})}(\theta_0 + \theta_X + \theta_M M + \theta_{\boldsymbol{W}}\boldsymbol{W})\} \\
&\quad - \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}_{M|(X=0,\boldsymbol{W})}(\theta_0 + \theta_X + \theta_M M + \theta_{\boldsymbol{W}}\boldsymbol{W})\} \\
&= \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}_{M|(X=1,\boldsymbol{W})}(\theta_M M)\} - \mathbb{E}_{\boldsymbol{W}}\{\mathbb{E}_{M|(X=0,\boldsymbol{W})}(\theta_M M) \\
&= \theta_M\mathbb{E}_{\boldsymbol{W}}\{[\mathbb{E}_{M|(X=1,\boldsymbol{W})}(M) - \mathbb{E}_{M|X=0,\boldsymbol{W}}(M)]\} \\
&= \theta_M\mathbb{E}_{\boldsymbol{W}}\{(\beta_0 + \beta_X + \beta_{\boldsymbol{W}}\boldsymbol{W} - \beta_0 - \beta_v\boldsymbol{W})\} \\
&= \theta_M\beta_X
\end{aligned}$$

En procédant de même, il vient facilement $ACE^{NIE}(0) = \theta_M\beta_X$.

Des méthodes d'estimations non paramétriques ont également été récemment développées, notamment dans l'objectif d'estimer des effets directs et indirects en présence de plusieurs médiateurs (Tchetgen Tchetgen, 2013, Nguyen et al., 2015)

Conclusion

Plusieurs conceptions philosophiques existent afin de définir la notion délicate de causalité : la conception régulariste, la conception mécaniste, et la conception inférentielle. L'établissement et la mesure de la causalité en épidémiologie ou biostatistique par les approches classiques s'appuient essentiellement sur la conception inférentielle de la causalité, et en particulier sur sa dimension probabiliste. X est une cause de Y si X rend plus probable l'apparition de Y. Fondées sur des données observationnelles, et non expérimentales, les relations causales mesurées en épidémiologie ont cependant une faible valeur causale, car elles sont soupçonnées d'être entachées de biais de confusion et de sélection. Or l'inférence causale, et en particulier la théorie des modèles structuraux (Pearl, 2000, 2009), montre qu'il est possible d'identifier et d'estimer des quantités causales définies à partir de données observationnelles. La théorie de l'inférence a l'intérêt de regrouper plusieurs dimensions de la causalité en intégrant sa dimension structurelle et mécaniste, sa dimension contrefactuelle, interventionniste et probabiliste. Le DAG intègre la notion de structure. Les équations structurelles, qui répercutent les changements entre fonctions, sous-entendent la dimension mécaniste. La définition des quantités causales repose sur sa dimen-

sion contrefactuelle et interventionniste.
Au delà de sa vision unifiée de la causalité, l'inférence causale a surtout l'intérêt de présenter une approche formalisée de la causalité, utile pour identifier et estimer les quantités causales sur données observationnelles. Les quantités causales peuvent être formellement définis grâce aux variables contrefactuelles ou l'opérateur-*do*. C'est ensuite la structure du DAG qui permet de déterminer l'identifiabilité d'une quantité causale et le choix des variables d'ajustement. Lorsque toutes les variables sont observées, les quantités causales sont toujours identifiables et donc estimables. En présence de facteurs inobservés, et selon la place de ces facteurs dans le DAG, les quantités causales peuvent être identifiables et estimables. Enfin, en présence d'un médiateur, les effets causaux totaux peuvent ensuite être décomposés en effet direct et indirect. Maintenant les principes de l'inférence causale énoncés, il s'agit de formaliser la causalité dans les analyses en responsabilité.

Chapitre 2

Formalisation des analyses en responsabilité

Ce chapitre a fait l'objet d'un article « Causal inference to formalize responsibility analyses » soumis dans Journal of causal inference, et également disponible sur Arxiv (Dufournet et al., 2017).

Dans le chapitre précédent, nous avons montré que l'identification et la mesure de relations causales étaient possibles sur données observationnelles, en supposant la structure du modèle causal connue. Ces modèles permettent notamment la détermination d'un ensemble de variables suffisantes pour mesurer une quantité causale sans biais de confusion.
Dans le cadre de la sécurité routière, l'estimation d'une quantité causale est d'autant plus complexe que les données disponibles concernent uniquement des accidentés. Les épidémiologistes du risque routier se heurtent donc à une sélection extrême des individus. Le phénomène de sélection a été très étudié dans la littérature (Robins, 2001, Hernán et al., 2004, Lajous et al., 2014), avec des exemples célèbres comme le biais de Berkson (Berkson, 1946). Ici, nous adopterons la terminologie de Bareinboim and Pearl (2012) et Bareinboim and Tian (2015) où la sélection se réfère à l'inclusion préférentielle de certains individus à partir de la population. Les études cas-témoins sont particulièrement touchées par le biais de sélection étant donné que la sélection des individus dépend de la survenue de l'événement d'intérêt.
Comme mentionné en introduction, pour contourner le problème d'absence de témoins en sécurité routière, il est maintenant classique d'utiliser des analyses en responsabilité (Smith and Popham, 1951, Perchonok, 1978, Terhune, 1986, Brubacher et al., 2014), et de comparer les responsables de l'accident aux non-responsables de l'accident. L'analyse en responsabilité repose notamment sur l'hypothèse que les non-responsables accidentés sont représentatifs de la population des non-responsables ou des non-impliqués dans un accident, selon ce que l'on cherche à estimer. Or cette hypothèse est discutable puisque certaines caractéristiques peuvent être sous ou sur-représentées chez les non-responsables accidentés. Ainsi, la question est de déterminer si les analyses en responsabilité ne laissent pas perdurer un biais de sélection résiduel, compromettant ainsi l'estimation de mesures causales.
L'objectif de ce chapitre est donc de déterminer quels odds ratios causaux sont identifiables et estimables à partir des analyses en responsabilité, et selon les données d'accidentés disponibles. Il s'agit de présenter tout d'abord les récents résultats de la littérature sur l'identifiabilité de

quantités causales en présence de sélection des individus. Nous décrirons ensuite de manière formelle la notion de responsabilité. Nous étudierons enfin l'identifiabilité d'une quantité causale lorsque l'inclusion dans l'étude est conditionnée sur la survenue de l'accident, indépendamment de sa gravité, puis lorsque l'inclusion dans l'étude ne se fait que suite à la survenue d'un accident grave. L'identifiabilité de l'odds ratio causal de l'alcool sur la responsabilité de l'accident sera considérée en tant qu'exemple tout le long de ce chapitre. Nous nous interrogerons enfin sur la validité externe et interne des mesures estimées en présence de sélection préférentielle d'individus.

2.1 Inférence causale en présence de sélection

Dans le chapitre précédent, nous avons considéré l'identifiabilité de quantités causales d'un point de vue général, et selon les facteurs de confusion observés. Dans cette partie, il s'agit d'aborder la question de l'identifiabilité de quantités causales lorsque la sélection des individus de l'étude dépend d'une ou plusieurs variables du DAG.

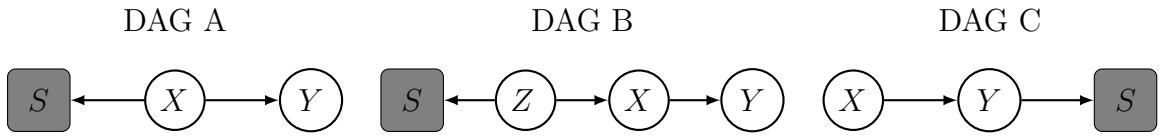
Une nouvelle fois, nous considérons X la variable d'exposition et Y l'événement comme binaires. $\boldsymbol{W}$ représente un vecteur de variables catégorielles. Notons $\boldsymbol{V} = (X, Y, \boldsymbol{W})$ le vecteur de variables observables. Le modèle causal $\mathcal{M}$ conduisant à la survenue de l'événement Y peut être graphiquement représenté par un DAG G. Pour des raisons explicitées plus tard dans ce chapitre, nous considérons l'effet causal défini sur l'échelle de l'odds ratio :

$$COR(X, Y) = \frac{\mathbb{P}(Y_1 = 1)/\mathbb{P}(Y_1 = 0)}{\mathbb{P}(Y_0 = 1)/\mathbb{P}(Y_0 = 0)}. \tag{2.1}$$

Toutefois, l'effet causal se définit de manière similaire sur l'échelle des excès de risque ou des risques relatifs.

En l'absence de sélection (cf. chapitre 1), le principal objectif de l'inférence causale est de déterminer si $\mathbb{P}(Y_x = 1)$ pour $x \in (0, 1)$, et donc une quantité causale, est identifiable selon la distribution des variables observées $\mathbb{P}(\boldsymbol{V} = v)$ (Bareinboim and Tian, 2015). Lorsque tous les facteurs de confusion sont observés, l'hypothèse de cohérence (Coh) et d'ignorabilité conditionnelle (C.Ign) garantissent l'expression de $\mathbb{P}(Y_x = 1)$ et donc des quantités causales à partir des variables observées. Lorsque certains facteurs sont inobservés, des critères comme les critères back-door et front-door permettent selon la structure du DAG, l'identifiabilité de quantités causales (Pearl, 1995, 2009).
En présence de sélection préférentielle d'individus, l'identifiabilité de $\mathbb{P}(Y_x = 1)$ et des quantités causales n'est plus garantie. Notons S la variable binaire indiquant si le sujet est sélectionné dans l'étude. Dans le DAG G, la variable S est ajoutée, constituant le DAG G_s. La représentation de S dans le DAG est classiquement différente de celle des autres variables, car S ne joue pas un rôle dans le mécanisme causal décrit par le DAG G. Elle joue un rôle dans le mécanisme de sélection des individus puisque les données $\boldsymbol{V} = (X, Y, \boldsymbol{W})$ sont disponibles uniquement pour les individus pour lesquels $S = 1$. Quelques exemples simples de phénomène de sélection sans facteur de confusion $\boldsymbol{W}$ sont donnés à travers la figure 2.1.

FIGURE 2.1 – Exemples de DAGs simples, sans facteur de confusion $\boldsymbol{W}$ en présence de processus de sélection



En présence de processus de sélection, l'identifiabilité de $\mathbb{P}(Y_x = 1)$ pour $x \in (0, 1)$ ne se fait plus à partir de la distribution des variables observées $\mathbb{P}(\boldsymbol{V} = v)$, mais à partir des variables observées parmi les individus où $S = 1$, *i.e.* $\mathbb{P}(\boldsymbol{V} = v|S = 1)$ (Bareinboim and Pearl, 2012, Bareinboim and Tian, 2015). La première intuition concernant l'identifiabilité de quantités causales en présence de sélection est donné par le théorème 1 de Bareinboim and Pearl (2012). Selon ce théorème, en l'absence de facteur de confusion $\boldsymbol{W}$, $\mathbb{P}(Y_x = 1)$ est identifiable, i.e. égal à $\mathbb{P}(Y = x|X = x, S = 1)$ si $Y \perp\!\!\!\perp S|X$.

Preuve du théorème 1 de Bareinboim and Pearl (2012) :

$$\begin{aligned}\mathbb{P}(Y_x = 1) &\overset{\text{Ign}}{=} \mathbb{P}(Y_x = 1|X = x)\\ &\overset{\text{Coh}}{=} \mathbb{P}(Y = 1|X = x)\\ &= \mathbb{P}(Y = 1|X = x, S = 1) \quad \text{si } Y \perp\!\!\!\perp S|X\end{aligned}$$

Cette condition $Y \perp\!\!\!\perp S|X$ est notamment vérifiée dans le DAG A et B de la figure 2.1, puisque dans les deux cas X bloque le chemin entre Y et S. En revanche, $Y \not\perp\!\!\!\perp S|X$ dans le DAG C. Or, les études cas-témoins se rapprochent du DAG C, étant donné que la sélection dépend de l'événement d'intérêt Y.

En présence de sélection, on parle plus rigoureusement de « recouvrabilité » (« recoverability ») et non plus d'identifiabilité de quantités causale, et le concept de recouvrabilité est une généralisation de celui d'identifiabilité en présence de sélection différentielle des individus. Cela étant, la traduction française n'étant pas adéquate, nous conserverons le terme d'identifiabilité dans la suite de cette thèse. C'est ensuite la structure du DAG G_s, et non plus celle de G, qui détermine si le mécanisme de sélection conduit ou non à des estimations sans biais de quantités causales (Hernán et al., 2004). Notons qu'en présence de facteurs de confusion, le biais survient lorsqu'il est impossible de conditionner les estimations sur certains facteurs inobservés, alors qu'en présence de sélection d'individus, le biais survient car la distribution des variables est toujours conditionnée sur $S = 1$ (Bareinboim and Pearl, 2012).

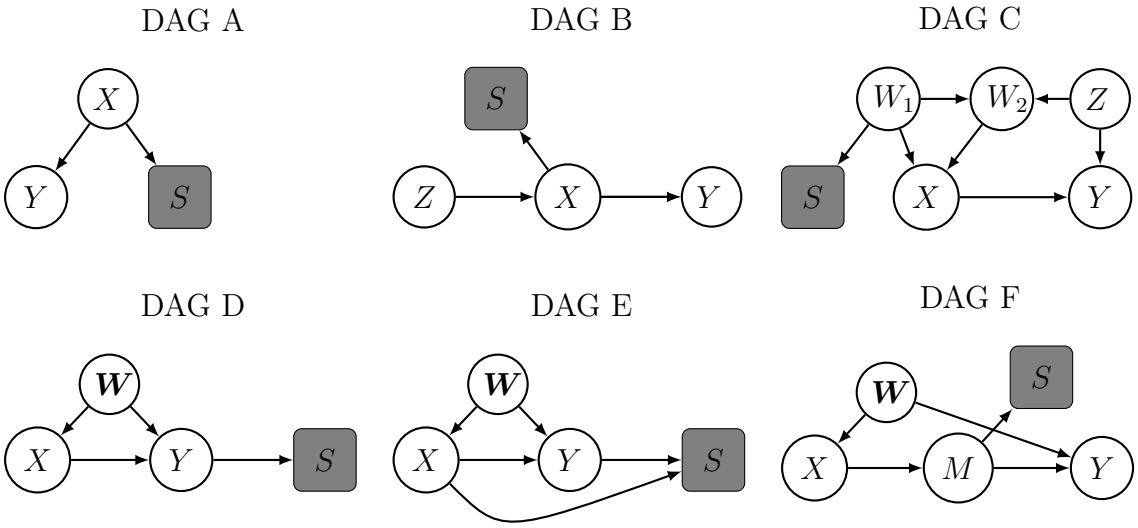
Deux types de résultats permettent de déterminer si des quantités causales sont identifiables à partir de la structure du DAG G_s. Tout d'abord, Bareinboim and Tian (2015) définissent des conditions nécessaires et suffisantes assurant l'identifiabilité de la quantité $\mathbb{P}(Y_x = 1)$, et donc par suite de quantités causales (excès de risque causal, risque relatif causal, odds ratio causal) en présence de sélection. Ensuite, sous certaines hypothèses, l'odds ratio causal peut tout de même être identifiable en présence de processus de sélection, alors que $\mathbb{P}(Y_x = 1)$ ou d'autres quantités causales ne le sont pas (Bareinboim and Pearl, 2012). Nous renvoyons le lecteur à Bareinboim et al. (2014) pour plus de détails sur l'identifiabilité d'autres mesures en présence de mécanisme de sélection.

2.1.1 Identifiabilité de $\mathbb{P}(Y_x = 1)$ en présence de sélection

Notons comme précédemment G le DAG causal, composé des variables observées $\boldsymbol{V}$, et par G_s le DAG obtenu après l'ajout de la variable S. Pour tout sous-ensemble de variables $\boldsymbol{C} \subseteq \boldsymbol{V}$, on note $G_{\boldsymbol{C}}$ le sous-graphe de G restreint aux variables de $\boldsymbol{C}$, c'est-à-dire où l'on a retiré les noeuds n'appartenant pas à $\boldsymbol{C}$ et les flèches en provenance ou à destination en dehors de $\boldsymbol{C}$. Pour tout $V_i \in \boldsymbol{V}$, on note par ailleurs $An(V_i)_G$ l'union de V_i et de ses ancêtres dans le DAG G. Le théorème 2 de Bareinboim and Tian (2015) établit alors que $\mathbb{P}(Y_x = y)$ est identifiable si et seulement si

$$An(Y)_{G_{\boldsymbol{V}\setminus X}} \cap An(S)_{G_s} = \emptyset. \tag{C.1}$$

FIGURE 2.2 – Exemples de DAGs en présence de processus de sélection



La condition (C.1) est notamment vérifiée dans les DAGs A et B de la figure 2.2. Dans les DAGs A et B, Y n'a pas d'ancêtre dans le DAG $G_{\boldsymbol{V}\setminus X}$ et $Y \notin An(S)_{G_s}$, et par conséquent la condition (C.1) est vérifiée. Dans le DAG C, Z est un ancêtre de Y dans le DAG $G_{\boldsymbol{V}\setminus X}$, mais Z n'est pas un ancêtre de S dans le DAG G_s, donc la condition (C.1) est également vérifiée. Dans les analyses en responsabilité, la sélection préférentielle des individus s'effectuant sur l'événement d'intérêt, l'accident, nous concentrerons par la suite nos exemples aux cas où la sélection dépend de l'événement Y, comme dans les DAG D et E de la figure 2.2. Nous considérons aussi le cas où S est un descendant d'un médiateur, ou facteur intermédiaire, comme dans le DAG F de la figure 2.2. La condition (C.1) n'est clairement pas vérifiée dans les trois DAGs D, E, F de la figure 2.2. Par exemple, dans les DAGs D et E nous avons $Y \in An(Y)_{G_{\boldsymbol{V}\setminus X}} \cap An(S)_{G_s}$. Elle n'est pas non plus vérifiée dans le DAG F où $M \in An(Y)_{G_{\boldsymbol{V}\setminus X}} \cap An(S)_{G_s}$. Par conséquent, $\mathbb{P}(Y_x = 1)$ n'est pas identifiable sous ces types de processus de sélection.

Remarque

Notons enfin que le théorème 2 de Bareinboim and Tian (2015) n'affirme pas que $\mathbb{P}(Y_x = y)$ n'est jamais identifiable selon les structures causales décrites par les DAG D, E et F de la figure 2.2. $\mathbb{P}(Y_x = y)$ n'est pas identifiable sans d'autres hypothèses ou informations non contenues par la structure du DAG G_s. Par exemple, le corollaire 4 de Bareinboim and Pearl (2012) établit

que $\mathbb{P}(Y_x = y)$ peut parfois être identifiable sous les structures causales du DAG D et E lorsque $\mathbb{P}(\boldsymbol{W})$ est connue et quand la dimension du vecteur $\boldsymbol{W}$ est égale à celle de $X \cup Y$. Considérons également le DAG C de la figure 2.1, qui équivaut au DAG D sans facteur de confusion de la figure 2.2. Le théorème 1 de (Bareinboim and Pearl, 2012) établit que $\mathbb{P}(Y_x = y)$ n'est pas identifiable. La condition (C.1) n'est également pas vérifiée et le théorème 2 de Bareinboim and Tian (2015) établit que $\mathbb{P}(Y_x = y)$ n'est pas identifiable d'après la structure du DAG G_s. Pour autant, la connaissance de $\mathbb{P}(X)$ est suffisante pour rendre identifiable $\mathbb{P}(Y_x = y)$, comme montré dans la section 3 de Bareinboim and Pearl (2012).

2.1.2 Identifiabilité de l'odds ratio causal en présence de biais de sélection

Supposons ici que $\boldsymbol{W}$ est un vecteur de facteurs de confusion. Nous considérons ici l'identifiabilité de l'odds ratio causal conditionné sur $\boldsymbol{W} = \boldsymbol{w}$ défini de la façon suivante :

$$COR(X, Y|\boldsymbol{W} = \boldsymbol{w}) = \frac{\mathbb{P}(Y_1 = 1|\boldsymbol{W} = \boldsymbol{w})/\mathbb{P}(Y_1 = 0|\boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(Y_0 = 1|\boldsymbol{W} = \boldsymbol{w})/\mathbb{P}(Y_0 = 0|\boldsymbol{W} = \boldsymbol{w})}.$$

L'odds ratio causal conditionné sur $\boldsymbol{W} = \boldsymbol{w}$, appelé également l'odds ratio causal $\boldsymbol{w}$-spécifique, correspond à l'odds ratio causal dans la strate de la population où $\boldsymbol{W} = \boldsymbol{w}$, si les variables composant le vecteur $\boldsymbol{W}$ sont discrètes. Quand $\boldsymbol{W}$ contient un ensemble suffisant de facteurs de confusion, l'hypothèse d'ignorabilité conditionnelle $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$ est vérifiée (Pearl, 2009, 2000, VanderWeele and Robins, 2009, VanderWeele and Shpitser, 2011), et donc $\mathbb{P}(Y_x = 1|\boldsymbol{W} = \boldsymbol{w}) = \mathbb{P}(Y = 1|X = x, \boldsymbol{W} = \boldsymbol{w})$. Dans ce cas, l'odds ratio causal $\boldsymbol{w}$-spécifique est donc égal à l'odd ratio ajusté sur $\boldsymbol{W} = \boldsymbol{w}$:

$$OR(X, Y|\boldsymbol{W} = \boldsymbol{w}) = \frac{\mathbb{P}(Y = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})/\mathbb{P}(Y = 0|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(Y = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})/\mathbb{P}(Y = 0|X = 0, \boldsymbol{W} = \boldsymbol{w})}.$$

D'après la définition 2 de Bareinboim and Pearl (2012), $COR(X, Y|\boldsymbol{W} = \boldsymbol{w})$ et donc $OR(X, Y|\boldsymbol{W} = \boldsymbol{w})$ sont identifiables en présence de biais de sélection si les hypothèses induites par la structure du DAG le rendent exprimables en fonction de la distribution de $\boldsymbol{V}|S = 1$. Bien entendu, lorsque (C.1) est vérifiée, et donc que $\mathbb{P}(Y_x = y)$ est identifiable, alors $COR(X, Y|\boldsymbol{W} = \boldsymbol{w})$ et $OR(X, Y|\boldsymbol{W} = \boldsymbol{w})$ sont également identifiables. La symétrie de l'odd ratio, $OR(X, Y|\boldsymbol{W} = \boldsymbol{w}) = OR(Y, X|\boldsymbol{W} = \boldsymbol{w})$, le rend identifiable dans certaines situations où $\mathbb{P}(Y_x = y)$ ne l'est pas. Plus précisément, le théorème 1 de Bareinboim and Pearl (2012), ou encore le corollaire 4 de Didelez et al. (2010), établit que $COR(X, Y|\boldsymbol{W} = \boldsymbol{w})$ est identifiable en présence de processus de sélection si et seulement si

$$\text{(C.2)} \qquad X \perp\!\!\!\perp S|(Y, \boldsymbol{W}) \quad \text{ou} \quad Y \perp\!\!\!\perp S|(X, \boldsymbol{W}).$$

Preuve de (C.2) :

$$\begin{aligned}
COR(X,Y|\boldsymbol{W}=\boldsymbol{w}) &= \frac{\mathbb{P}(Y_1=1|\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(Y_1=0|\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(Y_0=1|\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(Y_0=0|\boldsymbol{W}=\boldsymbol{w})} \\
&= \frac{\mathbb{P}(Y=1|X=1,\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(Y=0|X=1,\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(Y=1|X=0,\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(Y=0|X=0,\boldsymbol{W}=\boldsymbol{w})} \text{ si (C.Ign) et (Coh)} \\
&= OR(X,Y|\boldsymbol{W}=\boldsymbol{w}) \\
&= \frac{\mathbb{P}(Y=1|X=1,\boldsymbol{W}=\boldsymbol{w},S=1)/\mathbb{P}(Y=0|X=1,\boldsymbol{W}=\boldsymbol{w},S=1)}{\mathbb{P}(Y=1|X=0,\boldsymbol{W}=\boldsymbol{w},S=1)/\mathbb{P}(Y=0|X=0,\boldsymbol{W}=\boldsymbol{w},S=1)} \\
&\quad \text{si } Y \perp\!\!\!\perp S|(X,\boldsymbol{W})
\end{aligned}$$

De plus,

$$\begin{aligned}
OR(X,Y|\boldsymbol{W}=\boldsymbol{w}) &= OR(Y,X|\boldsymbol{W}=\boldsymbol{w}) \\
&= \frac{\mathbb{P}(X=1|Y=1,\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(X=0|Y=1,\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(X=1|Y=0,\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(X=0|Y=0,\boldsymbol{W}=\boldsymbol{w})} \\
&= \frac{\mathbb{P}(X=1|Y=1,\boldsymbol{W}=\boldsymbol{w},S=1)/\mathbb{P}(X=0|Y=1,\boldsymbol{W}=\boldsymbol{w},S=1)}{\mathbb{P}(X=1|Y=0,\boldsymbol{W}=\boldsymbol{w},S=1)/\mathbb{P}(X=0|Y=0,\boldsymbol{W}=\boldsymbol{w},S=1)} \\
&\quad \text{si } X \perp\!\!\!\perp S|(Y,\boldsymbol{W})
\end{aligned}$$

Cette preuve concerne la condition suffisante. Pour la condition nécessaire, nous laissons le lecteur se référer au materiel supplementaire de Bareinboim and Pearl (2012).

Nous nous focaliserons ici sur les DAGs D, E, F où $\mathbb{P}(Y_x = 1)$ n'était pas identifiable via la condition (C.1). La condition (C.2) est vérifiée sous le DAG D où $X \perp\!\!\!\perp S|(Y,\boldsymbol{W})$, mais n'est pas garantie sous les DAGs E et F. Ainsi, $COR(X,Y|\boldsymbol{W}=\boldsymbol{w})$ est identifiable sous le DAG D, mais ne l'est pas sous les DAGs E et F.

2.2 Application aux analyses en responsabilité

Ayant énoncé les conditions d'identifiabilité en présence de sélection préférentielle des individus, il s'agit maintenant de déterminer si ces conditions sont vérifiées dans le schéma d'étude particulier des analyses en responsabilité. A travers les parties suivantes, nous considérons l'identifiabilité de l'odds ratio causal de l'alcool X sur la responsabilité de l'accident parmi les conducteurs accidentés. Nous commencerons par préciser la notion de responsabilité d'un point de vue verbale puis formelle. Nous étudierons ensuite l'identifiabilité de certains odds ratio causaux dans les analyses en responsabilité dans deux situations : (i) lorsque les données disponibles concernent tous les accidents, indépendamment de leur gravité, (ii) lorsque seules les données concernant des accidents graves sont disponibles. Enfin, dans les cas où certains odds ratios causaux ne sont pas identifiables, nous nous interrogerons sur la validité externe et interne des mesures estimées.

2.2.1 Description formelle de la notion de responsabilité

Commençons d'abord par définir ce que signifie « être responsable » dans les analyses en responsabilité. Dans ce contexte, la notion de responsabilité est très différente de la conception légale ou assurantielle. Un conducteur impliqué dans un accident est considéré comme responsable de l'accident s'il a commis une faute de conduite (dépassement de ligne blanche, non respect des feux tricolores, conduite en sens inverse...) qui a déclenché l'accident. La responsabilité de l'accident, qui est assignée après que l'accident a lieu, est une mesure de la faute de conduite, qui elle a eu lieu avant l'accident. Un taux d'alcoolémie supérieur au seuil légal, même s'il constitue une infraction au code de la route, n'intervient pas dans la détermination de la responsabilité.

Nous notons F la variable binaire indiquant si le conducteur a commis une faute de conduite. Notons que cette variable est définie pour tous les conducteurs, pas seulement ceux impliqués dans un accident. A est la variable binaire indiquant si le conducteur est impliqué dans un accident. Considérons également la variable binaire R indiquant si le conducteur est responsable de l'accident. Par conséquent, $R = F \times A$, et nous avons :

$$\begin{cases} R = 1 & \text{si et seulement si } A = 1 \text{ et } F = 1 \\ R = 0 & \text{si } A = 1 \text{ et } F = 0 \\ R = 0 & \text{si } A = 0 \text{ même si } F = 1 \end{cases}$$

Lorsque les données sont conditionnées sur le fait d'être accidenté $A = 1$ alors $F = R$.
Notons que ni la responsabilité R ni la faute de conduite F ne sont des événements substitutifs à l'accident. En effet,

- le fait d'avoir commis une faute, n'implique pas nécessairement l'implication dans un accident ($F = 1 \not\Rightarrow A = 1$),
- le fait de ne pas être responsable d'un accident, n'implique pas le fait de ne pas avoir d'accident ($R = 0 \not\Rightarrow A = 0$),
- le fait de ne pas être responsable d'un accident, n'implique pas le fait de ne pas avoir commis de faute de conduite ($R = 0 \not\Rightarrow F = 0$).

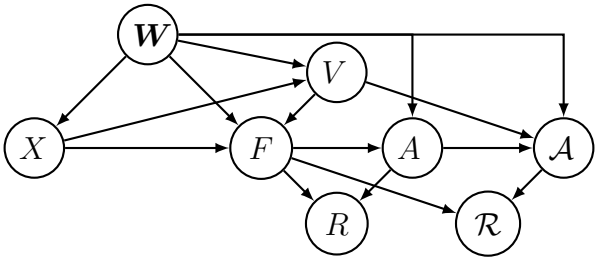
Dans la plupart des cas, les données disponibles dépendent de la gravité de l'accident, qui est reliée à la vitesse. Définissons donc $\mathcal{A}$ la variable binaire indiquant si le conducteur est impliqué dans un accident grave, *i.e.* un accident corporel ou mortel typiquement. Ensuite, nous pouvons définir de façon analogue la variable binaire $\mathcal{R}$ indiquant si le conducteur est responsable d'un accident grave. Ainsi, $\mathcal{R} = F \times \mathcal{A}$.

Le DAG G ci-dessous représente une version simplifiée du mécanisme causal conduisant potentiellement à un accident ou à un accident grave, du point de vue d'un usager de la route donné, quel que soit le nombre de véhicules impliqués dans l'accident. Comme déjà mentionné, un accident vient généralement d'une faute de conduite de l'usager, d'où la flèche de F vers A. Cette faute F est plus souvent impactée par le facteur d'exposition X que l'on étudie (substances psychoactives, hypovigilance...), d'où la flèche entre X et F. C'est en tant que cause de la faute de conduite que les facteurs X étudiés sont généralement qualifiés de causes de l'accident. Ici, la cause étudiée X est la variable binaire indiquant si le conducteur a une concentration d'alcool

supérieure au seuil légal. Le facteur d'exposition X et la faute F sont impactés par un ensemble de facteurs de confusion $\boldsymbol{W}$. $\boldsymbol{W}$ représente le vecteur de tous les facteurs de confusion (âge et sexe du conducteur, moment de l'accident, type de route, limitation de vitesse...). Les jeunes conducteurs ont par exemple plus de risque de conduire avec un taux d'alcoolémie positif, et ont également plus de risque de commettre une faute de conduite du fait de leur moindre expérience. Dans la mesure où X correspond à l'alcoolémie, et qu'il est communément admis que l'alcool augmente la vitesse de conduite, la vitesse est ici considérée comme un facteur intermédiaire ou médiateur dans la relation entre X et F. Notons que la vitesse est un médiateur pour de nombreux facteurs étudiés (cannabis, téléphone au volant...). Ajoutons que c'est une variable rarement disponible. Il existe ainsi une flèche entre X et V, V étant la variable binaire indiquant si le conducteur roule à une vitesse élevée. V est également impactée par $\boldsymbol{W}$ puisque la vitesse est liée à l'âge, sexe du conducteur, au type de route... On distingue ici la vitesse élevée de la vitesse excessive. La première constituant une variable à part entière dans le DAG, la seconde constituant une faute de conduite et donc étant incluse dans la variable F. Cette différence explique notamment le fait qu'il n'y ait pas de flèche entre V et A, mais qu'il y ait une flèche entre V et $\mathcal{A}$. En effet, si la vitesse élevée était considérée comme une cause de l'accident A, cela voudrait dire que la vitesse n'était pas appropriée, et donc qu'une faute de conduite aurait été commise. Par conséquent, V est une cause de A seulement à travers F, d'où la flèche entre V et F et l'absence de flèche entre V et A. Cependant, V a un effet sur $\mathcal{A}$, car la vitesse a un effet sur la gravité, sur la gravité lésionnelle de l'accident. Enfin, précisons l'absence de flèche entre X et R ou $\mathcal{R}$. Il est en effet admis que la responsabilité est assignée en ne prenant pas en considération toute information concernant X (Brubacher et al., 2014).
L'information relative à une faute de conduite commise par un autre usager, ou à un accident provoqué par une condition externe (traversée d'un animal, chute d'un arbre...) est également contenue dans le DAG présenté. Cette information est contenue dans le vecteur $\boldsymbol{W}$, puisqu'on fait l'hypothèse ici que $\boldsymbol{W}$ contient tous les facteurs de confusion possibles (moment de la journée, densité du trafic, type de véhicule, type de route...). La densité du trafic fournit l'information sur une faute commise par un autre usager, puisqu'il est en effet plus probable de croiser un autre conducteur aux horaires de pointe. Le type de route contient quant à lui l'information sur un accident provoqué par une condition externe, ce type d'incident ayant sûrement plus de risque de survenir sur les routes de campagne. Dans les deux cas, la densité du trafic ou le type de route a bien un impact sur la faute F, le fait de rouler à une vitesse élevée V et d'avoir un accident A ou un accident grave $\mathcal{A}$. Bien entendu, chacun des facteurs composant $\boldsymbol{W}$ n'a pas forcément d'impact à la fois sur X, V et A ou $\mathcal{A}$, mais dans la mesure où l'information principale d'un DAG réside dans l'absence de flèche, nous supposons qu'il existe les quatre flèches de $\boldsymbol{W}$ vers X, F, V, A ou $\mathcal{A}$.

Dans les parties suivantes, nous appliquons les principes de l'inférence causale en présence de sélection présentés en section 2.1 dans deux situations : *(i)* dans un cas idéal, même hypothétique, où l'inclusion dans l'étude est conditionnée à la survenue d'un accident, quelle que soit sa gravité $(S = 1) \Longrightarrow (A = 1)$ et odds ratio causal d'intérêt correspond à l'effet de X sur le risque d'être responsable d'un accident, $\mathbb{P}(R = 1)$ et *(ii)* dans le cas le plus commun où l'inclusion dans l'étude ne se fait que suite à un accident grave $(S = 1) \Longrightarrow (\mathcal{A} = 1)$ et l'odds ratio causal d'intérêt correspond à l'effet de X sur le risque d'être responsable d'un accident grave, $\mathbb{P}(\mathcal{R} = 1)$. Deux DAGs issus du précédent peuvent représenter ces deux cas et sont donnés par la figure 2.4. Dans le cas *(i)*, puisque V n'a pas d'effet sur A, l'effet de X sur A est entièrement médié par F,

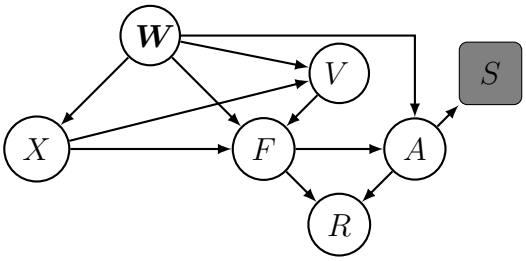
FIGURE 2.3 – DAG G représentant le mécanisme causal



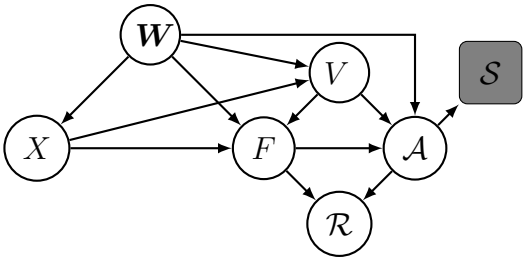
pendant que, dans le cas *(ii)*, l'effet de X sur $\mathcal{A}$ est médié à la fois par F et V.

FIGURE 2.4 – DAGs G_s dans les analyses en responsabilité

(i) données concernant l'ensemble des accidents



(ii) données concernant uniquement les accidents graves



X alcool, F faute de conduite, A accident, $\mathcal{A}$ accident grave, R responsabilité, $\mathcal{R}$ responsabilité d'un accident grave, V vitesse élevée, $\boldsymbol{W}$ ensemble des facteurs de confusion, S sélection sur l'accident, $\mathcal{S}$ sélection sur l'accident grave.

2.2.2 Application à l'analyse en responsabilité lorsque l'inclusion ne dépend pas de la gravité de l'accident

Le DAG de gauche de la figure 2.4 représente le mécanisme causal lorsque les données disponibles décrivent tous les types d'accidents, des plus légers aux plus graves, et que le processus d'inclusion ne dépend pas de leur gravité. Dans le cas *(i)*, l'effet de X sur A est entièrement médié par F et $X \perp\!\!\!\perp A|(F, W)$. L'ensemble $(F, \boldsymbol{W})$ bloque tous les chemins entre X et A.

Il s'agit de considérer tout d'abord l'identifiabilité de $\mathbb{P}(R_x = r)$ pour $(x, r) \in (0, 1)$, la probabilité d'être ou non responsable d'un accident dans le monde contrefactuel où $X = x$. Les résultats de la section 2.1 montrent que $\mathbb{P}(R_x = r)$ n'est pas identifiable puisque $A \in An(R)_{G_{\boldsymbol{V}\setminus X}} \cap An(S)_{G_s}$. Considérons maintenant l'identifiabilité de l'odd ratio causal $\boldsymbol{w}$-specifique $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$. $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ n'est pas identifiable puisque ni $X \perp\!\!\!\perp S|(R, \boldsymbol{W})$ ni $R \perp\!\!\!\perp S|(X, \boldsymbol{W})$ ne sont vérifiées. En utilisant le critère de d-séparation (cf. chapitre 1), d'une part, $X \not\perp\!\!\!\perp S|(R, \boldsymbol{W})$ car R est un collider sur le chemin $X \longrightarrow F \longrightarrow R \longleftarrow S$. D'autre part, $R \not\perp\!\!\!\perp S|(X, \boldsymbol{W})$, car $(X, \boldsymbol{W})$ ne bloque pas tous les chemins entre R et S. $(X, \boldsymbol{W})$ ne bloque notamment pas le chemin

$R \longleftarrow A \longrightarrow S$. Néanmoins, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ est identifiable puisque $X \perp\!\!\!\perp S|(F, \boldsymbol{W})$, car $X \perp\!\!\!\perp A|(F, \boldsymbol{W})$. Par conséquent, nous avons $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ qui vaut la quantité suivante :

$$= \frac{\mathbb{P}(F=1|X=1,\boldsymbol{W}=\boldsymbol{w},A=1)/\mathbb{P}(F=0|X=1,\boldsymbol{W}=\boldsymbol{w},A=1)}{\mathbb{P}(F=1|X=0,\boldsymbol{W}=\boldsymbol{w},A=1)/\mathbb{P}(F=0|X=0,\boldsymbol{W}=\boldsymbol{w},A=1)} \\ \frac{\mathbb{P}(R=1|X=1,\boldsymbol{W}=\boldsymbol{w},A=1)/\mathbb{P}(R=0|X=1,\boldsymbol{W}=\boldsymbol{w},A=1)}{\mathbb{P}(R=1|X=0,\boldsymbol{W}=\boldsymbol{w},A=1)/\mathbb{P}(R=0|X=0,\boldsymbol{W}=\boldsymbol{w},A=1)},$$

où la dernière égalité vient du fait que $F = R$ dans la sous-population où $\{A = 1\}$.

Par ailleurs, si les prévalences $\mathbb{P}(F_x = 1|\boldsymbol{W} = \boldsymbol{w})$ et $\mathbb{P}(R_x = 1|\boldsymbol{W} = \boldsymbol{w})$ sont faibles pour tout $x \in \{0, 1\}$, nous obtenons l'égalité suivante :

$$COR(X, R|\boldsymbol{W} = \boldsymbol{w}) \approx COR(X, F|\boldsymbol{W} = \boldsymbol{w}). \tag{2.2}$$

Preuve de l'approximation (2.2) :

$$\begin{aligned}
COR(X, R|\boldsymbol{W} = \boldsymbol{w}) &\approx CRR(X, R|\boldsymbol{W} = \boldsymbol{w}) \text{ si } \mathbb{P}(R_x = 1|\boldsymbol{W} = \boldsymbol{w}) \text{ petit} \\
&= \frac{\mathbb{P}(R_1 = 1|\boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(R_0 = 1|\boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F_1 = 1, A_1 = 1|\boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F_0 = 1, A_0 = 1|\boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F_1 = 1, A_1 = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F_0 = 1, A_0 = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F = 1, A = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F = 1, A = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \text{ puisque } (A_x, F_x) \perp\!\!\!\perp X|W \\
&= \frac{\mathbb{P}(A = 1|F = 1, X = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(A = 1|F = 1, X = 0, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(A = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(A = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \text{ puisque } X \perp\!\!\!\perp A|(F, \boldsymbol{W}) \\
&= \frac{\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F_1 = 1|\boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F_0 = 1|\boldsymbol{W} = \boldsymbol{w})} \text{ puisque } F_x \perp\!\!\!\perp X|W \\
&= CRR(X, F|\boldsymbol{W} = \boldsymbol{w}) \\
&\approx COR(X, F|\boldsymbol{W} = \boldsymbol{w}) \text{ si } \mathbb{P}(F_x = 1|\boldsymbol{W} = \boldsymbol{w}) \text{ petit}
\end{aligned}$$

Et par conséquent, $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$, qui n'est pas identifiable, peut être approximé par $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ qui est lui identifiable puisque $X \perp\!\!\!\perp S|(F, \boldsymbol{W})$.

Les conditions de l'identifiabilité des odds ratios causaux concernent les odds ratios causaux conditionnels. On peut noter que l'approximation de l'odd ratio causal $\boldsymbol{w}$-specifique $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ n'est pas suffisante pour approximer l'odd ratio causal dans l'ensemble de la population $COR(X, R)$. Néanmoins, l'approximation de $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ reste intéressante puisqu'elle permet par exemple de calculer des fractions de risque attribuable (PAF), et donc établir une

hiérarchisation des causes des accidents de la circulation routière, dans le cas présent où des données concernant l'ensemble des accidents, indépendamment de leur gravité, sont disponibles. En effet, le PAF se définit usuellement de la façon suivante (Benichou, 2001, Bruzzi et al., 1985) :

$$\begin{aligned}\text{PAF} = \frac{\mathbb{P}(R=1) - \mathbb{P}(R_0=1)}{\mathbb{P}(R=1)} &= \frac{\mathbb{P}(R=1) - \sum_{\boldsymbol{w}} \mathbb{P}(R=1|X=0,\boldsymbol{W}=\boldsymbol{w})\mathbb{P}(\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(R=1)} \\ &= 1 - \sum_{i=0}^{1}\sum_{\boldsymbol{w}} \frac{p_{i,\boldsymbol{w}}}{RR_{i\boldsymbol{w}}} \\ &= \mathbb{P}(X=1|R=1) - \sum_{\boldsymbol{w}} \frac{p_{1\boldsymbol{w}}}{RR_{1\boldsymbol{w}}}. \end{aligned} \tag{2.3}$$

où $p_{i,\boldsymbol{w}} = \mathbb{P}(X=i, \boldsymbol{W}=\boldsymbol{w}|R=1)$ et $RR_{i\boldsymbol{w}} = \frac{\mathbb{P}(R=1|X=i,\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(R=1|X=0,\boldsymbol{W}=\boldsymbol{w})}$.
Le fait d'être responsable d'un accident implique qu'il ait eu un accident, ie $R=1 \Rightarrow A=1$. Nous avons alors $p_{1,\boldsymbol{w}} = \mathbb{P}(X=1, \boldsymbol{W}=\boldsymbol{w}|R=1) = \mathbb{P}(X=1, \boldsymbol{W}=\boldsymbol{w}|R=1, A=1)$. $p_{1,\boldsymbol{w}}$ et $\mathbb{P}(X=1|R=1)$ peuvent donc être estimées sans biais à partir de données décrivant uniquement la sous-population d'accidentés $\{A=1\}$. D'autre part, le risque relatif ajusté sur W, $RR_{1\boldsymbol{w}}$, peut être approximé par l'odd ratio ajusté $OR(X,R|\boldsymbol{W}=\boldsymbol{w})$, sous l'hypothèse, qui semble naturelle, selon laquelle $\mathbb{P}(R=1|X=x,\boldsymbol{W}=\boldsymbol{w})$ est petit pour tout $x \in \{0,1\}$ et $\boldsymbol{w}$. Par conséquent, sous ces hypothèses, toutes les quantités de l'équation (2.3) peuvent être estimées sans biais ou bien approximées, et par conséquent les fractions de risque attribuables peuvent également être approximées dans le cas *(i)* où des données concernant l'ensemble des accidents, indépendamment de leur gravité, seraient disponibles.

2.2.3 Application à l'analyse en responsabilité lorsque l'inclusion dépend de la gravité de l'accident

En sécurité routière, la plupart des données disponibles concernent uniquement des accidents graves. Donnant lieu à un procès-verbal, la totalité des accidents mortels sont disponibles ainsi que les accidents corporels les plus graves. Les accidents corporels sont en effet sous-enregistrés car les données sur les blessés plus légers ne sont souvent pas disponibles. De même, les accidents matériels ne sont pas enregistrés, et si c'était le cas, seuls les plus graves seraient sans doute à disposition. Ainsi, l'inclusion des individus dépend de la gravité de l'accident, qui est elle-même essentiellement déterminée par la vitesse. Le DAG de droite de la figure 2.4 représente le mécanisme causal lorsque seules les données sur des accidents graves sont disponibles. Comme précédemment, $\mathcal{A}$ correspond à l'occurrence d'un accident grave, sur laquelle la sélection des données s'effectue. Nous avons alors $\mathcal{S}=1 \Rightarrow \mathcal{A}=1$. Dans le cas *(ii)*, l'effet de V sur $\mathcal{A}$ n'est plus entièrement médié par F, et la condition $X \perp\!\!\!\perp \mathcal{A}|(F,\boldsymbol{W})$ n'est plus vérifiée.

Comme dans la partie précédente, ni $\mathbb{P}(\mathcal{R}_x=r)$ ni $COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w})$ ne sont identifiables. Respectivement, car $\mathcal{A} \in An(\mathcal{R})_{G_{\boldsymbol{V}\setminus X}} \cap An(S)_{G_s}$, et parce qu'aucune des conditions suivantes n'est vérifiée : $X \perp\!\!\!\perp \mathcal{S}|(\mathcal{R},\boldsymbol{W})$ ou $\mathcal{R} \perp\!\!\!\perp \mathcal{S}|(X,\boldsymbol{W})$. Puisque $\mathcal{R}$ est un collider, alors $X \not\perp\!\!\!\perp S|(\mathcal{R},\boldsymbol{W})$. D'autre part, $\mathcal{R} \not\perp\!\!\!\perp \mathcal{S}|(X,\boldsymbol{W})$, car une nouvelle fois $(X,\boldsymbol{W})$ ne bloque pas tous les chemins entre $\mathcal{R}$ et $\mathcal{S}$. $(X,\boldsymbol{W})$ ne bloque notamment pas le chemin $\mathcal{R} \longleftarrow F \longrightarrow \mathcal{A} \longrightarrow \mathcal{S}$. De plus, dans le cas *(ii)*, $COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w})$ ne peut plus s'approximer par une autre odds ratio causal identifiable. D'une part, l'équation (2.2) n'est plus valable car $X \perp\!\!\!\perp \mathcal{A}|(F,\boldsymbol{W})$ n'est pas vérifiée dans le DAG

G_s correspondant au cas *(ii)*. D'autre part, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ n'est plus identifiable car ni $X \perp\!\!\!\perp \mathcal{A}|(F, \boldsymbol{W})$ ni $F \perp\!\!\!\perp \mathcal{A}|(X, \boldsymbol{W})$ n'est vérifiée d'après la structure du DAG G_s. $(F, \boldsymbol{W})$ ne bloque pas tous les chemins entre X et $\mathcal{A}$, puisqu'il ne bloque pas le chemin $X \longrightarrow V \longrightarrow \mathcal{A}$. De la même manière, $(X, \boldsymbol{W})$ ne bloque pas le chemin entre $F \longleftarrow V \longrightarrow \mathcal{A}$. Notons que s'il n'y avait pas de flèche entre X et V, on aurait alors $X \perp\!\!\!\perp \mathcal{A}|(F, \boldsymbol{W})$, et $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ serait identifiable, et $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$ pourrait alors être approximé. Dans le cas où X représente l'alcool, mais également d'autres facteurs comme le cannabis, le téléphone au volant..., cette absence de flèche entre X et V est peu vraisemblable.

Lorsqu'une quantité causale n'est pas identifiable, une solution alternative a été proposée par Hernán et al. (2004). Elle requiert notamment la connaissance ou l'estimation de la probabilité d'inclusion dans l'étude. Dans les analyses en responsabilité, le critère d'inclusion est la survenue de l'accident, et donc la probabilité d'inclusion reposerait au moins en partie sur la probabilité d'avoir un accident $\mathbb{P}(A = 1|X = x)$. Bien entendu, cette donnée n'est pas disponible puisqu'elle constitue la quantité d'intérêt que l'on souhaiterait estimer si l'on disposait d'individus non-accidentés. Si elle était disponible, les épidémiologistes du risque routier n'aurait sans doute pas recours aux analyses en responsabilité.

2.2.4 Validité externe et interne des mesures estimables

Lorsque l'odd ratio causal $\boldsymbol{w}$-spécifique n'est pas identifiable, la plupart des épidémiologistes considèrent que les mesures d'associations qu'ils estiment sur la population $S = 1$ ont une valeur causale uniquement sur $S = 1$. Ils considèrent ainsi la validité interne de leurs résultats et questionnent généralement seulement la validité externe. Or, une définition plus formelle de la validité interne et externe peut être fournie en utilisant les principes de l'inférence causale, et ainsi déterminer le type de validité des mesures estimées sur $S = 1$.

Rappelons la définition de la mesure d'association estimable en présence de sélection préférentielle d'individus :

$$OR(X, Y|\boldsymbol{W} = \boldsymbol{w}, S = 1) = \frac{\mathbb{P}(Y = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}, S = 1)/\mathbb{P}(Y = 0|X = 1, \boldsymbol{W} = \boldsymbol{w}, S = 1)}{\mathbb{P}(Y = 1|X = 0, \boldsymbol{W} = \boldsymbol{w}, S = 1)/\mathbb{P}(Y = 0|X = 0, \boldsymbol{W} = \boldsymbol{w}, S = 1)}.$$

Formellement, nous considèrerons

— la validé externe de $OR(X, Y|\boldsymbol{W} = \boldsymbol{w}, S = 1)$, si

$$OR(X, Y|\boldsymbol{W} = \boldsymbol{w}, S = 1) = COR(X, Y|W = w),$$

— la validité interne de $OR(X, Y|\boldsymbol{W} = \boldsymbol{w}, S = 1)$, si

$$OR(X, Y|\boldsymbol{W} = \boldsymbol{w}, S = 1) = COR(X, Y|W = w, S = 1),$$

$$\text{avec } COR(X, Y|\boldsymbol{W} = \boldsymbol{w}, S = 1) = \frac{\mathbb{P}(Y_1 = 1|\boldsymbol{W} = \boldsymbol{w}, S = 1)/\mathbb{P}(Y_1 = 0|\boldsymbol{W} = \boldsymbol{w}, S = 1)}{\mathbb{P}(Y_0 = 1|\boldsymbol{W} = \boldsymbol{w}, S = 1)/\mathbb{P}(Y_0 = 0|\boldsymbol{W} = \boldsymbol{w}, S = 1)}.$$

Si $OR(X, Y|\boldsymbol{W} = \boldsymbol{w}, S = 1) = COR(X, Y|\boldsymbol{W} = \boldsymbol{w})$ alors l'OR estimable a bien une valeur cau-

sale. Or, $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)=COR(X,Y|\boldsymbol{W}=\boldsymbol{w})$ implique que $COR(X,Y|\boldsymbol{W}=\boldsymbol{w})$ est identifiable. Donc si $COR(X,Y|W=w)$ n'est pas identifiable, la mesure estimée $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ n'a pas de valeur causale en dehors de la population d'étude $\{S=1\}$. Si l'on se réfère à la définition de Rothman et al. (2008), cela signifie que $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ n'a pas de validité externe, c'est-à-dire que les résultats ne peuvent se généraliser à la population cible. Formellement, nous dirons que $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ est valide d'un point de vue externe, si cet odd ratio ajusté correspond à l'odd ratio causal $\boldsymbol{w}$-spécifique dans la population cible, ie $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)=COR(X,Y|\boldsymbol{W}=\boldsymbol{w})$.
Ainsi dans le DAG D de la figure 2.2, $COR(X,Y|\boldsymbol{W}=\boldsymbol{w})$ est identifiable et vaut $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$, qui est donc valide d'un point de vue externe. Dans le DAG E et F de la figure 2.2, $COR(X,Y|\boldsymbol{W}=\boldsymbol{w})$ n'est pas identifiable, donc $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ n'a pas de validité externe.
Considérons maintenant le cas des analyses en responsabilité. Dans le cas (i) où le processus d'inclusion ne dépend pas de la gravité de l'accident, nous avons montré que $COR(X,R|\boldsymbol{W}=\boldsymbol{w})\approx COR(X,F|\boldsymbol{W}=\boldsymbol{w})$ était identifiable, et donc que $COR(X,R|\boldsymbol{W}=\boldsymbol{w})\approx OR(X,R|\boldsymbol{W}=\boldsymbol{w},S=1)$. $OR(X,R|\boldsymbol{W}=\boldsymbol{w},S=1)$ est alors approximativement valide d'un point de vue externe dans le cas (i). Dans le cas (ii) où l'inclusion des individus dépend de la gravité de l'accident, $COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w})$ n'est pas approximable par un autre odds ratio causal identifiable, et donc $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},S=1)$ n'a pas de validité externe. $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},S=1)$ est valide d'un point de vue externe dans le cas *(ii)* uniquement s'il n'existe pas de flèche entre X et V.

La validité externe de l'OR estimé dans les DAG E, F et celui correspondant au cas (ii) de l'analyse en responsabilité n'étant pas vérifiée, se pose alors la question de leur validité interne. Selon Rothman et al. (2008), « la validité interne implique la validité de l'inférence sur la population source de l'étude ». Plus formellement, nous dirons que $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ est valide d'un point de vue interne s'il est égal à l'odd ratio causal $\boldsymbol{w}$-spécifique dans la population source, ie $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)=COR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$.
L'égalité $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)=COR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ est donc vérifiée si $\mathbb{P}(Y_x=1|\boldsymbol{W}=\boldsymbol{w},S=1)=\mathbb{P}(Y=1|X=x,\boldsymbol{W}=\boldsymbol{w},S=1)$ pour $x\in(0,1)$, et donc si $Y_x \perp\!\!\!\perp X|(\boldsymbol{W},S)$. Dans le DAG E, $Y_x \perp\!\!\!\perp X|(\boldsymbol{W},S)$ n'est pas vérifiée. En effet, nous avons $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$ car $\boldsymbol{W}$ satisfait le critère back-door. En revanche, $(\boldsymbol{W},S)$ ne satisfait pas le critère car S est un collider et un descendant de X, donc $Y_x \not\perp\!\!\!\perp X|(\boldsymbol{W},S)$. Par conséquent, il n'est pas garanti que $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ n'ait de validité interne dans le DAG E.
Dans le DAG F, $Y_x \perp\!\!\!\perp X|(\boldsymbol{W},S)$ n'est pas non plus vérifiée, car S est un descendant de X. Il n'est donc pas garanti que $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ n'ait de validité interne ou externe dans le DAG F.
Un exemple de validité interne peut être fourni à travers le DAG C de la figure 2.2. Dans ce DAG, $Y_x \perp\!\!\!\perp X|(\boldsymbol{W},S)$ est vérifiée, et donc $\mathbb{P}(Y_x=1|\boldsymbol{W}=\boldsymbol{w},S=1)=\mathbb{P}(Y=1|X=x,\boldsymbol{W}=\boldsymbol{w},S=1)$. $OR(X,Y|\boldsymbol{W}=\boldsymbol{w},S=1)$ est alors valide d'un point de vue interne.
Considérons maintenant le cas (ii) de l'analyse en responsabilité où l'inclusion des individus dépend de la gravité de l'accident. Dans ce cas, notre échantillon est composé des individus pour lesquels $\{\mathcal{S}=1\}$, alors que la population source et la population cible sont constituées respectivement des individus pour lesquels $\{\mathcal{A}=1\}$ et $\{\mathcal{A}=1\}\cup\{\mathcal{A}=0\}$. Afin d'étudier la validité interne de $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)$, nous considérons pour simplifier que $\mathcal{S}=\mathcal{A}$, c'est-à-dire que notre échantillon correspond à l'ensemble des accidents graves. La question est donc

de savoir si :

$$OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) = \frac{\mathbb{P}(\mathcal{R} = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)/\mathbb{P}(\mathcal{R} = 0|X = 1, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)}{\mathbb{P}(\mathcal{R} = 1|X = 0, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)/\mathbb{P}(\mathcal{R} = 0|X = 0, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)}.$$

est égal à

$$COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) = \frac{\mathbb{P}(\mathcal{R}_1 = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)/\mathbb{P}(\mathcal{R}_1 = 0|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)}{\mathbb{P}(\mathcal{R}_0 = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)/\mathbb{P}(\mathcal{R}_0 = 0|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)}.$$

Cette égalité est vérifiée si $\mathbb{P}(\mathcal{R} = 1|X = x, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) = \mathbb{P}(\mathcal{R}_x = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$, c'est-à-dire si $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A})$ ou encore si $(\boldsymbol{W}, \mathcal{A})$ vérifie le critère back-door. Or, bien que $\boldsymbol{W}$ satisfasse le critère back-door et que $\mathcal{R}_x \perp\!\!\!\perp X|\boldsymbol{W}$, $(\boldsymbol{W}, \mathcal{A})$ ne satisfait pas le critère back-door, car $\mathcal{A}$ est un descendant de X. La condition $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A})$ n'est donc généralement pas vérifiée (cf. annexe .1 pour une preuve par l'absurde). Par conséquent, $\mathbb{P}(\mathcal{R} = 1|X = x, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ diffère généralement de $\mathbb{P}(\mathcal{R}_x = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$, et $OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) \neq COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ généralement. Il n'est donc pas garanti que $OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ n'ait de validité interne. En d'autres termes, même dans le cas le plus favorable où $\mathcal{S} = \mathcal{A}$, l'odd ratio ajusté observable dans les analyses en responsabilité où l'inclusion dans l'étude dépend de la gravité de l'accident $OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{S} = 1) = OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ ne peut pas être interprété comme l'odd ratio causal ajusté dans la population source $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$. Comme mentionné par Bareinboim and Pearl (2012), $OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ ne représente pas un odds ratio causal dans un échantillon stable d'individus, car $\mathcal{A}$ est un descendant de X. Plus précisément, $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A})$ n'est pas vérifiée, en revanche nous avons $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A}_x)$. En effet, cette indépendance peut rapidement être prouvée par l'approche SWIG (Richardson and Robins, 2013), et le SWIT correspondant au DAG G_s de la figure 2.4 (cf. figure 2.5). D'après la figure 2.5, $\mathcal{R}_x \perp\!\!\!\perp X|\boldsymbol{W}$ et $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A}_x)$. En effet, $\boldsymbol{W}$ bloque tous les chemins entre X et R_x, i.e. $\mathcal{R}_x \perp\!\!\!\perp X|\boldsymbol{W}$. De plus, $\mathcal{A}_x$ n'est pas un collider sur les chemins de X à R_x, et donc l'ensemble $(\boldsymbol{W}, \mathcal{A}_x)$ bloque également tous les chemins entre X et R_x, i.e. $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A}_x)$. Par conséquent,

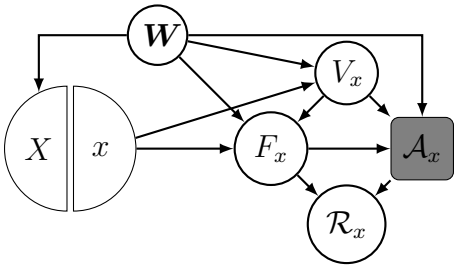
$$\begin{aligned}\mathbb{P}(\mathcal{R} = 1|X = x, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) &= \mathbb{P}(\mathcal{R}_x = 1|X = x, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_x = 1) \quad \text{par coherence} \\ &= \mathbb{P}(\mathcal{R}_x = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_x = 1) \quad \text{puisque } \mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A}_x).\end{aligned} \tag{2.4}$$

L'odd ratio ajusté et conditionné sur $\mathcal{A} = 1$ observé dans les analyses en responsabilité est alors formellement égal à :

$$OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) = \frac{\mathbb{P}(\mathcal{R}_1 = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_1 = 1)/\mathbb{P}(\mathcal{R}_1 = 0|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_1 = 1)}{\mathbb{P}(\mathcal{R}_0 = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_0 = 1)/\mathbb{P}(\mathcal{R}_0 = 0|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_0 = 1)}.$$

Les quantités $\mathbb{P}(R_x = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_x = 1)$ peuvent s'interpréter comme le risque d'être responsable dans le monde contrefactuel où l'alcool a été fixée à x pour les individus où $\boldsymbol{W} = \boldsymbol{w}$ et ayant un accident dans ce monde contrefactuel. Or, les sous-populations pour lesquelles $\{\mathcal{A}_0 = 1, \boldsymbol{W} = \boldsymbol{w}\}$ et $\{\mathcal{A}_1 = 1, \boldsymbol{W} = \boldsymbol{w}\}$ sont très différentes. La population $\{\mathcal{A}_0 = 1, \boldsymbol{W} = \boldsymbol{w}\}$

FIGURE 2.5 – SWIT résultant de l'intervention $do(X = x)$ dans le cas *(ii)*



est composée des individus qui auraient eu un accident dans le monde contrefactuel où ils n'auraient pas bu, tandis que la population $\{\mathcal{A}_1 = 1, \boldsymbol{W} = \boldsymbol{w}\}$ est composée des conducteurs qui auraient eu un accident dans le monde contrefactuel où ils auraient bu. On peut supposer que la population $\{\mathcal{A}_0 = 1, \boldsymbol{W} = \boldsymbol{w}\}$ est composée soit de conducteurs plus âgés (le vieillissement ralentissant le traitement de l'information), soit de conducteurs plus distraits que la population $\{\mathcal{A}_1 = 1, \boldsymbol{W} = \boldsymbol{w}\}$. Par conséquent, les probabilités intervenant dans le calcul de l'$OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ n'étant pas estimée sur une population identique, $OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ doit être interprétée avec prudence.
Dans le cas où X n'est pas une cause de V, alors $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A})$ est vérifiée. $\boldsymbol{W}$ bloque tous les chemins back-door, et $\mathcal{A}$ n'étant cette fois-ci pas un descendant de X, l'ensemble $(\boldsymbol{W}, \mathcal{A})$ vérifie également le critère back-door. $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A})$, et dans ce cas $OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ est valide d'un point de vue interne.

Notons que cette problématique n'est pas restreinte à l'analyse en responsabilité. Il est connu que sélectionner les individus à partir de l'événement d'intérêt, ou à partir d'une variable impactée par l'événement d'intérêt, peut conduire à un biais de sélection. L'exemple le plus célèbre correspond au biais de Berkson (Berkson, 1946). Par ailleurs, conditionner les estimations sur un descendant de l'exposition dans le DAG est connu comme un phénomène plus général appelé le biais de collision (« collider bias ») (Greenland, 2003, Rothman et al., 2008). C'est ce type de biais qui entre en compte dans l'explication du paradoxe de l'obésité (Lajous et al., 2014, Viallon and Dufournet, 2017a), comme nous le verrons au chapitre 4.

Conclusion

Communément adopté dans le champ de la sécurité routière, les analyses en responsabilité proposent d'estimer des quantités causales en présence d'un processus de sélection, c'est-à-dire en considérant uniquement les accidentés de la route. Dans ce chapitre, nous avons donc mis l'accent sur les conditions d'identifiabilité de quantités causales en présence de processus sélection afin de fournir une description formelle des analyses en responsabilité.
Nous avons considéré tout au long de ce chapitre l'identifiabilité d'un effet de X l'alcool sur la responsabilité de l'accident R puis sur la responsabilité de l'accident grave $\mathcal{R}$ dans deux cas. Le premier, lorsque l'inclusion des données ne dépend pas de la gravité. Le second, qui est le plus fréquent en pratique, lorsque les données concernent uniquement les accidents graves (corporels ou

mortels par exemple). Nous avons montré que l'odd ratio causal conditionnel $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ n'est pas identifiable dans les deux situations. Dans le cas où des données concernant l'ensemble des accidents, indépendamment de leur gravité, seraient disponibles, et sous certaines hypothèses, l'odd ratio causal conditionnel $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ est approximable par un autre odds ratio causal identifiable $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$. Néanmoins, la plupart du temps, seuls les accidents graves constituent les données disponibles et dans ce cas $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ n'est plus identifiable, et donc $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ n'est plus approximable. Ces derniers résultats sont valables pour n'importe quelle exposition X ayant un effet sur la vitesse V. Lorsque X n'est pas une cause de V, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ reste identifiable et donc $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ est approximable.
Dans le cas où seuls des accidents graves sont disponibles, les non-responsables accidentés ne semblent pas constituer un groupe de témoins représentatifs des non-responsables ($\mathcal{R} = 0$). Cette population témoin est en fait composée de trois types de conducteurs : ceux qui n'ont pas commis de faute et qui n'ont pas d'accident grave ($F = 0, \mathcal{A} = 0$), ceux qui n'ont pas commis de faute et qui ont un accident grave ($F = 0, \mathcal{A} = 1$), et ceux qui ont commis une faute et qui n'ont pas d'accident grave ($F = 1, \mathcal{A} = 0$). Or, dans les analyses en responsabilité, puisque seuls les accidents sont observables, le groupe témoin est uniquement composé des conducteurs qui n'ont pas commis de faute et qui ont un accident grave ($F = 0, \mathcal{A} = 1$). Du fait du lien causal supposé entre l'alcoolémie et vitesse élevée, les conducteurs non responsables étant impliqués dans un accident grave roulent certainement plus vite que ceux non impliqués dans un accident. Ainsi, les sous-populations $\{(F = 0, \mathcal{A} = 0), (F = 1, \mathcal{A} = 0)\}$ et $(F = 0, \mathcal{A} = 1)$ sont différentes au regard de l'exposition X. La mesure d'association estimée dans les analyses en responsabilité est donc probablement une sous-estimation de la mesure causale. C'est ce que nous chercherons à vérifier et préciser dans le chapitre suivant. D'autres chercheurs ont soulevé cette question de la non-représentativité des témoins non-responsables accidentés et ont proposé de transformer le groupe témoin afin de le rendre plus représentatif (Laumon et al., 2005). Toutefois, il n'y a à ce jour aucun moyen de déterminer si le nouveau groupe de témoins constitué est effectivement représentatif sans information additionnelle. La connaissance de la prévalence de l'exposition $\mathbb{P}(X)$, par exemple pourrait rendre l'odd ratio causal conditionnel $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$, $COR(X, \mathcal{R})$ et $\mathbb{P}(\mathcal{R}_x = r)$, pour $(x, r) \in \{0, 1\}$ identifiables à partir des données disponibles, même si la structure du DAG ne permet pas l'identifiabilité de ces quantités (cf. section 3 de Bareinboim and Pearl (2012))
Par conséquent, lorsque seuls les accidents graves sont disponibles, la seule quantité qu'il est possible d'estimer reste la mesure d'association $OR(X, R|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$. Or, cette mesure reste à interpréter avec prudence car ni sa validité externe ni sa validité interne n'est garantie. $OR(X, R|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ est en fait égal à

$$\frac{\mathbb{P}(\mathcal{R}_1 = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_1 = 1)/\mathbb{P}(\mathcal{R}_1 = 0|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_1 = 1)}{\mathbb{P}(\mathcal{R}_0 = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_0 = 1)/\mathbb{P}(\mathcal{R}_0 = 0|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_0 = 1)}$$

et non à

$$\frac{\mathbb{P}(\mathcal{R} = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}, A = 1)/\mathbb{P}(\mathcal{R} = 0|X = 1, \boldsymbol{W} = \boldsymbol{w}, A = 1)}{\mathbb{P}(\mathcal{R} = 1|X = 0, \boldsymbol{W} = \boldsymbol{w}, A = 1)/\mathbb{P}(\mathcal{R} = 0|X = 0, \boldsymbol{W} = \boldsymbol{w}, A = 1)}.$$

Les populations $\{\mathcal{A}_1 = 1, \boldsymbol{W} = \boldsymbol{w}\}$ et $\{\mathcal{A}_0 = 1, \boldsymbol{W} = \boldsymbol{w}\}$ pouvant être sensiblement différentes, l'odd ratio estimé n'a donc pas de valeur causale.
Les récents résultats de la littérature concernant l'identifiabilité de quantités causales en présence de sélection nous ont permis de formaliser l'analyse en responsabilité d'un point de vue théorique.

Il s'agit dans le chapitre 3 d'illustrer ces résultats théoriques par simulations numériques et de donner un premier ordre de grandeur du biais induit par les analyses en responsabilité.

Chapitre 3

Sens et quantification du biais dans les analyses en responsabilité

Ce chapitre fera l'objet d'un article que nous comptons soumettre dans Analytic Methods in Accident Research.

Dans le chapitre précédent, l'objectif était en particulier d'étudier l'identifiabilité des odds ratios causaux :

- $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$, l'odds ratio causal de X sur la responsabilité de l'accident R dans la strate de la population où le vecteur de facteurs de confusion $\boldsymbol{W} = \boldsymbol{w}$,
- $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$, l'odds ratio causal de X sur la responsabilité de l'accident grave $\mathcal{R}$ dans la strate de la population où le vecteur de facteurs de confusion $\boldsymbol{W} = \boldsymbol{w}$

Sous certaines hypothèses de prévalences faibles, $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ est approximable par $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$, l'odds ratio causal de X sur la faute de conduite, dans la strate où $\boldsymbol{W} = \boldsymbol{w}$. $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ est en effet identifiable car, d'après le DAG, $X \perp\!\!\!\perp A|(F, \boldsymbol{W})$, où A désigne la survenue de l'accident. Lorsque l'inclusion des données dépend de la sévérité de l'accident, ni $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$ ni $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ n'est identifiable, car $X \perp\!\!\!\perp \mathcal{A}|(F, \boldsymbol{W})$, où $\mathcal{A}$ désigne la survenue d'un accident grave, n'est pas vérifiée. Cette dépendance conditionnelle est liée à la présence de deux flèches dans le DAG : celle de X vers la vitesse élevée V, et celle de V vers $\mathcal{A}$. Dès lors que $X \not\perp\!\!\!\perp \mathcal{A}|(F, \boldsymbol{W})$, l'odds ratio estimable, $OR(X, F|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) = OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$ diffère de $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$. Par conséquent, il s'agit dans ce chapitre de s'intéresser à l'ampleur du biais entre la mesure d'association estimable dans les analyses en responsabilité et les véritables odds ratios causaux en fonction de l'écart à l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F, \boldsymbol{W})$. Nous vérifierons également que ce biais augmente lorsque l'on s'éloigne de cette indépendance.
Pour cela, nous spécifions les fonctions génératrices et perturbations associées à chaque variable du DAG 2.4 afin de pouvoir calculer les trois quantités théoriques d'intérêt $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$, et $OR(X, F|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) = OR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$, et les comparer. En considérant X binaire, et en faisant varier positivement le paramètre correspondant à l'effet de X sur la vitesse V, nous décrivons une situation où X pourrait représenter le fait d'avoir consommé de l'alcool. A l'inverse, en faisant varier négativement le paramètre correspondant à l'effet de X sur la vitesse V, nous décrivons une situation où X pourrait représenter le fait d'avoir consommé du cannabis. Enfin, en considérant X en catégorielle, l'idée est de se rapprocher d'une situation où X représenterait les différents taux d'alcoolémie. Dans ce dernier cas, nous

considérons un effet non linéaire de X sur la vitesse puisque l'augmentation de la vitesse sous l'emprise de l'alcool est beaucoup plus forte pour un faible taux d'alcoolémie (0,5-0,8 gr/l) que pour les doses plus élevées (Kelly et al., 2004, Ronen et al., 2010).
Nous considérons l'exposition X binaire puis catégorielle. Quelle que soit la configuration, nous décrivons un modèle causal simple. Ce modèle est décrit dans le cas le plus général, le cas *(ii)*, où l'inclusion dépend de la gravité de l'accident $\mathcal{A}$ et est considérée la responsabilité de l'accident grave $\mathcal{R}$. L'illustration du cas *(i)*, où l'inclusion ne dépend pas de la gravité, s'obtient en fixant certains paramètres à zéro.

3.1 Illustration numérique lorsque X est binaire

Dans cette première section, nous considérons une exposition X binaire, et considérons un cas très simple où il existe un seul facteur de confusion binaire W, afin d'illustrer les biais entre $COR(X, F|W = w)$, $COR(X, \mathcal{R}|W = w)$ et $OR(X, F|W = w, \mathcal{A} = 1)$. Nous présentons tout d'abord le modèle causal considéré, inspiré de l'étude numérique menée par Sperrin et al. (2016) dans le cadre de la quantification des biais en jeu dans le paradoxe de l'obésité (cf. chapitre 4). Nous exposons ensuite les résultats de l'étude en présence de facteurs de confusion puis en présence de facteurs de confusion et de certaines interactions.

3.1.1 Modèle causal considéré

L'objectif de cette partie est de comparer les odds ratios causaux $COR(X, F|W = w), COR(X, \mathcal{R}|W = w)$ à l'association mesurable dans les analyses en responsabilité $OR(X, F|W = w, \mathcal{A} = 1)$. Rappelons que ces mesures sont définies de la façon suivante :

$$
\begin{aligned}
COR(X, F|W = w) &= \frac{\mathbb{P}(F_1 = 1|W = w)/\mathbb{P}(F_1 = 0|W = w)}{\mathbb{P}(F_0 = 1|W = w)/\mathbb{P}(F_0 = 0|W = w)},\\
COR(X, \mathcal{R}|W = w) &= \frac{\mathbb{P}(\mathcal{R}_1 = 1|W = w)/\mathbb{P}(\mathcal{R}_1 = 0|W = w)}{\mathbb{P}(\mathcal{R}_0 = 1|W = w)/\mathbb{P}(\mathcal{R}_0 = 0|W = w)},\\
OR(X, F|W = w, \mathcal{A} = 1) &= \frac{\mathbb{P}(F = 1|X = 1, W = w, \mathcal{A} = 1)/\mathbb{P}(F = 0|X = 1, W = w, \mathcal{A} = 1)}{\mathbb{P}(F = 1|X = 0, W = w, \mathcal{A} = 1)/\mathbb{P}(F = 0|X = 0, W = w, \mathcal{A} = 1)},
\end{aligned}
$$

où $OR(X, F|W = w, \mathcal{A} = 1) = OR(X, R|W = w, \mathcal{A} = 1)$ puisque $F = R$ dès lors que $\mathcal{A} = 1$.

D'après le DAG de droite de la figure 2.4, il est possible de préciser les quantités intervenant dans le calcul de ces trois odds ratios. Notons $p_V(x, w) = \mathbb{P}(V = 1|X = x, W = w)$, $p_F(x, v, w) = \mathbb{P}(F = 1|X = x, V = v, W = w)$ et $p_A(f, v, w) = \mathbb{P}(\mathcal{A} = 1|F = f, V = v, W = w)$ pour $(x, v, w, f) \in \{0, 1\}^4$. Nous obtenons :

$$
\begin{aligned}
&\mathbb{P}(F_x = 1|W = w) = \mathbb{P}(F_x = 1|X = x, W = w) \quad \text{car } F_x \perp\!\!\!\perp X|W\\
&= \mathbb{P}(F = 1|X = x, W = w) \quad \text{par cohérence}\\
&= \sum_{v\in(0,1)} \{\mathbb{P}(F = 1|X = x, V = v, W = w)\mathbb{P}(V = v|X = x, W = w)\}\\
&= p_F(x, 1, w)p_V(x, w) + p_F(x, 0, w)(1 - p_V(x, w))
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}(\mathcal{R}_x = 1|W = w) = \mathbb{P}(\mathcal{A}_x = 1, F_x = 1|W = w) \\
&= \mathbb{P}(\mathcal{A}_x = 1, F_x = 1|X = x, W = w) \quad \text{car } (A_x, F_x) \perp\!\!\!\perp X|W \\
&= \mathbb{P}(\mathcal{A} = 1, F = 1|X = x, W = w) \quad \text{par cohérence} \\
&= \mathbb{P}(\mathcal{A} = 1|F = 1, X = x, W = w)\mathbb{P}(F = 1|X = x, W = w) \\
&= \sum_{v\in(0,1)} \{\mathbb{P}(\mathcal{A} = 1|F = 1, V = v, W = w)\mathbb{P}(F = 1|X = x, V = v, W = w)\mathbb{P}(V = v|X = x, W = w)\} \\
&= p_A(1, 1, w)p_F(x, 1, w)p_V(x, w) + p_A(1, 0, w)p_F(x, 0, w)(1 - p_V(x, w))
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}(F = 1|X = x, W = w, \mathcal{A} = 1) \\
&= \frac{\mathbb{P}(F = 1, X = x, W = w, \mathcal{A} = 1)}{\mathbb{P}(X = x, W = w, \mathcal{A} = 1)} \\
&= \frac{\mathbb{P}(\mathcal{A} = 1|F = 1, X = x, W = w)\mathbb{P}(F = 1, X = x, W = w)}{\mathbb{P}(\mathcal{A} = 1|X = x, W = w)\mathbb{P}(X = x|W = w)\mathbb{P}(W = w)} \\
&= \frac{\sum_{v\in(0,1)} \{\mathbb{P}(\mathcal{A} = 1|F = 1, V = v, W = w)\mathbb{P}(F = 1|X = x, V = v, W = w)\mathbb{P}(V = v|X = x, W = w)\}}{\sum_{v,f\in(0,1)} \{\mathbb{P}(\mathcal{A} = 1|F = f, V = v, W = w)\mathbb{P}(F = f|X = x, V = v, W = w)\mathbb{P}(V = v|X = x, W = w)\}} \\
&= [p_A(1, 1, w)p_F(x, 1, w)p_V(x, w) + p_A(1, 0, w)p_F(x, 0, w)(1 - p_V(x, w))] \\
&\quad \times [p_A(1, 1, w)p_F(x, 1, w)p_V(x, w) + p_A(1, 0, w)p_F(x, 0, w)(1 - p_V(x, w)) \\
&\quad + p_A(0, 1, w)(1 - p_F(x, 1, w))p_V(x, w) + p_A(0, 0, w)(1 - p_F(x, 0, w))(1 - p_V(x, w))]^{-1}
\end{aligned}
$$

Ensuite, en nous inspirant de l'étude numérique menée par Sperrin et al. (2016), nous spécifions les distributions des variables du DAG. Seules X et W sont des variables ne dépendant pas d'autres variables du DAG. Nous définissons donc deux variables aléatoires indépendantes suivant une loi de Bernoulli : $X \sim B(p_X)$ et $W \sim B(p_W)$. Nous fixons $p_X = p_W = 0.5$. Nous introduisons ensuite la fonction sigmoïde $\mathrm{h}(x) = (1+\exp(-x))^{-1}$, et posons, pour tout $(x, v, w, f) \in \{0, 1\}^4$ et pour les paramètres suivants $\alpha_0, \alpha_X, \alpha_W, \alpha_{XW}, \beta_0, \beta_X, \beta_V, \beta_W, \beta_{XV}, \beta_{XW}, \beta_{VW}, \gamma_0, \gamma_F, \gamma_V, \gamma_W$, $\gamma_{FV}, \gamma_{FW}, \gamma_{VW}$:

$$
\begin{aligned}
p_V(x, w) &= \mathrm{h}(\alpha_0 + \alpha_X x + \alpha_W w + \alpha_{XW} xw) \\
p_F(x, v, w) &= \mathrm{h}(\beta_0 + \beta_X x + \beta_V v + \beta_W w + \beta_{XV} xv + \beta_{XW} xw + \beta_{VW} vw) \\
p_A(f, v, w) &= \mathrm{h}(\gamma_0 + \gamma_F f + \gamma_V v + \gamma_W w + \gamma_{FV} fv + \gamma_{FW} fw + \gamma_{VW} vw).
\end{aligned}
$$

Par ailleurs, pour $\nu = 13$, nous posons les constantes suivantes de telle sorte que la probabilité de rouler à une vitesse élevée $\mathbb{P}(V = 1)$ soit égale à 0.5 et que les prévalences de F et A soient

inférieures à $10^{-6}\%$, c'est-à-dire des prévalences réalistes (Bouaoun et al., 2015).

$$\begin{aligned}
\alpha_0 &= -\frac{1}{2}\left(\alpha_X + \alpha_W + \frac{1}{2}\alpha_{XW}\right) \\
\beta_0 &= -\frac{1}{2}\left(\beta_X + \beta_V + \beta_W + \frac{1}{2}(\beta_{XV} + \beta_{XW} + \beta_{VW}) - \nu\right) \\
\gamma_0 &= -\frac{1}{2}\left(\gamma_F + \gamma_V + \gamma_W \frac{1}{2}(\beta_{FV} + \beta_{FW} + \beta_{VW}) - \nu\right)
\end{aligned}$$

Sous cette configuration, il est donc possible d'obtenir les formules analytiques de $\mathbb{P}(F_x = 1|W = w)$, $\mathbb{P}(\mathcal{R}_x = 1|W = w)$ et $\mathbb{P}(F = 1|X = x, W = w, \mathcal{A} = 1)$ et donc celle de $COR(X, F|W = w)$, $COR(X, \mathcal{R}|W = w)$ et $OR(X, F|W = w, \mathcal{A} = 1)$. Dans la mesure où $\mathrm{h}(-x) = 1 - \mathrm{h}(x)$, nous obtenons par exemple sous le modèle considéré ci-dessus :

$$\begin{aligned}
&\mathbb{P}(F_x = 1|W = w) = p_F(x, 1, w)p_V(x, w) + p_F(x, 0, w)(1 - p_V(x, w)) \\
&= \mathrm{h}(\beta_0 + \beta_X x + \beta_V + \beta_W w + \beta_{XV} x + \beta_{XW} xw + \beta_{VW} w)\mathrm{h}(\alpha_0 + \alpha_X x + \alpha_W w + \alpha_{XW} xw) + \\
&\quad \mathrm{h}(\beta_0 + \beta_X x + \beta_W w + \beta_{XW} xw)\mathrm{h}(-(\alpha_0 + \alpha_X x + \alpha_W w + \alpha_{XW} xw))
\end{aligned}$$

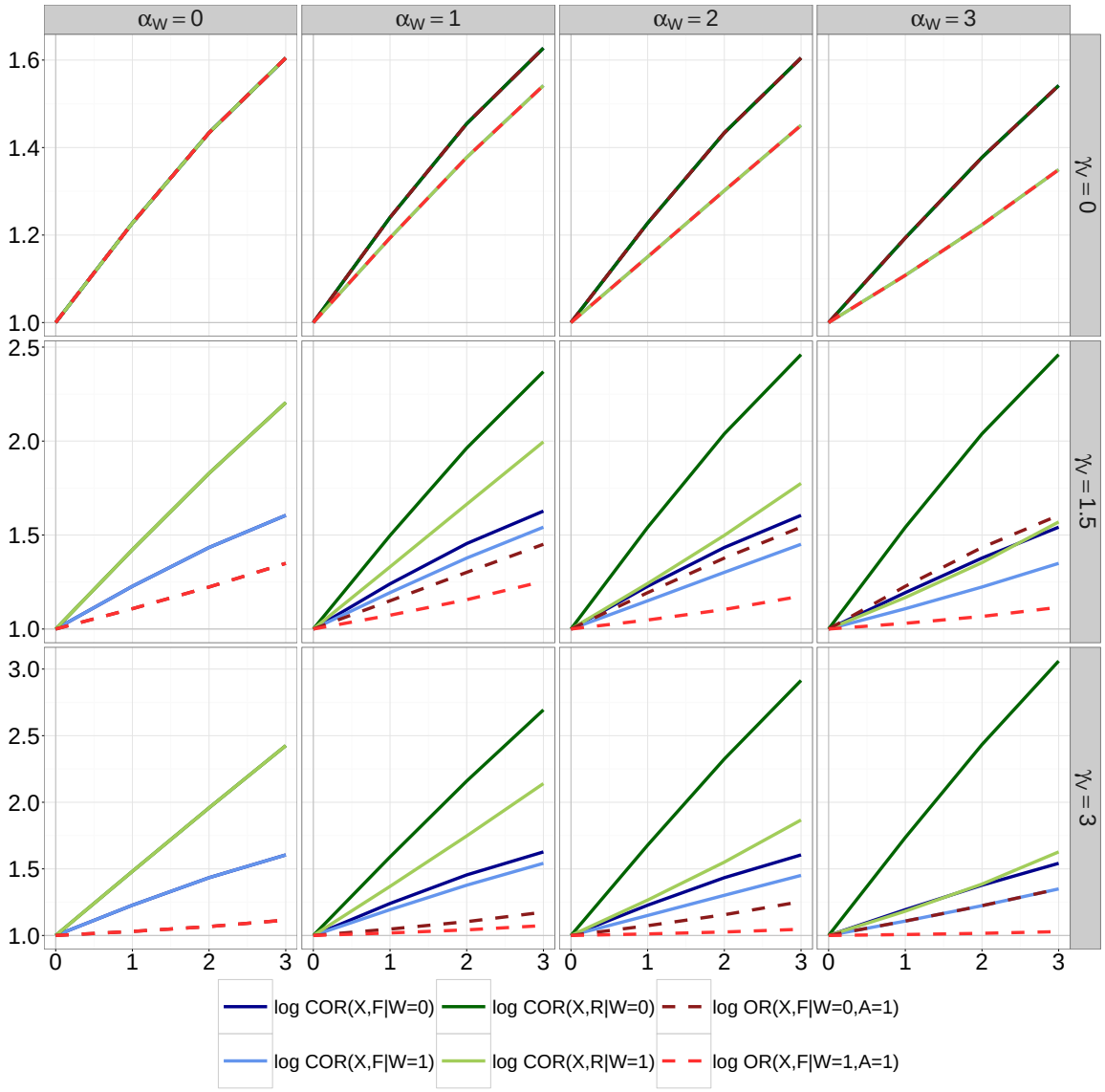
Etant donné que l'identifiabilité de $COR(X, F|W = w)$ et l'approximation de $COR(X, \mathcal{R}|W = w)$ repose en particulier sur la condition $X \perp\!\!\!\perp \mathcal{A}|(F, W)$, les résultats du chapitre 2 suggèrent que l'importance du biais entre les trois odds ratios d'intérêt dépend de l'écart à l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F, W)$. Le fait que cette indépendance ne soit pas vérifiée, est liée à la présence d'une flèche entre X et V et d'une flèche entre V et $\mathcal{A}$, et donc en premier lieu à la valeur des paramètres α_X et γ_V. Nous faisons également varier les autres paramètres du modèle afin de déterminer dans quel sens et quelle ampleur sont modifiés les biais entre les trois quantités.

3.1.2 Résultats avec facteurs de confusion

Puisque l'ampleur des biais réside en premier lieu dans la valeur des paramètres α_X et γ_V, ce sont ces deux paramètres que nous faisons toujours varier dans nos illustrations.
Concernant γ_V, nous considérons trois situations : $\gamma_V = 0$, qui correspond à l'absence de flèche entre V et $\mathcal{A}$, $\gamma_V = 1.5$ et $\gamma_V = 3$. Concernant α_X, d'une part nous faisons varier α_X entre 0 et 3, ce qui permet de décrire une situation où X représenterait le fait d'avoir consommé de l'alcool, et augmenterait la vitesse. D'autre part, nous faisons varier α_X entre -3 et 0, permettant de nous rapprocher d'une situation où X représenterait le fait d'avoir consommé du cannabis, et à l'inverse diminuerait la vitesse. Considérer α_X variant entre 0 et 3, respectivement -3 et 0, multiplierait, ou respectivement diviserait, entre 0 et 20 le risque de rouler à une vitesse élevée. Nous présentons nos résultats en posant $\beta_V = 1$ et $\gamma_F = 4$, puisque V augmente le risque de commettre une faute de conduite et F augmente fortement le risque d'avoir un accident.

Afin d'illustrer les résultats du chapitre 2, considérons pour commencer le cas où il n'y a pas d'effet du facteur de confusion, et qui correspond à la première colonne des figures 3.1 et 3.2 où $\alpha_W = 0$ (la variation de β_W et γ_W ne modifiant que très marginalement les estimations. Rappelons qu'ils sont fixés ici à 1).

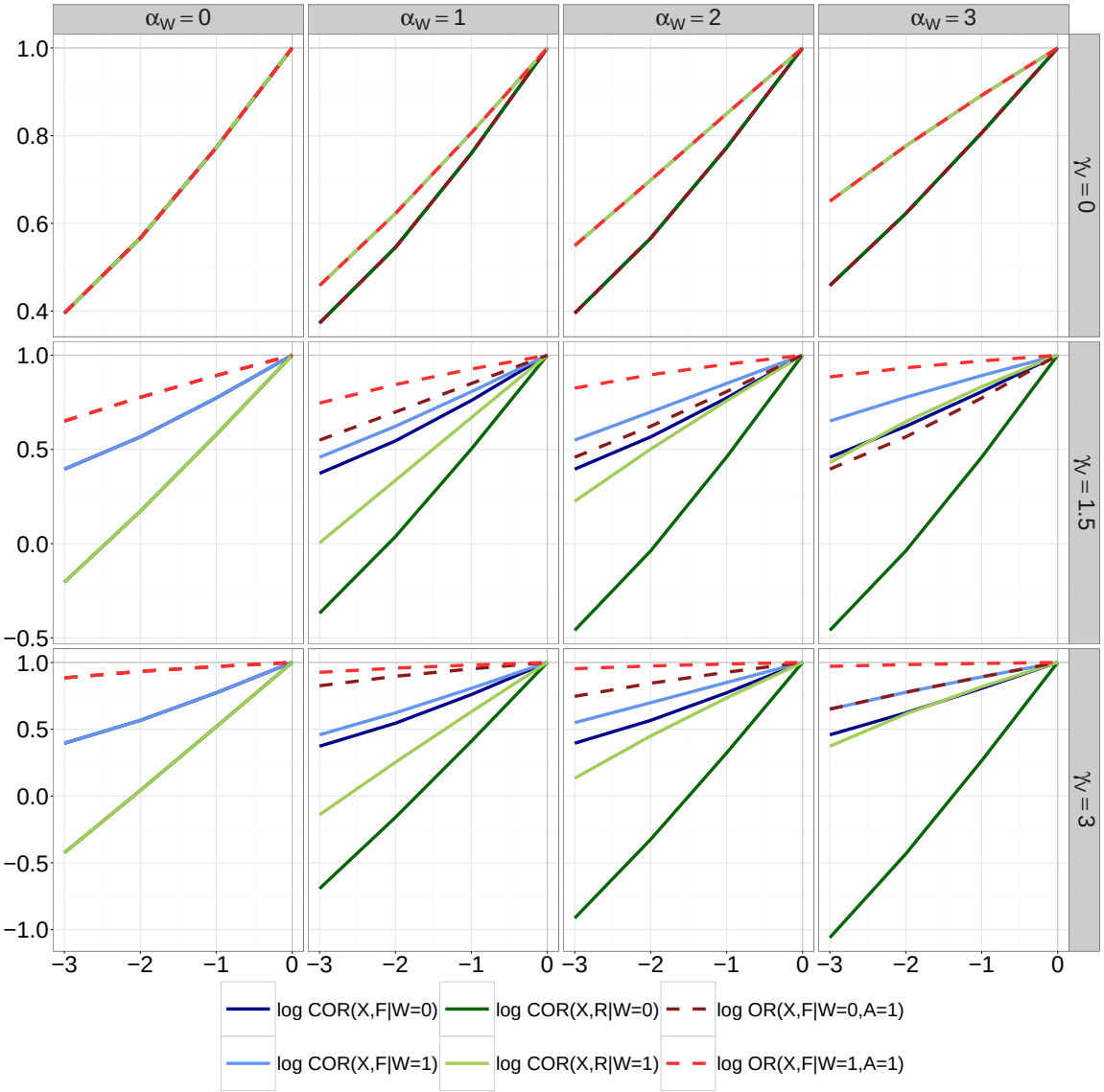
FIGURE 3.1 – Odds ratios causaux et odds ratio estimable sur l'échelle logarithmique dans le cas où $\beta_V = \beta_X = \beta_W = \gamma_W = 1, \gamma_F = 4$, sans interaction, et pour des valeurs variables des autres paramètres $\alpha_X, \alpha_W, \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_X varie de 0 à 3



Lecture (3ème cadran de la premiere colonne) : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et lorsque le facteur de confusion n'a pas d'effet sur la vitesse ($\alpha_W = 0$). Pour $\alpha_X = 1$, le logarithme de l'odds ratio estimable (en pointillé rouge) est proche de 1, le logarithme de l'odds ratio causal sur la faute (en trait plein bleu) est environ égal à 1,2 et le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave (en trait plein vert) est à 1,5. Les trois quantités n'étant pas égales, il existe un biais.

Lorsque l'inclusion des données ne dépend pas de la sévérité de l'accident, i.e. $\gamma_V = 0$, nos résultats confirment l'absence de biais entre $COR(X, F|W = w)$ et $OR(X, F|W = w, \mathcal{A} = 1)$, quelle que soit la valeur de α_X. Ce résultat est attendu puisque lorsque $\gamma_V = 0$, $X \perp\!\!\!\perp \mathcal{A}|(F, W)$ est vérifiée. L'illustration numérique confirme également dans ce cas que $COR(X, \mathcal{R}|W = w) \simeq$

FIGURE 3.2 – Odds ratios causaux et odds ratio estimable sur l'échelle logarithmique dans le cas où $\beta_V = \beta_X = \beta_W = \gamma_W = 1, \gamma_F = 4$, sans interaction, et pour des valeurs variables des autres paramètres $\alpha_X, \alpha_W, \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_X varie de -3 à 0



Lecture (3ème cadran de la premiere colonne) : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et lorsque le facteur de confusion n'a pas d'effet sur la vitesse ($\alpha_W = 0$). Pour $\alpha_X = -3$, le logarithme de l'odds ratio estimable (en pointillé rouge) est proche de 0,9, le logarithme de l'odds ratio causal sur la faute (en trait plein bleu) est environ égal à 0,4 et le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave (en trait plein vert) est presque égal à -0,5. Les trois quantités n'étant pas égales, il existe un biais.

$COR(X, F|W = w)$. Rappelons que nous nous sommes placés dans un cas où les prévalences de $\mathcal{R}$ et F sont faibles pour tout w, x, v.

Lorsque l'inclusion des données dépend de la sévérité de l'accident, i.e. $\gamma_V \neq 0$, il existe un biais entre $COR(X, F|W = w)$ et $OR(X, F|W = w, \mathcal{A} = 1)$ dès lors que $\alpha_X \neq 0$, et l'amplitude de

ce biais augmente avec α_X. Comme attendu, les biais sont d'autant plus grands que γ_V et α_X sont grands, puisqu'on s'éloigne alors d'autant plus de l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F,W)$. Dans ce cas où $\gamma_V \neq 0$, $COR(X,\mathcal{R}|W=w)$ ne peut plus être approximé par $COR(X,F|W=w)$. Et l'écart entre ces deux odds ratios causaux est d'autant plus grand que γ_V est grand, puisque qu'on s'éloigne une fois de plus de $X \perp\!\!\!\perp \mathcal{A}|(F,W)$, hypothèse sur laquelle est en partie fondée l'approximation. Il existe un cas particulier où il n'existe pas de biais entre les trois quantités, c'est lorsque que $\alpha_X = 0$. Cela correspond à l'absence de flèche entre X et V et donc à l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F,W)$.
Observons maintenant le sens des biais. Dans le cas où X augmente la probabilité de rouler à une vitesse élevée V (α_X variant entre 0 et 3), l'OR estimable $OR(X,F|W=w,\mathcal{A}=1)$ dans les analyses en responsabilité sous-estime les deux odds ratios causaux $COR(X,F|W=w)$ et $COR(X,\mathcal{R}|W=w)$, comme attendu (cf. discussion du Chapitre 2). Dans le cas où X diminue la probabilité de rouler à une vitesse élevée V (α_X variant entre -3 et 0), il sur-estime les deux odds ratios causaux. Nous avons essentiellement abordé le biais quantitatif jusqu'ici, mais notons que dans le cas où α_X varie dans les négatifs, il peut y avoir également un biais qualitatif. Dans le cas où α_X est inférieur à -2, l'odds ratio total de X sur $\mathcal{R}$ ajusté sur W serait protecteur ($\log COR(X,\mathcal{R}|W=w) < 0$) alors que l'OR estimable à partir des analyses en responsabilité suggèrerait un effet délétère ($\log OR(X,F|W=w,\mathcal{A}=1) > 0$). En d'autres termes, il est possible, sous certaines configurations du modèle causal simple considéré ici, d'observer un OR supérieur à 1 et conclure à l'effet délétère du cannabis sur la responsabilité de l'accident grave, alors qu'il aurait en réalité un effet protecteur.

Considérons maintenant un facteur de confusion W, binaire, ayant un impact positif sur la vitesse élevée V, la faute F et l'accident grave $\mathcal{A}$. Les paramètres modélisés peuvent décrire par exemple une situation où W serait le fait d'être jeune conducteur. Les jeunes conducteurs roulant généralement plus vite, et ayant plus de risque de commettre une faute ou d'être impliqué dans un accident du fait de leur manque d'expérience. Notons que la variation des paramètres β_W et γ_W ne modifient que très marginalement les résultats, c'est pourquoi ils ne sont pas présentés. Par conséquent, nous faisons varier α_W entre 0 et 3 et fixons β_W et γ_W à 1.

D'après les figures 3.1 et 3.2, quels que soit γ_V et α_W, et pour tout $w \in (0,1)$, l'écart entre $COR(X,\mathcal{R}|W=w)$ et $COR(X,F|W=w)$ augmente avec α_X, comme attendu. Il en est de même pour le biais entre ces odds ratios causaux et $OR(X,F|W=w,\mathcal{A}=1)$. Pour un γ_V donné, le biais entre $COR(X,\mathcal{R}|W=0)$ et $COR(X,F|W=0)$ augmente avec α_W, alors qu'il diminue entre $COR(X,\mathcal{R}|W=1)$ et $COR(X,F|W=1)$. En revanche, le biais entre $COR(X,F|W=w)$ et $OR(X,F|W=w,\mathcal{A}=1)$ diminue avec l'augmentation de α_W, quelle que soit la strate W. Il semble qu'à une variation près, on s'éloigne de l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F,W)$ dans la strate où $\{W=0\}$, i.e. $\mathbb{P}(\mathcal{A}=1|X=1,F=f,W=0)/\mathbb{P}(\mathcal{A}=1|X=0,F=f,W=0)$ s'éloigne de 1, lorsque α_W augmente. Et à l'inverse, dans la strate $\{W=1\}$, on se rapproche de $X \perp\!\!\!\perp \mathcal{A}|(F,W)$, et donc le risque relatif $\mathbb{P}(\mathcal{A}=1|X=1,F=f,W=1)/\mathbb{P}(\mathcal{A}=1|X=0,F=f,W=1)$ se rapproche de 1. Ces hypothèses sont validées par le calcul analytique des risques relatifs dans chaque strate de W (cf. tableau 3.1). Le risque relatif s'obtient en précisant la quantité $\mathbb{P}(\mathcal{A}=1|X=x,F=f,W=1)$. En effet, $\mathbb{P}(\mathcal{A}=1|X=x,F=f,W=1) = \sum_{v\in(0,1)}\{\mathbb{P}(\mathcal{A}=1|X=x,V=v,W=w)\mathbb{P}(F=1|X=x,V=v,W=w)\mathbb{P}(V=v|X=x,W=w) = p_A(f,1,w)p_V(x,w) + p_A(f,0,w)(1-p_V(x,w))$. Quel que soit $f \in (0,1)$, les risques relatifs va-

rient très marginalement, par conséquent, dans les tableaux de ce chapitre, ils sont présentés pour $f = 1$.

TABLE 3.1 – Risques relatifs $\mathbb{P}(\mathcal{A} = 1|X = 1, F = f, W = w)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = f, W = w)$ pour $\alpha_X = 1$ et $\gamma_V = 3$ et différentes valeurs de α_W

	$\alpha_W = 0$	$\alpha_W = 1$	$\alpha_W = 2$	$\alpha_W = 3$
$\{W = 1\}$	1,57	1,42	1,29	1,19
$\{W = 0\}$	1,57	1,72	1,83	1,87

Lecture : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et lorsque le facteur de confusion a un faible effet sur la vitesse ($\alpha_W = 1$). Le risque relatif dans la strate où $\{W = 1\}$, $\mathbb{P}(\mathcal{A} = 1|X = 1, F = 1, W = 1)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = 1, W = 1)$ est alors égal à 1,42. Le risque relatif dans la strate où $\{W = 0\}$, $\mathbb{P}(\mathcal{A} = 1|X = 1, F = 1, W = 0)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = 1, W = 0)$ est alors égal à 1,72.

Par ailleurs, pour $\gamma_V \neq 0$, l'OR estimable dans les analyses en responsabilité semble également mieux estimer $COR(X, F|W = w)$ que $COR(X, \mathcal{R}|W = w)$. Prenons tout d'abord l'exemple où X représenterait l'alcool, et à un effet positif sur le fait de rouler à une vitesse élevée (α_X variant dans les positifs). D'une part, on peut montrer que lorsque $\gamma_V \neq 0$, $COR(X, F|W = w) \leq COR(X, \mathcal{R}|W = w)$. En effet, en se référant à la preuve que $COR(X, \mathcal{R}|W = w) \simeq COR(X, F|W = w)$ du chapitre 2, on établit que dans le cas où $X \perp\!\!\!\perp A|(F, W)$:

$$
\begin{aligned}
COR(X, R|\boldsymbol{W} = \boldsymbol{w}) &\approx CRR(X, R|\boldsymbol{W} = \boldsymbol{w}) \\
&= \frac{\mathbb{P}(A = 1|F = 1, X = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(A = 1|F = 1, X = 0, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(A = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(A = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \text{ puisque X } \perp\!\!\!\perp \text{ A|(F, W)} \\
&= CRR(X, F|\boldsymbol{W} = \boldsymbol{w}) \\
&\approx COR(X, F|\boldsymbol{W} = \boldsymbol{w})
\end{aligned}
$$

Lorsque $\gamma_V \neq 0$, et donc $X \not\perp\!\!\!\perp \mathcal{A}|(F, W)$, alors

$$
\frac{\mathbb{P}(\mathcal{A} = 1|F = 1, X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(\mathcal{A} = 1|F = 1, X = 0, \boldsymbol{W} = \boldsymbol{w})} \neq \frac{\mathbb{P}(\mathcal{A} = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(\mathcal{A} = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})}.
$$

Dans le cas où X augmente la probabilité de rouler à une vitesse élevée, alors

$$
\frac{\mathbb{P}(\mathcal{A} = 1|F = 1, X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(\mathcal{A} = 1|F = 1, X = 0, \boldsymbol{W} = \boldsymbol{w})} \geq \frac{\mathbb{P}(\mathcal{A} = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(\mathcal{A} = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})},
$$

et donc $COR(X, F|W = w) \leq COR(X, \mathcal{R}|W = w)$. D'autre part, d'après les résultats, $OR(X, F|W = w, \mathcal{A} = 1) \leq COR(X, F|W = w)$ et $OR(X, F|W = w, \mathcal{A} = 1) \leq COR(X, \mathcal{R}|W = w)$. Il n'est donc pas surprenant que $OR(X, F|W = w, \mathcal{A} = 1)$ se rapproche davantage de $COR(X, F|W = w)$ que de $COR(X, \mathcal{R}|W = w)$.

Si X représente le cannabis et réduit la probabilité de rouler à une vitesse élevée (α_X variant dans les négatifs), le raisonnement est similaire. D'une part,

$$\frac{\mathbb{P}(\mathcal{A}=1|F=1,X=1,\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(\mathcal{A}=1|F=1,X=0,\boldsymbol{W}=\boldsymbol{w})} \leq \frac{\mathbb{P}(\mathcal{A}=1|F=1,\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(\mathcal{A}=1|F=1,\boldsymbol{W}=\boldsymbol{w})},$$

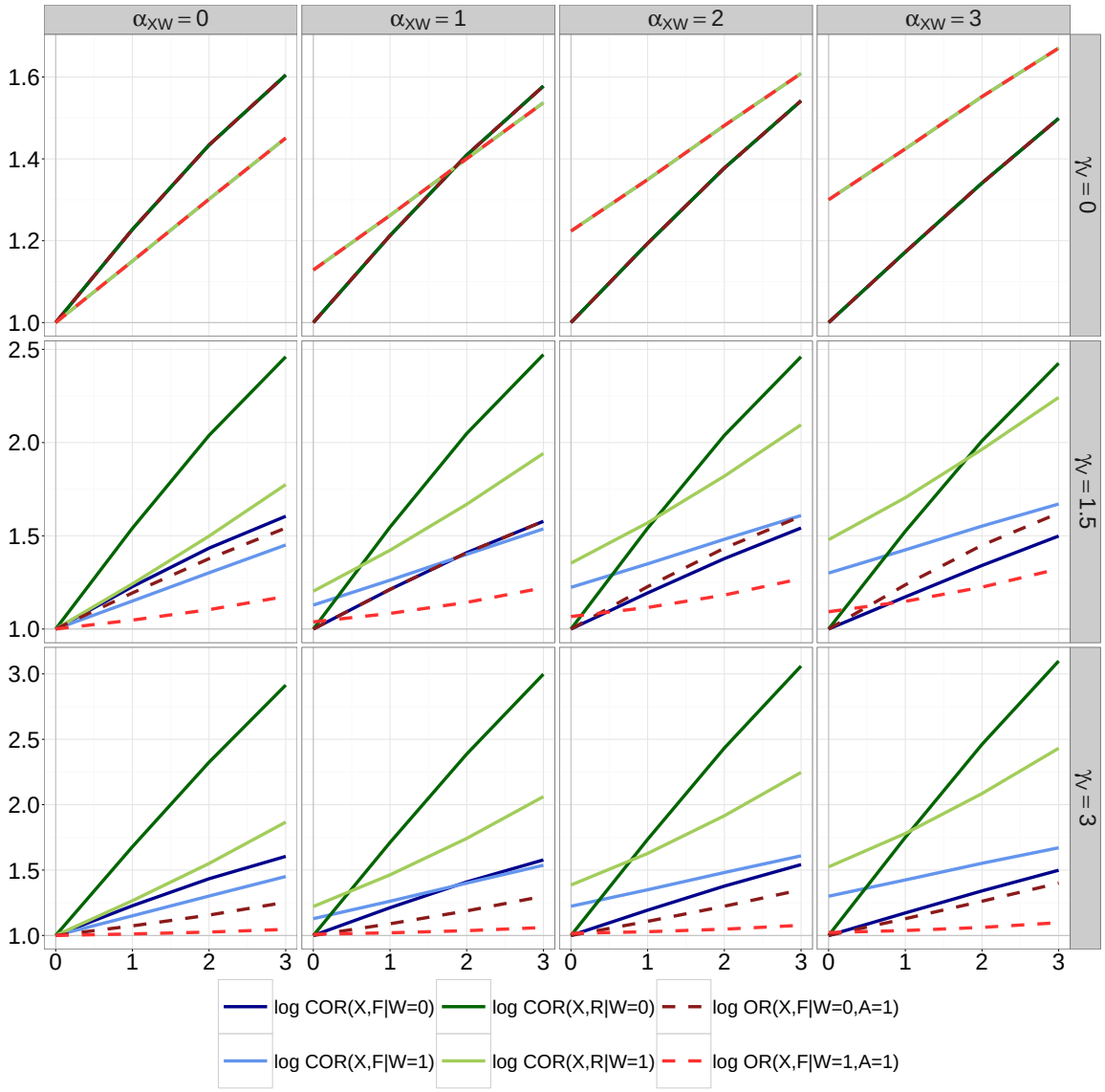
et $COR(X,F|W=w) \geq COR(X,\mathcal{R}|W=w)$. D'autre part, d'après les résultats, $OR(X,F|W=w,\mathcal{A}=1) \geq COR(X,F|W=w)$ et $OR(X,F|W=w,\mathcal{A}=1) \geq COR(X,\mathcal{R}|W=w)$. $OR(X,F|W=w,\mathcal{A}=1)$ a tendance à mieux estimer $COR(X,F|W=w)$ que $COR(X,\mathcal{R}|W=w)$.
Enfin, bien que la variation de α_W modifie les odds ratios causaux et la mesure d'association observée la variation de biais la plus importante observée reste toujours liée à la variation de α_X et γ_V, et non à l'introduction d'un facteur de confusion ou à la variation de α_W.

3.1.3 Résultats en présence d'un facteur de confusion W et d'interaction

En présence d'un facteur de confusion W, l'introduction et la variation des termes d'interactions $\beta_{XW}, \beta_{VW}, \gamma_{FW}, \gamma_{FV}$ modifient marginalement la valeur des odds ratios et l'amplitude du biais, c'est pourquoi ils ne sont pas présentés. Par conséquent, ces coefficients sont fixés par la suite à 0, et nous étudierons la variation du biais en présence des interactions α_{XW} ou β_{XV} pour $\alpha_W = 2$, $\beta_W = 1$ et $\gamma_W = 1$.

L'interaction entre X et W sur la vitesse α_{XW} peut par exemple illustrer une interaction entre le fait de consommer de l'alcool ou du cannabis et être jeune conducteur. On peut faire l'hypothèse que les jeunes conducteurs alcoolisés ont plus tendance à prendre des risques, ce qui se traduit par une vitesse plus élevée. L'écart de probabilité de rouler à une vitesse élevée entre les deux classes d'âge est alors plus important chez les alcoolisés que les non-alcoolisés.
D'après la figure 3.3, et donc quand α_X varie entre 0 et 3, la variation des biais entre les trois quantités dépend de la strate W. Dans la strate où $\{W=0\}$, les biais sont relativement stables : l'écart entre $COR(X,\mathcal{R}|W=0)$, $COR(X,F|W=0)$ et $OR(X,F|W=0,\mathcal{A}=1)$ apparait globalement constant. En revanche, dans la strate où $\{W=1\}$, le biais entre les trois quantités augmente avec α_{XW}. Dans $\{W=0\}$, l'écart à l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F,W)$ est constant alors que dans la strate $\{W=1\}$, on s'éloignerait de cette indépendance. Cette hypothèse est en accord avec les résultats de la table 3.2 où, pour la strate $\{W=1\}$, $\mathbb{P}(\mathcal{A}=1|X=1,F=f,W=1)/\mathbb{P}(\mathcal{A}=1|X=0,F=f,W=1)$ s'éloigne de 1 (cf. la première ligne du tableau 3.2), et pour la strate $\{W=0\}$, $\mathbb{P}(\mathcal{A}=1|X=1,F=f,W=0)/\mathbb{P}(\mathcal{A}=1|X=0,F=f,W=0)$ reste constant (cf. la deuxième ligne du tableau 3.2).
L'augmentation de α_{XW} a davantage d'impact sur $COR(X,\mathcal{R}|W=1)$, et dans une moindre mesure sur $COR(X,F|W=1)$, que sur les autres odds ratios conditionnels. Ce résultat est attendu, puisque comme nous l'avons déjà montré, dans la strate où $\{W=1\}$, on s'éloigne de la condition $X \perp\!\!\!\perp \mathcal{A}|(F,W)$ avec l'augmentation de α_{XW}. Par conséquent, $\mathbb{P}(\mathcal{A}=1|F=1,X=1,W=1)/\mathbb{P}(\mathcal{A}=1|F=1,X=0,W=1)$ devient de plus en plus supérieur à 1, et donc, $COR(X,\mathcal{R}|W=1)$ devient de plus en plus supérieur à $COR(X,F|W=1)$.
Notons également qu'avec cette interaction, il existe un biais entre les trois quantités étudiées

FIGURE 3.3 – Odds ratios causaux et odds ratio estimable sur l'échelle logarithmique dans le cas où $\beta_V = \beta_X = \beta_W = \gamma_W = 1$, $\gamma_F = 4, \alpha_W = 2$, et pour des valeurs variables des autres paramètres $\alpha_X, \alpha_{XW}, \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_X varie de 0 à 3



Lecture (3ème cadran de la premiere colonne) : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et qu'il n'existe pas d'interaction entre X et W sur la vitesse($\alpha_X W = 0$). Pour $\alpha_X = 1$, les logarithmes de l'odds ratio estimable, conditionné sur W (en pointillé rouge) sont proches de 1, les logarithmes de l'odds ratio causal sur la faute (en trait plein bleu), ainsi que le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave dans la strate où $\{W = 1\}$ (en trait plein vert clair) sont légèrement supérieurs. Le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave dans la strate où $\{W = 0\}$ (en trait plein vert foncé) est environ égal à 1,70. Les trois quantités conditionnées sur $\{W = w\}$ pour $w \in (0, 1)$ n'étant pas égales, il existe un biais.

même lorsque $\alpha_X = 0$ dans la strate où $\{W = 1\}$. En effet, dans ce cas, $X \not\perp\!\!\!\perp \mathcal{A}|(F, W)$ et le risque relatif $\mathbb{P}(\mathcal{A} = 1|X = 1, F = f, W = 1)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = f, W = 1)$ est différent

de 1, car il dépend de α_{XW}, et non plus seulement de α_X. L'écart entre ces trois quantités est d'autant plus grand que α_{XW} est grand. En présence de l'interaction α_{XW}, les odds ratios causaux conditionnés sur $W = 1$ sont donc supérieurs aux odds ratios causaux conditionnés sur $W = 0$ lorsque $\alpha_X = 0$. Lorsque α_X augmente, les courbes relatives aux odds ratios causaux se croisent. Plus α_{XW} est grand, plus le seuil α_X où les courbes se croisent est grand.

TABLE 3.2 – Risques relatifs $\mathbb{P}(\mathcal{A} = 1|X = 1, F = f, W = w)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = f, W = w)$ pour $\alpha_X = 1$ et $\gamma_V = 3$ et différentes valeurs de α_{XW}

	$\alpha_{XW} = 0$	$\alpha_{XW} = 1$	$\alpha_{XW} = 2$	$\alpha_{XW} = 3$
$\{W = 1\}$	1,28	1,55	1,81	2,09
$\{W = 0\}$	1,83	1,86	1,87	1,86

Lecture : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et lorsque l'interaction entre X et W a un faible effet sur la vitesse ($\alpha_X W = 1$). Le risque relatif dans la strate où $\{W = 1\}$, $\mathbb{P}(\mathcal{A} = 1|X = 1, F = 1, W = 1)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = 1, W = 1)$ est alors égal à 1,55. Le risque relatif dans la strate où $\{W = 0\}$, $\mathbb{P}(\mathcal{A} = 1|X = 1, F = 1, W = 0)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = 1, W = 0)$ est alors égal à 1,86.

Lorsque α_X varie entre -3 et 0 (cf. figure 3.4), et dès lors que $\gamma_V \neq 0$, l'interaction α_{XW} introduit aussi un biais entre les trois quantités d'intérêt lorsque $\alpha_X \neq 0$. Dans la strate où $\{W = 0\}$, les biais sont relativement stables avec l'augmentation de α_{XW}. Dans cette strate, l'augmentation de α_{XW} modifie très marginalement l'écart à l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F, W)$, i.e. $\mathbb{P}(\mathcal{A} = 1|X = 1, F = f, W = 0)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = f, W = 0)$ est constant (cf. deuxième ligne du tableau 3.3). En revanche, dans la strate où $\{W = 1\}$, l'augmentation de α_{XW} a tendance, dans un premier temps, à réduire les biais entre les trois quantités puisque $\mathbb{P}(\mathcal{A} = 1|X = 1, F = f, W = 1)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = f, W = 1)$ se rapproche, et atteint 1, puis ensuite s'en éloigne (cf. première ligne du tableau 3.3). Ce résultat est attendu puisque α_X varie de 0 à -3 alors que α_{XW} varie de -1 à 2. Par conséquent, $\mathbb{P}(\mathcal{A} = 1|X = 1, F = f, W = 1)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = f, W = 1) = 1$ lorsque $\alpha_X + \alpha_{XW} = 0$.

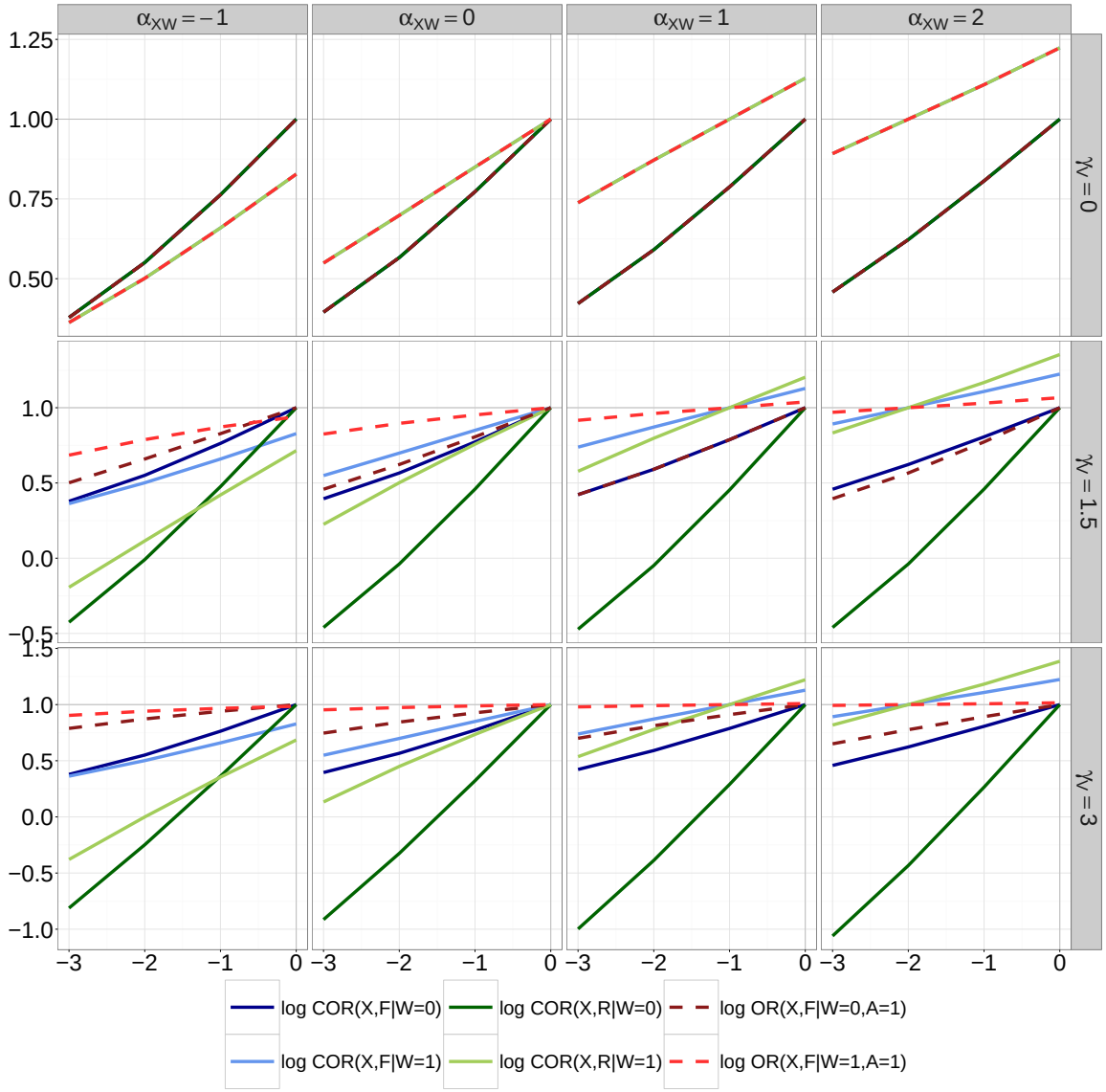
TABLE 3.3 – Risques relatifs $\mathbb{P}(\mathcal{A} = 1|X = 1, F = f, W = w)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = f, W = w)$ pour $\alpha_X = -1$ et $\gamma_V = 3$ et différentes valeurs de α_{XW}

	$\alpha_{XW} = -1$	$\alpha_{XW} = 0$	$\alpha_{XW} = 1$	$\alpha_{XW} = 2$
$\{W = 1\}$	0,54	0,77	1,00	1,19
$\{W = 0\}$	0,56	0,55	0,53	0,53

Lecture : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et lorsque l'interaction entre X et W a un faible effet sur la vitesse ($\alpha_X W = -1$). Le risque relatif dans la strate où $\{W = 1\}$, $\mathbb{P}(\mathcal{A} = 1|X = 1, F = 1, W = 1)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = 1, W = 1)$ est alors égal à 0,77. Le risque relatif dans la strate où $\{W = 0\}$, $\mathbb{P}(\mathcal{A} = 1|X = 1, F = 1, W = 0)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = 1, W = 0)$ est alors égal à 0,55.

L'introduction du terme d'interaction β_{XV} modifie fortement la valeur des odds ratios mais pas l'ampleur des biais (cf figures 3.5 et 3.6). Cela signifie que l'introduction et la variation de l'interaction β_{XV} ne modifie pas l'écart à l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F, W)$, i.e. le risque relatif $\mathbb{P}(\mathcal{A} = 1|X = 1, F = f, W = w)/\mathbb{P}(\mathcal{A} = 1|X = 0, F = f, W = w)$ reste stable. Lorsque α_X varie de 0 à 3, la valeur des risques relatifs pour les différentes valeurs de β_{XV} est présentée dans

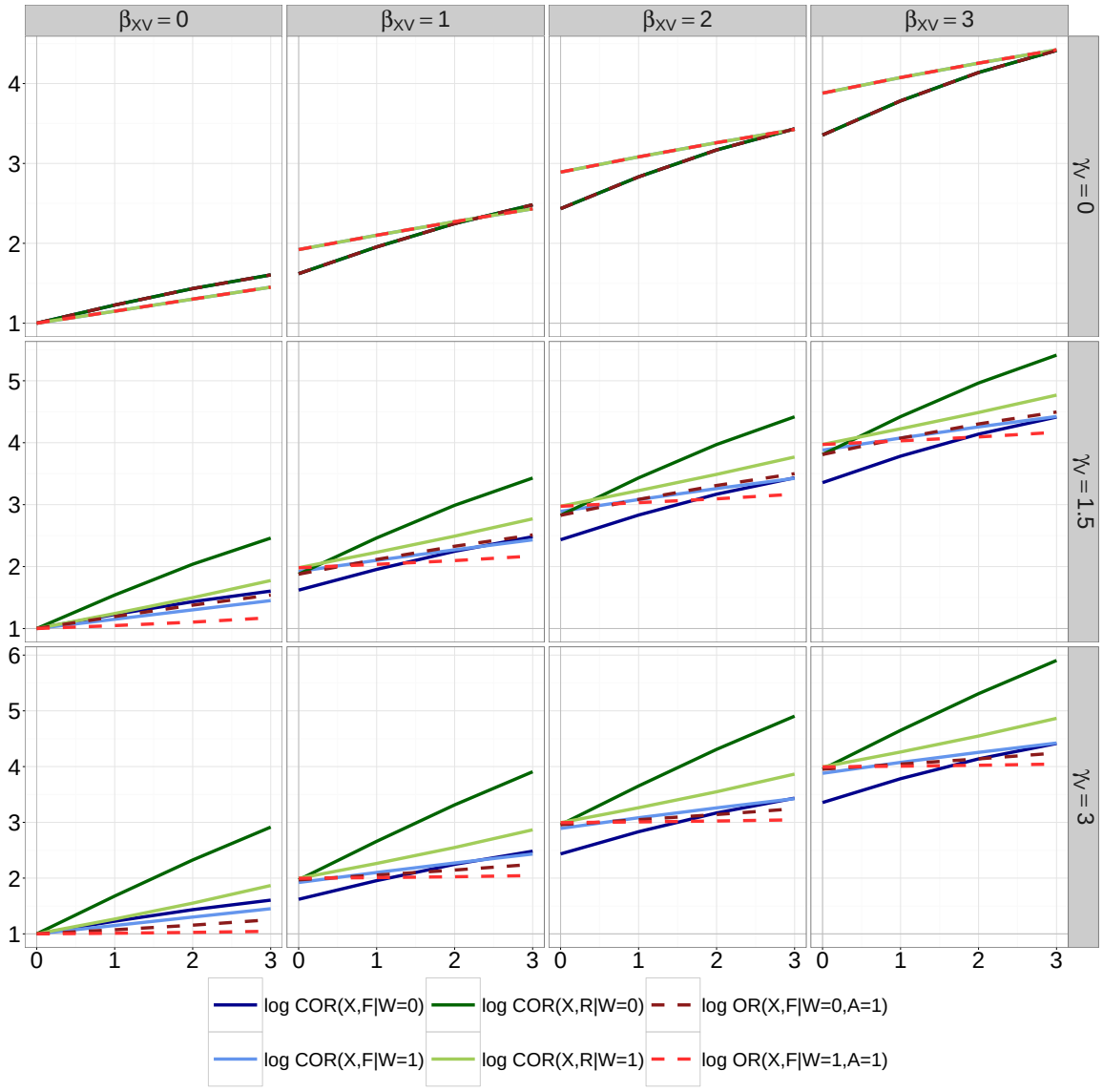
FIGURE 3.4 – Odds ratios causaux et odds ratio estimable sur l'échelle logarithmique dans le cas où $\beta_V = \beta_X = \beta_W = \gamma_W = 1$, $\gamma_F = 4, \alpha_W = 2$, et pour des valeurs variables des autres paramètres $\alpha_X, \alpha_{XW}, \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_X varie de -3 à 0.



Lecture (3ème cadran de la premiere colonne) : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et qu'il n'existe pas d'interaction entre X et W sur la vitesse ($\alpha_X W = 0$). Pour $\alpha_X = -3$, les logarithmes de l'odds ratio estimable, conditionné sur W (en pointillé rouge) sont proches de 0,9, les logarithmes de l'odds ratio causal sur la faute (en trait plein bleu) sont environ égaux à 0,4. Les logarithme de l'odds ratio causal sur la responsabilité de l'accident grave sont inférieurs à 0. Les trois quantités conditionnées sur $\{W = w\}$ pour $w \in (0, 1)$ n'étant pas égales, il existe un biais.

le tableau 3.4. Quand α_X varie de -3 à 0, les résultats sont présentés dans le tableau 3.5. Et effectivement, dans les deux cas, les risques relatifs restent stables. Notons que l'introduction de l'interaction introduit un biais entre les trois quantités lorsque $\alpha_X = 0$ dès que $\gamma_V \neq 0$.

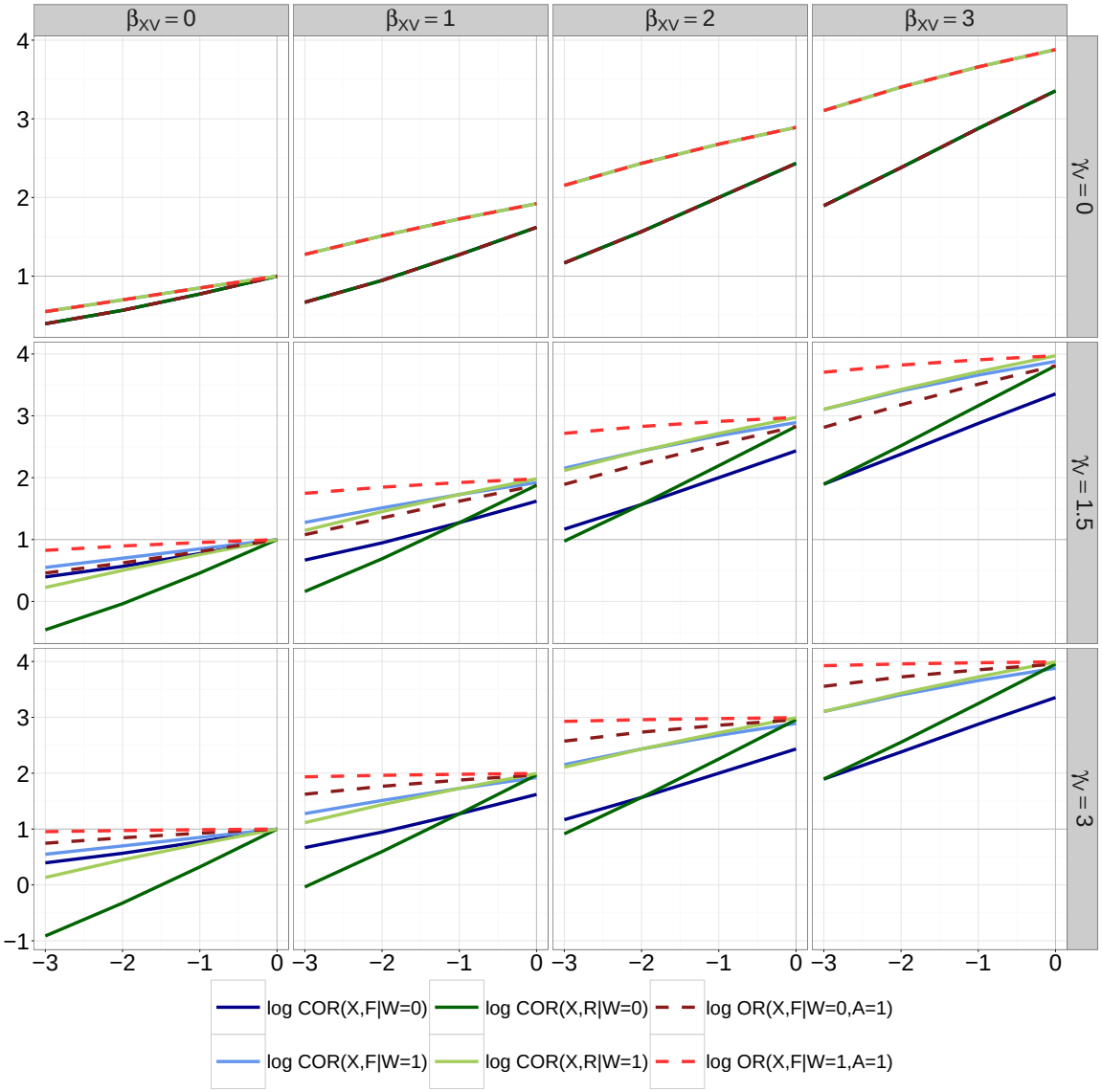
FIGURE 3.5 – Odds ratios causaux et odds ratio estimable sur l'échelle logarithmique dans le cas où $\beta_V = \beta_X = \beta_W = \gamma_W = 1$, $\gamma_F = 4, \alpha_W = 2$, et pour des valeurs variables des autres paramètres $\alpha_X, \beta_{XV}, \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_X varie de 0 à 3.



Lecture (3ème cadran de la premiere colonne) : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et qu'il n'existe pas d'interaction entre X et V sur la faute ($\beta_X V = 0$). Pour $\alpha_X = 1$, les logarithmes de l'odds ratio estimable, les logarithmes de l'odds ratio causal sur la faute (en trait plein bleu), ainsi que le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave dans la strate où $\{W = 1\}$ (en trait plein vert clair) sont compris entre 1 et 1,3. Le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave dans la strate où $\{W = 0\}$ (en trait plein vert foncé) est presque égal à 1,70. Les trois quantités conditionnées sur $\{W = w\}$ pour $w \in (0, 1)$ n'étant pas égales, il existe un biais.

Dans le cas binaire, nous avons montré que l'ampleur du biais entre les trois quantités d'intérêt dépendait de l'écart à l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F, W)$, et donc en particulier des paramètres α_X

FIGURE 3.6 – Odds ratios causaux et odds ratio estimable sur l'échelle logarithmique dans le cas où $\beta_V = \beta_X = \beta_W = \gamma_W = 1$, $\gamma_F = 4, \alpha_W = 2$, et pour des valeurs variables des autres paramètres $\alpha_X, \beta_{XV}, \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_X varie de -3 à 0.



Lecture (3ème cadran de la premiere colonne) : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et qu'il n'existe pas d'interaction entre X et V sur la faute ($\beta_X V = 0$). Pour $\alpha_X = -1$, les logarithmes de l'odds ratio estimable, les logarithmes de l'odds ratio causal sur la faute (en trait plein bleu), ainsi que le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave dans la strate où $\{W = 1\}$ (en trait plein vert clair) sont compris entre 0,5 et 1. Le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave dans la strate où $\{W = 0\}$ (en trait plein vert foncé) est égal à 0,2. Les trois quantités conditionnées sur $\{W = w\}$ pour $w \in (0, 1)$ n'étant pas égales, il existe un biais.

et γ_V. La présence d'un facteur de confusion W, mais surtout l'augmentation de son effet sur le fait de rouler à une vitesse élevée V modifie cet écart à l'indépendance, mais dans une moindre

TABLE 3.4 – Risques relatifs $\mathbb{P}(\mathcal{A}=1|X=1,F=f,W=w)/\mathbb{P}(\mathcal{A}=1|X=0,F=f,W=w)$ pour $\alpha_X=1$ et $\gamma_V=3$ et différentes valeurs de β_{XV}

	$\beta_{XV}=0$	$\beta_{XV}=1$	$\beta_{XV}=2$	$\beta_{XV}=3$
$\{W=1\}$	1,28	1,29	1,29	1,29
$\{W=0\}$	1,83	1,83	1,83	1,83

Lecture : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V=3$), et lorsque l'interaction entre X et V a un faible effet sur la faute ($\beta_{XV}=1$). Le risque relatif dans la strate où $\{W=1\}$, $\mathbb{P}(\mathcal{A}=1|X=1,F=1,W=1)/\mathbb{P}(\mathcal{A}=1|X=0,F=1,W=1)$ est alors égal à 1,29. Le risque relatif dans la strate où $\{W=0\}$, $\mathbb{P}(\mathcal{A}=1|X=1,F=1,W=0)/\mathbb{P}(\mathcal{A}=1|X=0,F=1,W=0)$ est alors égal à 1,83.

TABLE 3.5 – Risques relatifs $\mathbb{P}(\mathcal{A}=1|X=1,F=f,W=w)/\mathbb{P}(\mathcal{A}=1|X=0,F=f,W=w)$ pour $\alpha_X=-1$ et $\gamma_V=3$ et différentes valeurs de β_{XV}

	$\beta_{XV}=0$	$\beta_{XV}=1$	$\beta_{XV}=2$	$\beta_{XV}=3$
$\{W=1\}$	0,77	0,77	0,78	0,78
$\{W=0\}$	0,55	0,54	0,55	0,55

Lecture : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V=3$), et lorsque l'interaction entre X et V a un faible effet sur la faute ($\beta_{XV}=1$). Le risque relatif dans la strate où $\{W=1\}$, $\mathbb{P}(\mathcal{A}=1|X=1,F=1,W=1)/\mathbb{P}(\mathcal{A}=1|X=0,F=1,W=1)$ est alors égal à 0,77. Le risque relatif dans la strate où $\{W=0\}$, $\mathbb{P}(\mathcal{A}=1|X=1,F=1,W=0)/\mathbb{P}(\mathcal{A}=1|X=0,F=1,W=0)$ est alors égal à 0,54.

mesure que la variation de α_X ou γ_V. Il en est de même avec l'interaction entre X et W sur V. En revanche, l'introduction et l'augmentation de l' interaction entre X et V sur la faute F ne modifie pas l'ampleur des biais mais seulement la valeur des odds ratios.
Concernant le sens des biais, lorsque X a un effet positif sur la vitesse, et donc que α_X varie de 0 à 3, l'OR estimable à partir des analyses en responsabilité a tendance à sous-estimer les odds ratios causaux, et ce d'autant plus que α_X et γ_V sont grands. Lorsque X a un effet négatif sur la vitesse, et donc que α_X varie de -3 à 0, l'OR estimable a au contraire tendance à sur-estimer les odds ratios causaux. Le rapport entre les logarithmes des différents odds ratios varie entre 0 et 1.5, ce qui signifie que le biais entre les différentes quantités se situe dans un rapport de 1 à 5. Enfin, dans le cas où α_X varie dans les négatifs, un biais d'ordre qualitatif est observable. Il serait donc possible de conclure à un effet délétère de l'exposition considérée, alors qu'elle aurait en réalité un effet protecteur.

3.2 Illustration lorsque X est catégorielle

Dans cette partie, nous considérons cette fois-ci X comme variable catégorielle à $J+1$ modalités $x_j \in (x_0, x_1, ..., x_J)$. X est alors transformée en J variables binaires, notée X_j, valant 1 si la variable X vaut x_j et 0 sinon, pour $j \geq 1$. Par conséquent, il s'agit pour chaque variable binaire X_j de comparer les quantités $COR(X_j, F|W=w)$, $COR(X_j, \mathcal{R}|W=w)$ et $OR(X_j, F|W=w, \mathcal{A}=1)$ définis de la façon suivante :

$$COR(X_j, F|W=w) = \frac{\mathbb{P}(F_j=1|W=w)/\mathbb{P}(F_j=0|W=w)}{\mathbb{P}(F_0=1|W=w)/\mathbb{P}(F_0=0|W=w)},$$
$$COR(X_j, \mathcal{R}|W=w) = \frac{\mathbb{P}(\mathcal{R}_j=1|W=w)/\mathbb{P}(\mathcal{R}_j=0|W=w)}{\mathbb{P}(\mathcal{R}_0=1|W=w)/\mathbb{P}(\mathcal{R}_0=0|W=w)},$$
$$OR(X_j, F|W=w, \mathcal{A}=1) = \frac{\mathbb{P}(F=1|X=x_j, W=w, \mathcal{A}=1)/\mathbb{P}(F=0|X=x_j, W=w, \mathcal{A}=1)}{\mathbb{P}(F=1|X=x_0, W=w, \mathcal{A}=1)/\mathbb{P}(F=0|X=x_0, W=w, \mathcal{A}=1)}.$$

où F_j, respectivement $\mathcal{R}_j$, est l'issue contrefactuelle F, respectivement $\mathcal{R}$, que l'on aurait observé si l'exposition avait été fixée à $X = x_j$ pour $x_j \in (x_0, x_1, x_2, x_3, x_4)$.

Nous décrivons rapidement le modèle causal considéré ici. Ensuite, nous faisons varier différents paramètres d'intérêt, sans, puis avec un facteur de confusion W. Cette analyse se fait dans l'idée de représenter par exemple le biais induit par les analyses en responsabilité lorsque X représente les différents taux d'alcoolémie.

3.2.1 Modèle causal considéré

X est cette fois-ci une variable catégorielle à $j = 5$ modalités : x_0 la référence, et x_1, x_2 ,x_3 ,x_4 les autres modalités. Les 5 variables binaires X_j, où $j \in (0, 1, 2, 3, 4)$ sont définies telles que $p_{X_j} = 0.2$ pour tout j. $W \sim B(p_W)$ suit par ailleurs toujours une loi de Bernoulli où $p_W = 0.5$. Nous introduisons ensuite pour les paramètres suivants $\alpha_0, \alpha_{X_1}, \alpha_{X_2}, \alpha_{X_3}, \alpha_{X_4}, \alpha_W, \beta_0, \beta_{X_1}, \beta_{X_2}, \beta_{X_3}, \beta_{X_4}, \beta_V, \beta_W, \beta_{VW}, \gamma_0, \gamma_F, \gamma_V, \gamma_W, \gamma_{FV}, \gamma_{FW}, \gamma_{VW}$:

$$\begin{aligned} p_V(x_j, w) &= \mathrm{h}(\alpha_0 + \alpha_{X_j} x_j + \alpha_W w) \\ p_F(x_j, v, w) &= \mathrm{h}(\beta_0 + \beta_{X_j} x_j + \beta_V v + \beta_W w + \beta_{VW} vw) \\ p_A(f, v, w) &= \mathrm{h}(\gamma_0 + \gamma_F f + \gamma_V v + \gamma_W w + \gamma_{FV} fv + \gamma_{FW} fw + \gamma_{VW} vw). \end{aligned}$$

Pour simplifier, nous ne considérons pas les interactions entre les différentes modalités de X et les autres variables W ou V.

Par ailleurs, pour $\nu = 13$, nous posons les constantes suivantes :

$$\begin{aligned} \alpha_0 &= -\frac{1}{5}\left(\alpha_{X_1} + \alpha_{X_2} + \alpha_{X_3} + \alpha_{X_4} + \frac{5}{2}\alpha_W\right) \\ \beta_0 &= -\frac{1}{5}\left(\beta_{X_1} + \beta_{X_2} + \beta_{X_3} + \beta_{X_4} + \frac{5}{2}(\beta_V + \beta_W) + \frac{5}{4}\beta_{VW} - \nu\right) \\ \gamma_0 &= -\frac{1}{2}\left(\gamma_F + \gamma_V + \gamma_W + \frac{1}{2}(\beta_{FV} + \beta_{FW} + \beta_{VW}) - \nu\right) \end{aligned}$$

De la même manière que dans le cas binaire, l'importance du biais entre les trois odds ratios d'intérêt dépendra de l'écart à l'indépendance $X_j \perp\!\!\!\perp \mathcal{A}|(F, W)$. Le fait que cette indépendance ne soit pas vérifiée dépend donc en premier lieu de la valeur des paramètres α_{X_j} et γ_V. Nous faisons également varier les autres paramètres du modèle afin de déterminer dans quel sens et

quelle ampleur sont modifiés les biais entre les trois quantités.

3.2.2 Résultats en l'absence de facteurs de confusion

Nous optons pour un modèle génératif visant à se rapprocher d'une situation où X représenterait le taux d'alcoolémie. Le taux d'alcoolémie contribue à augmenter la vitesse pour les faibles doses, et à la diminuer pour les doses les plus fortes (Kelly et al., 2004, Ronen et al., 2010). D'autre part, plus le taux d'alcool est élevé, plus le risque de commettre une faute de conduite est grand. Nous choisissons les relations suivantes entre les différents coefficients afin d'obtenir des mesures d'associations proches de celles observées sur la base de données VOIESUR, qui a l'avantage de présenter une estimation fiable de la vitesse initiale :

$$\begin{cases} \alpha_{X_1} \\ \alpha_{X_2} = 0.80\alpha_{X_1} \\ \alpha_{X_3} = 0.25\alpha_{X_1} \\ \alpha_{X_4} = 0.20\alpha_{X_1} \end{cases} \qquad \begin{cases} \beta_{X_1} \\ \beta_{X_2} = 2.50\beta_{X_1} \\ \beta_{X_3} = 3.50\beta_{X_1} \\ \beta_{X_4} = 3.40\beta_{X_1} \end{cases}$$

Nous présentons nos résultats toujours en posant $\beta_V = 1$ et $\gamma_F = 4$, puisque V augmente le risque de commettre une faute de conduite et F augmente fortement le risque d'avoir un accident. Ensuite, nous faisons varier α_{X_1} et γ_V entre 0 et 3 en fixant $\beta_{X_1} = 1$. Les résultats sont présentés à travers la figure 3.7.

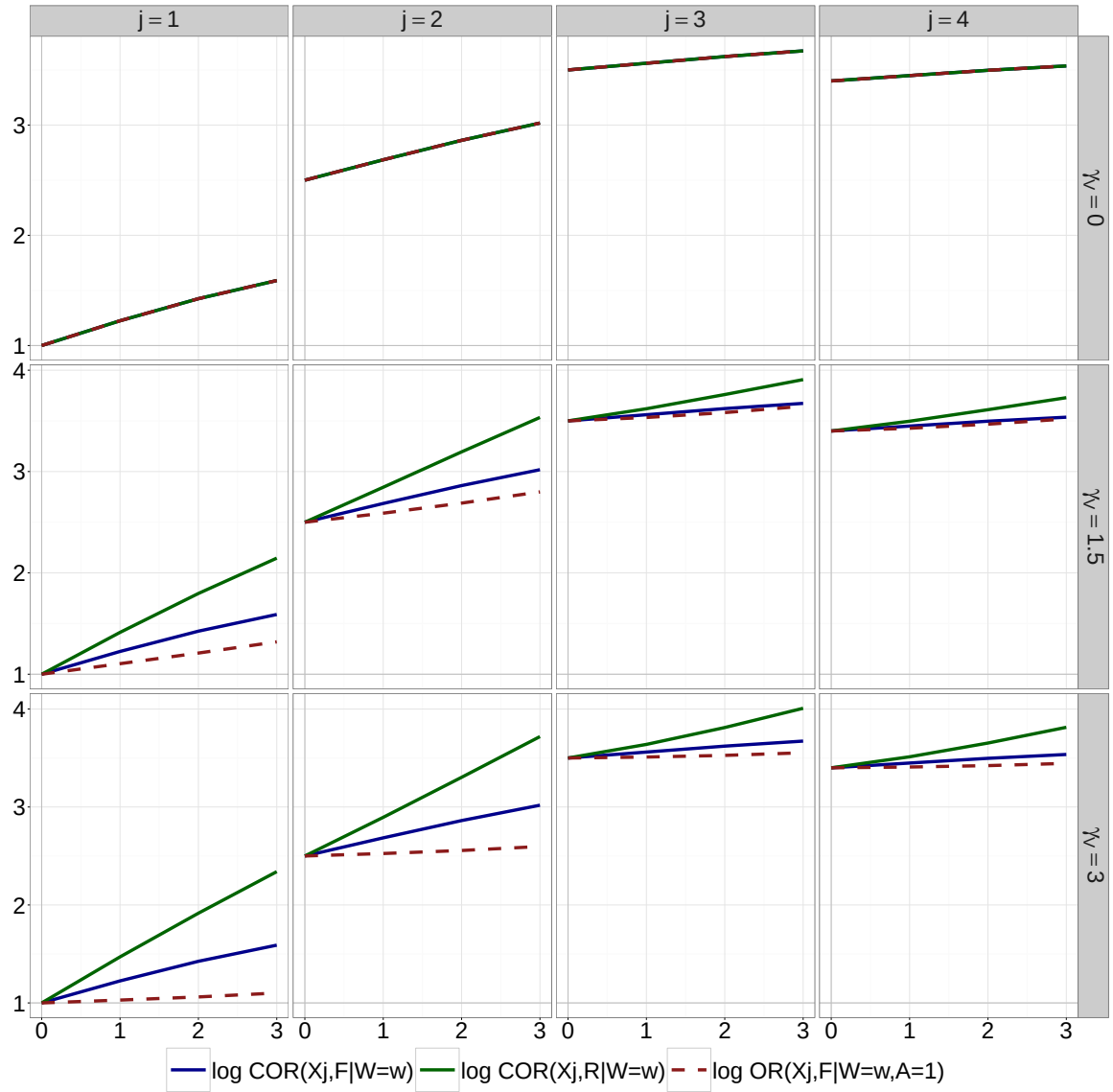
Nous retrouvons les résultats observés dans le cas binaire. Pour chaque modalité j, quand $\gamma_V = 0$, $X_j \perp\!\!\!\perp \mathcal{A}|(F, W)$ et nous observons l'absence de biais entre $COR(X_j, F|W = w)$, $COR(X_j, \mathcal{R}|W = w)$ et $OR(X, F|W = w, \mathcal{A} = 1)$, quelle que soit la valeur de α_{X_1}. Dès que $\gamma_V \neq 0$ et dès lors que $\alpha_{X_1} \neq 0$, il existe un biais entre les trois quantités considérées, et l'amplitude de ce biais augmente avec α_{X_1}, et donc avec α_{X_j}. Les biais sont en effet d'autant plus grands que α_{X_j} et γ_V sont grands, car on s'éloigne de l'indépendance $X_j \perp\!\!\!\perp \mathcal{A}|(F, W)$. Notons également que plus la relation entre α_{X_j} et α_{X_1} est faible, plus le biais entre les trois quantités est faible, ce qui est également attendu. Enfin, puisque chaque variable X_j augmente la probabilité de rouler à une vitesse élevée V, $OR(X, F|W = w, \mathcal{A} = 1)$ sous estime les deux oods ratios causaux $COR(X_j, F|W = w)$ et $COR(X_j, \mathcal{R}|W = w)$.

3.2.3 Résultats avec facteurs de confusion

Nous introduisons un facteur de confusion W ayant un impact positif sur V, et faisons varier α_W entre 0 et 3. Les résultats pour la modalité $j = 1$ sont illustrées par la figure 3.8. Les résultats pour les autres modalités sont présentées en annexes.

De la même manière que dans le cas binaire, quel que soit γ_V et α_W, et pour tout $W \in (0, 1)$, l'écart entre $COR(X_1, \mathcal{R}|W = w)$, $COR(X_1, F|W = w)$, et $OR(X_1, F|W = w, \mathcal{A} = 1)$ augmente avec α_{X_1}, comme attendu. Les variations de biais lorsque α_W dépendent de la strate W. A une variation près, lorsque α_W augmente, on s'éloigne de l'indépendance $X_j \perp\!\!\!\perp \mathcal{A}|(F, W)$ dans la strate $\{W = 0\}$, puisque le biais augmente, et à l'inverse on s'en rapproche dans la strate $\{W = 1\}$, le biais diminuant. Enfin, notons qu'en présence du facteur de confusion, l'OR estimable dans

FIGURE 3.7 – Odds ratios causaux et odds ratio estimable pour chaque modalité j de X sur l'échelle logarithmique dans le cas où $\beta_V = \beta_{X_1} = 1$, $\gamma_F = 4$ et pour des valeurs variables des autres paramètres α_{X_1} et γ_V. Dans chaque cadran, le long de l'axe des abscisses, α_{X_1} varie de 0 à 3.

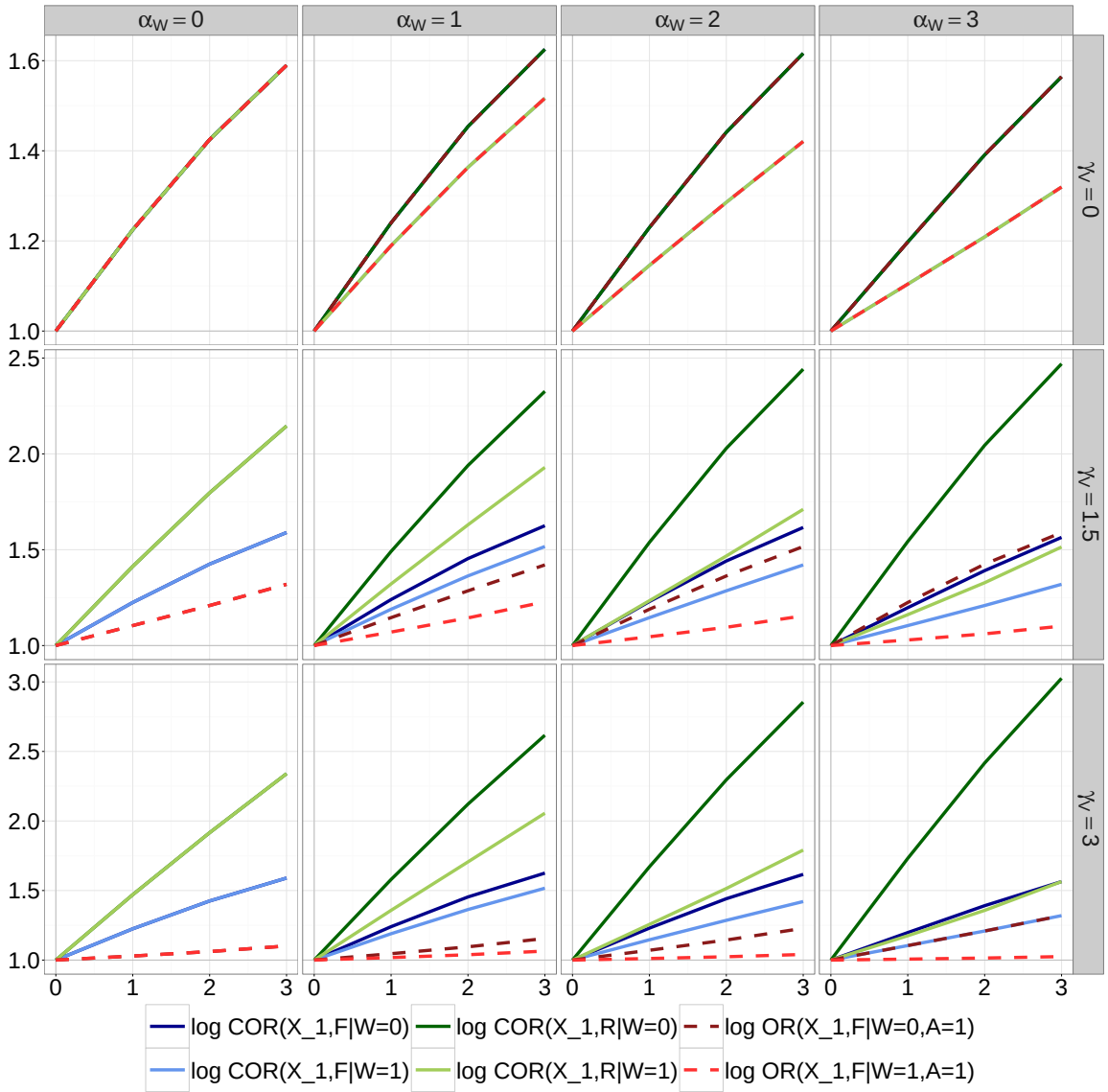


Lecture (3ème cadran de la premiere colonne) : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et les odds ratios considérés concernent la première modalité de la variable X ($j = 1$). Pour $\alpha_X = 1$, le logarithme de l'odds ratio estimable (en pointillé rouge) est proche de 1, le logarithme de l'odds ratio causal sur la faute (en trait plein bleu) est environ égal à 1,2 et le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave (en trait plein vert) est à 1,5. Les trois quantités n'étant pas égales, il existe un biais.

les analyses en responsabilité correspond une fois de plus à une sous-estimation des odds ratios causaux lorsque X_1 a un effet positif sur la vitesse.

Les mêmes conclusions sont observables pour les autres variables X_j (cf. annexe .2), et les biais

FIGURE 3.8 – Odds ratios causaux et odds ratio estimable pour la modalité $j = 1$ de X sur l'échelle logarithmique dans le cas où $\beta_V = \beta_{X_1} = \beta_W = \gamma_W = 1$, $\gamma_F = 4$ et pour des valeurs variables des autres paramètres α_{X_1} et γ_V. Dans chaque cadran, le long de l'axe des abscisses, α_{X_1} varie de 0 à 3.



Lecture (3ème cadran de la premiere colonne) : Nous considérons le cas où l'inclusion des données dépend fortement de la gravité de l'inclusion ($\gamma_V = 3$), et lorsque le facteur de confusion n'a pas d'effet sur la vitesse ($\alpha_W = 0$). Pour $\alpha_X = 1$, le logarithme de l'odds ratio estimable (en pointillé rouge) est proche de 1, le logarithme de l'odds ratio causal sur la faute (en trait plein bleu) est environ égal à 1,2 et le logarithme de l'odds ratio causal sur la responsabilité de l'accident grave (en trait plein vert) est à 1,5. Les trois quantités n'étant pas égales, il existe un biais.

entre les trois quantités est d'autant plus faibles que la relation entre α_X et α_{XW} est faible.

Conclusion

Dans le chapitre 2, nous avons notamment montré que l'identifiabilité de $COR(X, F|W = w)$, puis l'approximation de $COR(X, \mathcal{R}|W = w)$, dépendait de l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F, W)$. Dès que cette condition n'est pas vérifiée, l'OR estimable dans les analyses en responsabilité $OR(X, F|W = w, \mathcal{A} = 1)$ n'a aucune valeur causale et reste une mesure biaisée de $COR(X, F|W = w)$ ou $COR(X, \mathcal{R}|W = w)$. Dans ce chapitre 3, nous avons décrit un modèle génératif simple pour illustrer l'ampleur, le sens et la variation des biais entre les trois quantités d'intérêt en fonction de cet écart à l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F, W)$. L'indépendance n'est pas vérifiée lorsqu'il existe une flèche entre X et V et une flèche entre V et $\mathcal{A}$, par conséquent, l'importance du biais dépend en premier lieu des paramètres α_X et γ_V. Nous faisons également varier les autres paramètres du modèle afin de déterminer dans quel sens et quelle ampleur sont modifiés les biais entre les trois quantités. Dans le cas binaire, et en l'absence d'interaction, les résultats numériques confirment l'absence de biais entre $COR(X, F|W = w)$, $COR(X, \mathcal{R}|W = w)$ et $OR(X, F|W = w, A = 1)$ lorsque $\alpha_X = 0$ ou $\gamma_V = 0$. D'autre part, dès que ces paramètres sont non nuls, le biais entre les trois quantités est d'autant plus grand que α_X et γ_V sont grands puisqu'on s'éloigne de l'indépendance $X \perp\!\!\!\perp \mathcal{A}|(F, W)$. L'introduction d'un facteur de confusion W ou d'interactions peut modifier l'ampleur du biais mais pas autant que la variation de α_X ou γ_V. Précisons que, pour des valeurs plausibles des autres coefficients et interactions, l'ampleur des biais entre les trois quantités, dans le modèle simple considéré ici, est compris dans un rapport de 1 à 5. Quand X est corrélée positivement à V, l'OR observé $OR(X, F|W = w, A = 1)$ dans les analyses en responsabilité correspond à une sous-estimation des odds ratios causaux $COR(X, F|W = w)$ et $COR(X, \mathcal{R}|W = w)$. A l'inverse, quand X est corrélée négativement à V, l'OR estimable $OR(X, F|W = w, A = 1)$ dans les analyses en responsabilité correspond à une sur-estimation des odds ratios causaux $COR(X, F|W = w)$ et $COR(X, \mathcal{R}|W = w)$. Dans ce deuxième exemple, le biais peut même être qualitatif : il est possible d'observer un OR supérieur à 1 et conclure à l'effet délétère de l'exposition considérée sur la responsabilité de l'accident grave, alors qu'elle aurait en réalité un effet protecteur. Dans le cas où X est catégorielle, les résultats sont semblables au cas binaire lorsque X est corrélée positivement à V et à F. Cela veut dire que le biais entre l'odds ratio estimable dans les analyses en responsabilité et les odds ratios causaux diminue avec le taux d'alcoolémie, car plus l'alcoolémie est élevée, moins la probabilité de rouler à une vitesse élevée est forte (Kelly et al., 2004, Ronen et al., 2010). Avec ou sans facteur de confusion, l'OR estimé dans les analyses en responsabilité pour n'importe quel taux d'alcoolémie est une sous-estimation des odds ratios causaux $COR(X, F|W = w)$ et $COR(X, \mathcal{R}|W = w)$.

Ces résultats restent des résultats préliminaires, obtenus à l'aide d'un modèle génératif simple. Ils montrent notamment qu'il n'est pas si aisé d'anticiper l'ampleur et le sens de variation des biais entre les différentes quantités conditionnellement à W. D'autre part, notre modèle comporte des limites. Certaines hypothèses peuvent être remises en question comme l'équirépartition des modalités de X dans le cas binaire ou catégoriel, ou encore la prévalence de V, $\mathbb{P}(V = 1) = 0.5$. Notons également que le modèle considère un seul facteur de confusion binaire. Il serait ainsi intéressant de considérer dans d'autres recherches $W \in \mathbb{R}^p$.

Chapitre 4

Biais de collision et paradoxe de l'obésité

Ce chapitre a fait l'objet d'une lettre de recherche « Re :Collider bias is only a partial explanation for the obesity paradox » publiée dans Epidemiology (Viallon and Dufournet, 2017b), en réponse à un article de Sperrin et al. (2016) publiée dans la même revue. Une version plus complète de cette lettre est également disponible sur Arxiv (Viallon and Dufournet, 2017a).

L'exemple des analyses en responsabilité a montré que conditionner sur un descendant à la fois de l'événement d'intérêt et de l'exposition conduisait à introduire un biais entre la mesure d'association estimée en présence de sélection préférentielle des individus et le véritable effet causal. Les illustrations numériques du chapitre 3 ont mis en avant la majorité du temps un biais quantitatif. Lorsque l'exposition X augmente la probabilité de rouler à une vitesse élevée, la mesure d'association estimée dans les analyses en responsabilité où l'inclusion dépend de la sévérité de l'accident correspond à une sous-estimation de l'effet causal. Tandis que, lorsque l'exposition X diminue la probabilité de rouler à une vitesse élevée, la mesure d'association estimée est une sur-estimation de l'effet causal.

Dans d'autres cas, le biais de collision peut aboutir à un biais qualitatif, et donc à des résultats sans valeur causale. C'est le cas du célèbre paradoxe de l'obésité, où lorsque l'inférence est réalisée chez des patients souffrant d'une maladie chronique (diabète (Carnethon et al., 2012) ou insuffisance cardiaque (Oreopoulos et al., 2008)), l'obésité protègerait d'un décès précoce. Des arguments biologiques ont été mis en avant pour expliquer ce paradoxe, puis des arguments fondés sur la théorie des DAGs (Lajous et al., 2014) : puisque les maladies chroniques sont sur le chemin causal entre l'obésité et la mortalité, conditionner sur une population atteinte de maladies chroniques peut conduire à créer un biais. Nous retrouvons ici la thématique de l'identifiabilité des effets causaux en présence de sélection préférentielle d'individus (ici, les patients atteints d'une maladie chronique). La différence avec les analyses en responsabilité réside dans le fait que le conditionnement se réalise ici sur un facteur intermédiaire ou médiateur (et non un descendant de l'événement Y) de la relation entre X et Y. D'autre part, on se pose davantage la question de la validité interne de la mesure estimée que celle de sa validité externe.

Alors que l'élucidation du paradoxe de l'obésité semblait close, Sperrin et al. (2016) rouvrent le débat en affirmant, à travers des considérations causales, que le biais de collision ne peut expliquer à lui seul le paradoxe de l'obésité. Cependant, leur conclusion se fonde en particulier

sur une hypothèse qui n'est pas validée dans leur modèle causal, hypothèse similaire au cas où l'inclusion dépend de la sévérité de l'accident dans les analyses en responsabilité.
Dans ce chapitre, nous ferons état de considérations théoriques sur le mécanisme causal et les quantités causales d'intérêt à l'oeuvre dans l'étude de l'effet de l'obésité sur le risque de décès précoce parmi les patients atteints d'une maladie chronique. Nous fournirons ensuite une illustration numérique invalidant les conclusions de Sperrin et al. (2016), et montrant que le biais de collision peut expliquer à lui seul du paradoxe de l'obésité.

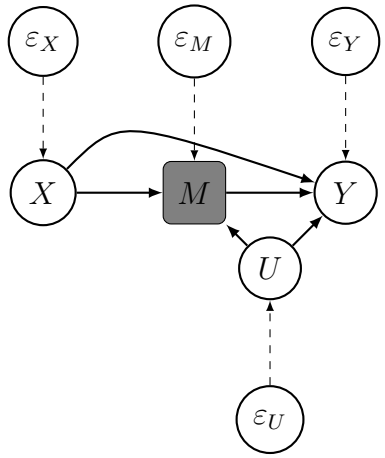
4.1 Considérations théoriques

4.1.1 Présentation du modèle causal

Le paradoxe de l'obésité a émergé suite à l'étude du lien entre l'obésité et la mortalité précoce parmi des patients atteints d'une maladie chronique (diabète ou insuffisance cardiaque). Plusieurs études observationnelles suggèrent une association négative entre l'obésité et le risque de décès prématuré parmi ces patients, concluant ainsi que l'obésité protègerait d'une mortalité précoce parmi les patients atteints d'une maladie chronique. Or, l'obésité est connue pour être un problème de santé publique majeur. Elle est récemment devenue la première cause de décès prématurés aux Etats-Unis (Cleveland Clinic, 22 avril 2017). C'est la conclusion d'une équipe de l'Ecole de médecine de New-York (Etats-Unis), qui a présenté ses travaux au Congrès annuel de la Society of General Internal Medicine, à Washington (D.C., Etats-Unis) du 19 au 22 avril 2017. Dans ce contexte, conclure à une association négative entre obésité et décès prématuré apparaît paradoxal.
Pour élucider ce paradoxe, des arguments issus de l'inférence causale ont été proposés (Lajous et al., 2014) : la maladie chronique étant sur le chemin causal entre l'obésité et le décès prématuré, elle agit comme un médiateur. Or, le conditionnement sur un médiateur crée un biais, dit de collision. Ainsi, l'effet protecteur observé de l'obésité sur le risque de décès prématuré parmi les patients atteints de maladie chronique 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 l'ampleur du biais de collision induit par le conditionnement sur le médiateur, et selon leurs analyses, il est peu probable que le biais de collision explique à lui seul le paradoxe de l'obésité. Néanmoins, leur raisonnement est entaché d'une erreur qui vient remettre en question leur conclusion.

Sperrin et al. (2016) fondent leur conclusion sur le modèle causal décrit par la figure 4.1. Les quatre variables X, M, Y et U, composant l'ensemble $\boldsymbol{V}$, sont considérées comme binaires, et représentent respectivement l'obésité, le diabète, le décès prématuré et un facteur de confusion potentiellement inobservé. Les variables $\varepsilon_{\boldsymbol{V}}$, pour $\boldsymbol{V} \in (X, Y, M, U)$, correspondent aux perturbations associées à chacune des variables X, M, Y et U. Ces perturbations sont supposées mutuellement indépendantes.

FIGURE 4.1 – Le modèle causal considéré par Sperrin et al. (2016)



X obésité, Y décès, M maladie chronique (diabète ou insuffisance cardiaque), U facteurs non observés, (ε_X, ε_M, ε_Y, ε_U) perturbations.

A ce DAG est associé un système d'équations structurelles. Chaque variable du DAG étant décrite par une fonction autonome invoquant les parents de cette variable, et la variable exogène ε. Le système d'équations structurelles reliée au DAG de la figure 4.1 correspond au système suivant :

$$\begin{cases} X &= f_X(\varepsilon_X) \\ U &= f_U(\varepsilon_U) \\ M &= f_M(X, U, \varepsilon_M) \\ Y &= f_Y(X, M, U, \varepsilon_Y), \end{cases}$$

où, f_X, f_U, f_M et f_Y sont des fonctions autonomes qui seront spécifiées dans la section 4.2.1.

Comme dans les chapitres précédents, ce système d'équations structurelles est particulièrement utile pour décrire les variables que l'on aurait observé en fixant certaines variables à une valeur fixe. En particulier, dans le monde contrefactuel Ω^x, ce que l'on aurait observé après l'intervention $do(X = x)$ pour $x \in (0, 1)$ (Pearl, 2000, 2009) est décrit par le système suivant :

$$\begin{cases} U &= f_U(\varepsilon_U) \\ M_x &= f_M(x, U, \varepsilon_M) \\ Y_x &= f_Y(x, M_x, U, \varepsilon_Y). \end{cases}$$

A travers ces équations, l'hypothèse de cohérence est bien vérifiée puisque si $X = x$, alors $M = f_M(x, U, \varepsilon_M) = M_x$ et $Y = f_Y(x, M, U, \varepsilon_Y) = f_Y(x, M_x, U, \varepsilon_Y) = Y_x$.

Pour la suite de ce chapitre, il est également utile d'introduire les variables que l'on aurait observés dans le monde contrefactuel $\Omega^{x,m}$ après la double intervention $do(X = x, M = m)$ (cf.

section 1.3) :

$$\begin{cases} U & = f_U(\varepsilon_U) \\ Y_{x,m} & = f_Y(x, m, U, \varepsilon_Y). \end{cases}$$

Cette définition implique que $Y_x = Y_{x,M_x}$. Ici, Y_{x,M_x} correspond à l'issue contrefactuelle que l'on aurait observé après les interventions $do(X = x, M = M_x)$. Ces interventions ont lieu dans le monde contrefactuel Ω^{x,M_x}, qui est précisément égal au monde contrefactuel Ω^x.

Remarque :
Notons que l'application de l'inférence causale dans le cas de l'obésité est délicate (Hernán and Taubman, 2008). Dans le cas d'études observationnelles mesurant l'effet de l'obésité sur la mortalité, l'hypothèse de cohérence n'est en effet pas vérifiée. D'une part, car la plupart des auteurs ne définissent pas quelles interventions aboutissant à la réduction de l'indice de masse corporelle sont comparées. Et d'autre part, car plusieurs méthodes de réduction de l'indice de masse corporelle peuvent exister (régime alimentaire, pratique sportive, mais également famine, dépression, tabac...) et conduire à la même réduction de l'indice de masse corporelle sans pour autant conduire au même risque contrefactuel de décès prématuré. Dans les études observationnelles, nous ne connaissons pas la ou les raisons pour lesquelles les individus non obèses ne le sont pas au moment de leur décès, et par conséquent l'issue contrefactuelle correspondante devient un concept vague. Le non-respect de l'hypothèse de cohérence affecte également les deux autres conditions nécessaires à l'identifiabilité des effets causaux : l'hypothèse d'ignorabilité (ou échangeabilité) et la positivité. Néanmoins, dans la suite de ce chapitre, nous supposerons qu'une intervention unique et bien définie existe et utiliserons les principes de l'inférence causale comme habituellement.

Afin d'étudier le biais entre la mesure d'association et l'effet causal, Sperrin et al. (2016) considèrent ces mesures sur une échelle additive, l'échelle de l'excès de risque. La mesure d'association entre X et Y conditionnée sur $M = 1$ est définie de la façon suivante :

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

et est comparée à la mesure causale de X sur Y conditionnée sur $M = 1$:

$$\Delta_{CE} = \mathbb{E}[Y_1|M = 1] - \mathbb{E}[Y_0|M = 1].$$

Le paradoxe de l'obésité prend son sens lorsque $\Delta_{AS} \leq 0$. Mais pour expliquer l'existence de ce paradoxe, Sperrin et al. (2016) proposent de comparer Δ_{AS} à Δ_{CE}, c'est-à-dire d'étudier le biais $\Delta_{CE} - \Delta_{AS}$.

4.1.2 Différence entre la mesure d'association et l'effet défini par Sperrin (2016)

Il s'agit donc de comparer Δ_{AS} à Δ_{CE}. D'une part, en présence du facteur de confusion U, Δ_{AS} est égal à

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

D'autre part, Δ_{CE} est égal à

$$\begin{aligned}\Delta_{CE} &= \mathbb{E}[Y_1|M=1] - \mathbb{E}[Y_0|M=1] \\ &= \sum_u \{\mathbb{E}[Y_1|M=1,U=u] - \mathbb{E}[Y_0|M=1,U=u]\}\mathbb{P}(U=u|M=1)\end{aligned}$$

Sperrin et al. (2016) tentent alors d'établir une expression de Δ_{CE} qui ne dépendent que des variables observables, et non des variables contrefactuelles. Ils établissent alors, de manière erronée, que Δ_{CE} vaut Δ_{Sp} où

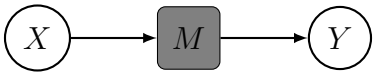
$$\Delta_{Sp} = \sum_u \{\mathbb{E}[Y|X=1,M=1,U=u] - \mathbb{E}[Y|X=0,M=1,U=u]\}\mathbb{P}(U=u|M=1)$$

Par conséquent, selon Sperrin et al. (2016), la différence entre Δ_{AS} à Δ_{CE} viendrait simplement de la différence entre $\mathbb{P}(U=u|X=x,M=1)$ et $\mathbb{P}(U=u|M=1)$.

Or l'égalité $\Delta_{CE} = \Delta_{Sp}$ est garantie si $\mathbb{E}[Y_x|M=1,U=u] = E[Y|X=x,M=1,U=u]$ pour $x \in (0,1)$. Et cette dernière égalité est vérifiée si $Y_x \perp\!\!\!\perp X|(M,U)$. On retrouve ici la même problématique que dans la section 2.2.4 du chapitre 2. Il s'agit donc de déterminer si cette égalité est vérifiée en utilisant par exemple le critère back-door (Pearl, 2000). Or, M étant sur le chemin causal entre X et Y, l'ensemble (M,U) ne satisfait pas le critère back-door. M étant un collider sur le chemin $X \longrightarrow M \longleftarrow U \longrightarrow Y$, le conditionnement sur M génère le biais de collision. Ainsi, $Y_x \not\perp\!\!\!\perp X|(M,U)$, et $E[Y_x|M=1,U=u] \neq E[Y|X=x,M=1,U=u]$.

Deux preuves alternatives, sur un modèle causal plus simple, peuvent être avancées pour montrer que $Y_x \not\perp\!\!\!\perp X|(M,U)$. Considérons le DAG de la figure 4.2, qui correspond au DAG considéré par Sperrin et al. (2016) sans facteur de confusion ni effet direct entre X et Y. Si la condition $Y_x \perp\!\!\!\perp X|M$ n'est pas garantie dans ce modèle simplifié alors elle ne peut pas l'être dans le modèle plus général qui inclut la figure 4.2 comme cas particulier.

FIGURE 4.2 – DAG simplifié du DAG considéré par Sperrin et al., où il n'y a pas de facteur de confusion ni d'effet direct entre X et Y.



Dans ce modèle causal simplifié où X n'est pas un parent de Y, et où il n'y a pas de facteur de confusion U, le système d'équations structurelles dans le monde réel devient :

$$\begin{cases} X &= f_X(\varepsilon_X) \\ M &= f_M(X, \varepsilon_M) \\ Y &= f_Y(M, \varepsilon_Y). \end{cases}$$

Et les variables contrefactuelles Y_x, M_x et $Y_{x,m}$ sont définies de la façon suivante :

$$\begin{cases} M_x &= f_M(x, \varepsilon_M) \\ Y_x &= f_Y(M_x, \varepsilon_Y) \\ Y_{x,m} &= f_Y(m, \varepsilon_Y). \end{cases}$$

Premièrement, une preuve par l'absurde montre que $Y_x \not\perp\!\!\!\perp X|M$ sous le modèle causal simplifié. D'après la structure du DAG de la figure 4.2, deux indépendances sont vérifiées : $Y_x \perp\!\!\!\perp X$ et $Y \perp\!\!\!\perp X|M$. En effet, d'une part, l'ensemble vide vérifie le critère back-door et donc $Y_x \perp\!\!\!\perp X$. Selon cette hypothèse, $\mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1) = \mathbb{P}(Y = 1|X = 1) - \mathbb{P}(Y = 1|X = 0)$, et est généralement non nul. D'autre part, M bloque le chemin entre Y et X donc $Y \perp\!\!\!\perp X|M$. Si l'on considère par erreur, que de plus, $Y_x \perp\!\!\!\perp X|M$, nous aboutissons aux égalités suivantes :

$$\begin{aligned} \mathbb{P}(Y_x = 1) &= \sum_m \mathbb{P}(Y_x = 1|M = m)\mathbb{P}(M = m) \\ &= \sum_m \mathbb{P}(Y_x = 1|M = m, X = x)\mathbb{P}(M = m) \quad \text{si } Y_x \perp\!\!\!\perp X|M \\ &= \sum_m \mathbb{P}(Y = 1|M = m, X = x)\mathbb{P}(M = m) \quad \text{par coherence} \\ &= \sum_m \mathbb{P}(Y = 1|M = m)\mathbb{P}(M = m), \quad \text{car } Y \perp\!\!\!\perp X|M \\ &= \mathbb{P}(Y = 1). \end{aligned}$$

Considérer par erreur que $Y_x \perp\!\!\!\perp X|M$, correspond à conclure que $\mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1) = \mathbb{P}(Y = 1) - \mathbb{P}(Y = 1) = 0$, ce qui est généralement faux. L'obésité a généralement un effet non nul sur le risque de décès prématuré, à moins que l'effet de l'obésité sur le risque de diabète soit nul, ou que le diabète n'ait pas d'impact sur le risque de décès.

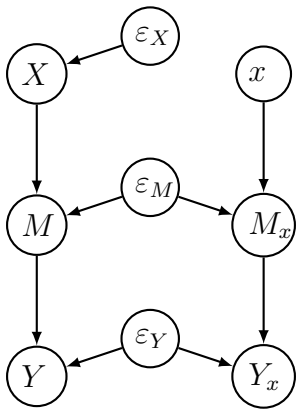
Deuxièmement, la méthode Twin Network (Pearl, 2000, Balke and Pearl, 2011) permet de montrer

que $Y_x \not\perp\!\!\!\perp X|M$ et que $Y_x \perp\!\!\!\perp X|M_x$. Cette méthode permet de représenter sur un même graphique le monde réel et le monde contrefactuel, relié par les variables exogènes ε_Y, ε_X, ε_M et ε_U, et de tester, comme l'approche SWIG, des indépendances entre variables contrefactuelles. La figure 4.3 présente le Twin network correspondant modèle simplifié 4.2. De cette représentation, il en découle $X \perp\!\!\!\perp (Y_x, M_x)$, et donc également que $Y_x \perp\!\!\!\perp X|M_x$. Dans ce cas,

$$\begin{aligned}\mathbb{P}(Y=1|X=x,M=1) &= \mathbb{P}(Y_x=1|X=x,M=1) \quad \text{par coherence}\\ &= \mathbb{P}(Y_x=1|M_x=1) \quad \text{puisque } Y_x \perp\!\!\!\perp X|M_x.\end{aligned}$$

$\mathbb{P}(Y=1|X=x,M=1)$ n'est donc pas égale à $\mathbb{P}(Y_x=1|M=1)$. Les populations $\{M_0=1\}$, $\{M_1=1\}$ et $\{M=1\}$ sont généralement différentes, et donc les probabilités conditionnelles le sont aussi. Par exemple, $\{M_0=1\}$ correspond aux individus qui auraient souffert du diabète dans le monde contrefactuel Ω_0, alors que la population $\{M=1\}$ correspond aux individus souffrant de diabètes dans le monde réel, qu'ils soient obèses ou non. On retrouve ici la question de la validité interne de la mesure d'association estimée (cf. section 2.2.4). La différence entre les populations $\{M_0=1\}$, $\{M_1=1\}$ et $\{M=1\}$ est davantage explicitée dans la section 4.2.2 de ce chapitre.

FIGURE 4.3 – Twin network associé au modèle simplifié 4.2.



En conclusion, $\Delta_{CE} \neq \Delta_{Sp}$ car $E[Y_x|M=1,U=u] \neq E[Y|X=x,M=1,U=u]$ pour $x \in (0,1)$, étant donné que $Y_x \perp\!\!\!\perp X|(M,U)$ n'est pas vérifiée. La différence entre Δ_{AS} et Δ_{CE} ne vient donc pas seulement de la différence entre $\mathbb{P}(U=u|X=x,M=1)$ et $\mathbb{P}(U=u|M=1)$, mais également de celle entre $\mathbb{E}[Y_x|M=1,U=u] \neq \mathbb{E}[Y|X=x,M=1,U=u]$, engendrée par le biais de collision. Par conséquent, lorsque Sperrin et al. (2016) tentent de quantifier le biais entre Δ_{AS} et l'expression erronée de Δ_{CE} dû au biais de collision, ils négligent finalement le biais de collision. Suite à notre lettre (Viallon and Dufournet, 2017b), Sperrin et al. (2017) ont reconnu leur erreur dans le calcul de Δ_{CE}. Ils ont par ailleurs établi que Δ_{Sp} était en fait égal à un effet direct contrôlé, conditionné sur $M=1$, et qui est peut être plus pertinent à considérer (cf. section suivante).

Par ailleurs, dans le DAG considéré par Sperrin et al. (2016), Δ_{CE} n'est en fait pas identifiable, c'est-à-dire exprimable en fonction des distributions des variables observables X, M, U et Y sans hypothèse supplémentaire sur le modèle causal. Dans la section 4.2, nous considérons un modèle génératif cohérent avec le modèle causal considéré par Sperrin et al. (2016). En spécifiant les fonctions f_Y, f_M, f_X et f_U et la distributions des perturbations ε_Y, ε_X, ε_M et ε_U, il est possible d'obtenir une formule analytique de Δ_{CE} (cf. section 4.2.3), puis d'évaluer le réel biais entre Δ_{CE} et Δ_{AS} (cf. section 4.2.4).

4.1.3 Autres quantités d'intérêt : des effets directs contrôlés

La pertinence d'étudier Δ_{CE} dans le contexte présent où M agit comme médiateur entre X et Y est également discutable. D'autres quantités causales comme des effets directs contrôlés peuvent être plus adaptés dans cette situation

Comme défini dans la section 4.1.1, notons M_x la variable contrefactuelle sur le diabète que l'on aurait observée dans le monde contrefactuel Ω^x. De plus, Y_{x,M_x} correspond à l'issue contrefactuelle que l'on aurait observée après les interventions $do(X = x, M = M_x)$. Dès lors, $Y_x = Y_{x,M_x}$, et $\mathbb{P}(Y_x = 1|M = 1) = \mathbb{P}(Y_{x,M_x} = 1|M = 1)$. En particulier, si l'on considère une intervention éliminant l'obésité, $\mathbb{P}(Y_0 = 1|M = 1) = \mathbb{P}(Y_{0,M_0} = 1|M = 1)$ est le risque de mortalité prématuré dans le monde contrefactuel Ω^0, chez les individus souffrant de diabète dans le monde réel. Or, $M = 1 \nRightarrow M_0 = 1$. Un patient souffrant de diabète dans le monde réel n'aurait pas forcément contracté cette maladie dans le monde contrefactuel Ω^0. Par conséquent, observer Δ_{CE} positif peut par exemple être expliqué par le fait qu'une certaine proportion d'individus souffrant d'une maladie chronique dans le monde réel, n'aurait pas souffert de cette maladie chronique dans le monde contrefactuel Ω^0, et aurait par conséquent vécu plus longtemps dans Ω^0 que dans Ω^1. Δ_{CE} reste malgré son expression, un effet total, et peut être décomposer en effet direct et indirect (cf. chapitre 1 section 1.3.2).

$$
\begin{aligned}
\Delta_{CE} &= \mathbb{E}[Y_1|M=1] - \mathbb{E}[Y_0|M=1] \\
&= \mathbb{E}[Y_{1,M_1}|M=1] - \mathbb{E}[Y_{0,M_0}|M=1] \\
&= \mathbb{E}[Y_{1,M_1}|M=1] - \mathbb{E}[Y_{1,M_0}|M=1] + \mathbb{E}[Y_{1,M_0}|M=1] - \mathbb{E}[Y_{0,M_0}|M=1] \\
&= \Delta_{NIE|M=1}(1) + \Delta_{NDE|M=1}(0) \\
&= \mathbb{E}[Y_{1,M_1}|M=1] - \mathbb{E}[Y_{0,M_1}|M=1] + \mathbb{E}[Y_{0,M_1}|M=1] - \mathbb{E}[Y_{0,M_0}|M=1] \\
&= \Delta_{NDE|M=1}(1) + \Delta_{NIE|M=1}(0)
\end{aligned}
$$

En d'autres termes, même si Δ_{CE} est conditionné sur $M = 1$, Δ_{CE} représente un effet total, et non un effet direct. C'est un effet total de X sur Y au sein d'un sous groupe de la population défini par une variable observable dans le monde réel, à l'instar de l'effet causal chez les traités $\mathbb{P}(Y_1 - Y_0|X = 1)$.
En présence d'un médiateur M, la littérature sur l'analyse causale en médiation (Pearl, 2000, Imai et al., 2010, VanderWeele and Vansteelandt, 2010) met à disposition d'autres définitions de quantités causales à étudier. En particulier, dans notre cas où il s'agit d'étudier l'effet de X

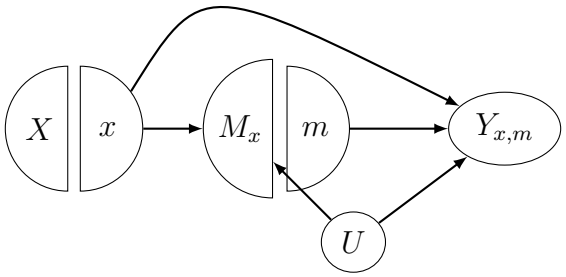
sur Y chez les individus pour lesquels $M = 1$, les effets directs contrôlés semblent plus adaptés. Cet effet défini l'effet causal d'une variable sur un événement pour une valeur fixe de M. Dans le cadre de l'étude des effets de l'obésité chez les malades chroniques, nous nous intéresserons au cas où M est donc fixé à 1. Trois quantités causales seront par la suite considérées :

$$\begin{aligned}\Delta_{CDE} &= \mathbb{P}(Y_{1,1} = 1) - \mathbb{P}(Y_{0,1} = 1)\\ \Delta_{CDE|M=1} &= \mathbb{P}(Y_{1,1} = 1|M = 1) - \mathbb{P}(Y_{0,1} = 1|M = 1)\\ \Delta_{CDE|X=1,M=1} &= \mathbb{P}(Y_{1,1} = 1|X = 1, M = 1) - \mathbb{P}(Y_{0,1} = 1|X = 1, M = 1).\end{aligned}$$

La première mesure correspond à un effet direct contrôlé de X sur Y, en fixant $M = 1$. Cet effet causal compare le risque de mortalité prématurée dans les mondes contrefactuels $\Omega^{1,1}$ et $\Omega^{0,1}$, où ont eu lieu respectivement les doubles interventions $do(X = 1, M = 1)$ et $do(X = 0, M = 1)$. Dans ces deux mondes contrefactuels, tous les individus souffrent de diabète, mais tous sont obèses dans le monde contrefactuel $\Omega^{1,1}$, alors qu'aucun ne l'est dans $\Omega^{0,1}$. Les deux autres mesures sont des versions conditionnelles de Δ_{CDE}. Plus précisément, $\Delta_{CDE|M=1}$ correspond à l'effet direct contrôlé Δ_{CDE} chez les patients souffrant de diabète dans le monde réel. $\Delta_{CDE|X=1,M=1}$ se focalise sur les patients souffrant à la fois de l'obésité et de diabète dans le monde réel.

Sous le modèle causal correspondant à la figure 4.1, Δ_{CDE} et $\Delta_{CDE|X=1,M=1}$ sont identifiables, exprimables à partir des distributions des variables observables X, M, Y, U. En effet, nous avons $Y_{x,m} \perp\!\!\!\perp X|U$ (VanderWeele and Vansteelandt, 2010), car il n'existe pas de facteurs de confusion inobservés entre X et Y. De plus, $Y_{x,m} \perp\!\!\!\perp M|(X, U)$, car il n'existe pas de facteurs de confusion inobservés entre M et Y (à l'exception éventuelle de U) : cf. figure 4.4.

FIGURE 4.4 – SWIT résultant de l'intervention $do(X = x), do(M = m)$



Par conséquent, les probabilités contrefactuelles $\mathbb{P}(Y_{x,m} = 1)$ pour $(x, m) \in (0, 1)$ sont exprimables à partir des probabilités observables :

$$\begin{aligned}
\mathbb{P}(Y_{x,m}=1) &= \sum_u \mathbb{P}(Y_{x,m}=1|U=u)\mathbb{P}(U=u) \\
&= \sum_u \mathbb{P}(Y_{x,m}=1|X=x,U=u)\mathbb{P}(U=u) \quad \text{car } Y_{x,m} \perp\!\!\!\perp X|U \\
&= \sum_u \mathbb{P}(Y_{x,m}=1|X=x,M=m,U=u)\mathbb{P}(U=u) \quad \text{car } Y_{x,m} \perp\!\!\!\perp M|(X,U) \\
&= \sum_u \mathbb{P}(Y=1|X=x,M=m,U=u)\mathbb{P}(U=u) \quad \text{par cohérence.}
\end{aligned}$$

Et Δ_{CDE} correspond à l'expression suivante :

$$\Delta_{CDE} = \sum_u \{\mathbb{P}(Y=1|X=1,M=1,U=u) - \mathbb{P}(Y=1|X=1,M=1,U=u)\}P(U=u).$$

Considérons maintenant $\Delta_{CDE|X=1,M=1}$, et l'identifiabilité des probabilités $\mathbb{P}(Y_{x,m}=1|X=x^*,M=m^*)$:

$$\begin{aligned}
&\mathbb{P}(Y_{x,m}=1|X=x^*,M=m^*) \\
&= \sum_u \mathbb{P}(Y_{x,m}=1|X=x^*,M=m^*,U=u)\mathbb{P}(U=u|X=x^*,M=m^*) \\
&= \sum_u \mathbb{P}(Y_{x,m}=1|X=x^*,U=u)\mathbb{P}(U=u|X=x^*,M=m^*) \quad \text{car } Y_{x,m} \perp\!\!\!\perp M|(X,U) \\
&= \sum_u \mathbb{P}(Y_{x,m}=1|X=x,U=u)\mathbb{P}(U=u|X=x^*,M=m^*) \quad \text{car } Y_{x,m} \perp\!\!\!\perp X|U \\
&= \sum_u \mathbb{P}(Y_{x,m}=1|X=x,M=m,U=u)\mathbb{P}(U=u|X=x^*,M=m^*) \quad \text{car } Y_{x,m} \perp\!\!\!\perp M|(X,U) \\
&= \sum_u \mathbb{P}(Y=1|X=x,M=m,U=u)\mathbb{P}(U=u|X=x^*,M=m^*) \quad \text{par cohérence.}
\end{aligned}$$

Et $\Delta_{CDE|X=1,M=1}$ correspond à l'expression suivante :

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

La différence essentielle entre Δ_{AS} et Δ_{CE} vient de celle entre $\mathbb{E}[Y_a|M=1,U=u]$ et $\mathbb{E}[Y|X=x,M=1,U=u]$, engendrée par le biais de collision. En revanche, la différence entre les quatres mesures Δ_{AS}, Δ_{Sp}, Δ_{CDE} et $\Delta_{CDE|X=1,M=1}$ est uniquement due à la distribution du facteur de confusion U utilisée afin de marginaliser sur u les quantités $\mathbb{P}(Y=1|X=x,M=1,U=u)$ pour $x \in (0,1)$. C'est $\mathbb{P}(U=u|X=x,M=1)$ qui est utilisée dans le cas de Δ_{AS}, $\mathbb{P}(U=u|M=1)$ pour Δ_{Sp}, $\mathbb{P}(U=u)$ pour Δ_{CDE}, et $\mathbb{P}(U=u|X=1,M=1)$ pour $\Delta_{CDE|X=1,M=1}$. Cela signifie également qu'en l'absence de facteur de confusion U, $\Delta_{AS} = \Delta_{Sp} = \Delta_{CDE} = \Delta_{CDE|X=1,M=1}$. Et par conséquent le biais entre ces trois quantités peut s'interpréter comme un biais de confusion. Cela confirme également qu'en ajustant sur un médiateur, on obtient un effet direct, et non un effet total comme le prétend Sperrin et al. (2016).

4.2 Illustrations numériques

Nous fournissons dans cette partie des exemples numériques illustrant les écarts entre les différentes quantités introduites. Nous présentons tout d'abord le système génératif de données considéré. Grâce à ce système, nous illustrons la différence entre les quantités $\mathbb{P}(Y_x = 1|M_x = 1)$ et $\mathbb{P}(Y_x = 1|M = 1)$ puis nous dérivons une formule analytique de la véritable quantité causale Δ_{CE}. Enfin, nous présentons des résultats selon lesquels l'observation de $\Delta_{AS} < 0$ peut être reliée à des quantités causales positives : $\Delta_{CE} > 0$, $\Delta_{CDE} > 0$, et $\Delta_{CDE|M=1} > 0$.

4.2.1 Mécanisme de génération des données

Nous décrivons dans cette section le modèle génératif que nous avons considéré, cohérent avec la figure 4.1, et qui correspond également à une spécification du modèle considéré par Sperrin et al. (2016). Plus précisément, notre système de génération des données s'obtient en spécifiant les fonctions f_X, f_U, f_M et f_Y et la distributions des perturbations $\varepsilon_X, \varepsilon_U, \varepsilon_M$ et ε_Y du système d'équations structurelles.

Soit $\mathbb{I}[\cdot]$ la fonction indicatrice. Les quatre perturbations, indépendantes entre elles, $\varepsilon_X, \varepsilon_U, \varepsilon_M$ et ε_Y ont générées selon une loi uniforme $[0, 1]$. Pour $(p_X, p_U) \in (0, 1)^2$, nous définissons $X = \mathbb{I}[\varepsilon_X \leq p_X]$ et $U = \mathbb{I}[\varepsilon_U \leq p_U]$ de telle sorte que $X \sim B(p_X)$ et $U \sim B(p_U)$ soient deux variables indépendantes suivant une loi de Bernoulli. Comme Sperrin et al. (2016), nous considérons le cas particulier où $p_X = p_U = 0.5$. Nous introduisons ensuite la fonction sigmoïde $\mathrm{h}(x) = (1 + \exp(-x))^{-1}$, et posons, pour tout $(x, u, m) \in \{0, 1\}^3$ et des paramètres réels $\alpha_0, \alpha_X, \alpha_U, \alpha_{XU}, \beta_0, \beta_X, \beta_U, \beta_M, \beta_{XM}, \beta_{XU}, \beta_{UM}$ et β_{XUM},

$$\begin{aligned} p_M(x, u) &= \mathrm{h}(\alpha_0 + \alpha_X x + \alpha_U u + \alpha_{XU} xu) \\ p_Y(x, m, u) &= \mathrm{h}(\beta_0 + \beta_X x + \beta_U u + \beta_M m + \beta_{XU} xu + \\ &\qquad \beta_{XM} xm + \beta_{UM} um + \beta_{XUM} xum). \end{aligned}$$

Finalement, nous définissons les variables M et Y par

$$\begin{aligned} M &= f_M(X, U, \varepsilon_M) = \mathbb{I}[\varepsilon_M \leq p_M(X, U)], \\ Y &= f_Y(X, M, U, \varepsilon_Y) = \mathbb{I}[\varepsilon_Y \leq p_Y(X, M, U)]. \end{aligned}$$

De la même manière que Sperrin et al. (2016), nous définissons les constantes suivantes :

$$\begin{aligned} \alpha_0 &= -\frac{1}{2}\left(\alpha_X + \alpha_U + \frac{1}{2}\alpha_{XU}\right) \\ \beta_0 &= -\frac{1}{2}\left(\beta_X + \beta_M + \beta_U + \frac{1}{2}(\beta_{XM} + \beta_{AU} + \beta_{UM}) + \frac{1}{4}\beta_{xUM} - \nu\right) \end{aligned}$$

pour $\nu \geq 0$.
Sperrin et al. (2016) fixent $\nu = 0$, ce qui garantit une prévalence de Y autour de 0.5. Ici, les résultats seront également présentés en considérant $\nu = 0$.

Nous calculons ensuite la mesure d'association et les quantités causales sur l'échelle des odds

ratios, plutôt que sur l'échelle additive de l'excès de risque. Nous définissons les quantités correspondantes à Δ_{AS}, Δ_{CE}, Δ_{Sp}, Δ_{CDE}, $\Delta_{CDE|X=1,M=1}$ par OR_{AS}, OR_{CE}, OR_{Sp}, OR_{CDE}, $OR_{CDE|X=1,M=1}$. Plus précisément,

$$
\begin{aligned}
OR_{AS} &= \frac{\mathbb{P}(Y=1|X=1,M=1)/\mathbb{P}(Y=0|X=1,M=1)}{\mathbb{P}(Y=1|X=0,M=1)/\mathbb{P}(Y=0|X=0,M=1)} \\
OR_{CE} &= \frac{\mathbb{P}(Y_1=1|M=1)/\mathbb{P}(Y_1=0|M=1)}{\mathbb{P}(Y_0=1|M=1)/\mathbb{P}(Y_0=0|M=1)} \\
OR_{CDE} &= \frac{\mathbb{P}(Y_{1,1}=1)/\mathbb{P}(Y_{1,1}=0)}{\mathbb{P}(Y_{0,1}=1)/\mathbb{P}(Y_{0,1}=0)} \\
OR_{CDE|X=1,M=1} &= \frac{\mathbb{P}(Y_{1,1}=1|X=1,M=1)/\mathbb{P}(Y_{1,1}=0|X=1,M=1)}{\mathbb{P}(Y_{0,1}=1|X=1,M=1)/\mathbb{P}(Y_{0,1}=0|X=1,M=1)}.
\end{aligned}
$$

Mais avant de présenter les résultats issus de l'étude des biais entre ces différentes écarts, utilisons la description du système de génération des données afin d'illustrer les différences entre les populations $\{M_0=1\}$, $\{M_1=1\}$ et $\{M=1\}$.

4.2.2 Illustration de la différence entre $\mathbb{P}(Y_x=1|M_x=1)$ et $\mathbb{P}(Y_x=1|M=1)$

Dans la section 4.1.2, nous avons montré que dans un DAG plus simple que celui considéré par Sperrin et al. (2016), $Y_x \not\perp\!\!\!\perp X|M$ mais que $X \perp\!\!\!\perp (Y_x, M_x)$. Il en découle donc que $\mathbb{P}(Y=1|M=1,X=x)$, quantité composant les mesures d'associations est formellement égale à $\mathbb{P}(Y_x=1|M_x=1)$. Et cette probabilité est généralement différente de la probabilité $\mathbb{P}(Y_x=1|M=1)$.

Le modèle génératif permet d'illustrer le fait que $\mathbb{P}(Y_x=1|M_x=1) \neq \mathbb{P}(Y_x=1|M=1)$ en caractérisant concrètement les populations $\{M_x=1\}$ et $\{M=1\}$. Considérons notre modèle génératif défini dans la section précédente et fixons $\beta_X=\beta_U=\alpha_U=0$ et $\beta_{XM}=\beta_{XU}=\beta_{XUM}=\beta_{UM}=\alpha_{XU}=\alpha_{UM}=0$, ce qui correspond à l'absence de flèche entre X et Y, c'est-à-dire à l'absence d'effet direct, et à l'absence de facteur de confusion U.
Sous cette configuration,

- $\{M_0=1\}=\{\varepsilon_M \leq \text{expit}(\alpha_0)\}$, de telle sorte que $\mathbb{P}(M_0=1)=\text{expit}(\alpha_0)$, et
- $\{M_1=1\}=\{\varepsilon_M \leq \text{expit}(\alpha_0+\alpha_X)\}$, de telle sorte que $\mathbb{P}(M_1=1)=\text{expit}(\alpha_0+\alpha_X)$

Et pour $\{M=1\}$, nous avons :

$$
\begin{aligned}
\{M=1\} &= \{\varepsilon_M \leq \text{expit}(\alpha_0+\alpha_X X)\} \\
&= (\{\varepsilon_M \leq \text{expit}(\alpha_0+\alpha_X)\} \cap \{\varepsilon_X \leq p_X\}) \cup (\{\varepsilon_M \leq \text{expit}(\alpha_0)\} \cap \{\varepsilon_X > p_X\}),
\end{aligned}
$$

de telle sorte que $\mathbb{P}(M=1)=p_X\text{expit}(\alpha_0+\alpha_X)+(1-p_X)\text{expit}(\alpha_0)$.

Par conséquent, $\{M=1\} \neq \{M_0=1\}$, $\{M=1\} \neq \{M_1=1\}$ et $\{M_0=1\} \neq \{M_1=1\}$ dès lors que $\alpha_X \neq 0$ sous le modèle causal simplifié.

4.2.3 Expression analytique de la quantité causale sous le mécanisme de génération de données considéré

Comme mentionné dans la section 4.1.2, Δ_{CE} n'est pas identifiable sous le modèle considéré par Sperrin et al. (2016), dans le sens où $\mathbb{P}(Y_x = 1|M = 1)$ ne peut s'exprimer en fonction de la distribution des variables observées X, M, U, Y sans hypothèse supplémentaire sur le modèle causal. Une fois spécifiées la forme des fonctions structurelles, ou génératrices, ainsi que la distribution des perturbations, il est toutefois possible d'obtenir une expression analytique de Δ_{CE}, afin d'évaluer le véritable biais entre Δ_{CE} et Δ_{AS}.

$$
\begin{aligned}
&\mathbb{P}(Y_x = 1|M = 1) \\
&= \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \mathbb{P}(Y_x = 1|M = 1, X = x^*, U = u)\mathbb{P}(X = x^*, U = u|M = 1) \\
&= \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \frac{\mathbb{P}(Y_x = 1, M = 1|X = x^*, U = u)}{\mathbb{P}(M = 1|X = x^*, U = u)}\ \mathbb{P}(X = x^*, U = u|M = 1) \\
&= \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \frac{\mathbb{P}(Y_x = 1, M = 1|X = x^*, U = u)}{\mathbb{P}(M = 1|X = x^*, U = u)}\ \frac{\mathbb{P}(X = x^*, U = u, M = 1)}{\mathbb{P}(M = 1)} \\
&= \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \frac{\mathbb{P}(Y_x = 1, M = 1|X = x^*, U = u)}{\mathbb{P}(M = 1|X = x^*, U = u)}\ \frac{\mathbb{P}(M = 1|X = x^*, U = u)\ \mathbb{P}(X = x^*, U = u)}{\mathbb{P}(M = 1)} \\
&= \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \mathbb{P}(Y_x = 1, M = 1|X = x^*, U = u)\ \frac{\mathbb{P}(X = x^*, U = u)}{\mathbb{P}(M = 1)} \\
&= \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \int_0^{p_M(x^*,u)} \mathbb{P}(Y_x = 1|\varepsilon_M = \varepsilon, X = x^*, U = u)\ f_{\varepsilon_M|X=x^*,U=u}\ \frac{\mathbb{P}(X = x^*, U = u)}{\mathbb{P}(M = 1)} d\varepsilon \\
&= \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \int_0^{\mathbb{P}_M(x^*,i_U)} \mathbb{P}(Y_x = 1|\varepsilon_M = \varepsilon, X = x^*, U = u)\ f_{\varepsilon_M|X=x^*,U=i_U}\ \frac{1}{4\mathbb{P}(M = 1)} d\varepsilon \\
&\overset{(*)}{=} \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \int_0^{p_M(x^*,u)} \frac{\mathbb{P}(Y_x = 1|\varepsilon_M = \varepsilon, X = x^*, U = u)}{4\mathbb{P}(M = 1)} d\varepsilon \\
&= \sum_{\substack{x^*\in\{0,1\}\\ u\in\{0,1\}}} \int_0^{p_M(x^*,u)} \frac{\mathbb{P}(\varepsilon_Y \leq p_Y(x, M_x, U)|\varepsilon_M = \varepsilon, X = x^*, U = u)}{4\mathbb{P}(M = 1)} d\varepsilon
\end{aligned}
$$

où $(*)$ est établi en utilisant le fait que la densité conditionnelle de ε_M sachant $(X = x^*, U = u)$, $f_{\varepsilon_M|X=x^*,U=u}$, est uniformément égale à 1 sur l'intervalle $[0, 1]$.
On pose alors $x \wedge y = \min(x, y)$. En utilisant le fait que (i) $M_x = \mathbb{I}[\varepsilon_M \leq p_M(x, U)]$, et (ii)

$\mathbb{P}(\varepsilon_Y \leq \rho) = \rho$ pour $\rho \in [0,1]$, alors

$$\begin{aligned}
&\mathbb{P}(Y_x = 1|M=1) \\
&= \sum_{\substack{x^* \in \{0,1\} \\ u \in \{0,1\}}} \frac{\int_0^{[p_M(x^*,u) \wedge p_M(x,u)]} \mathbb{P}(\varepsilon_Y \leq p_Y(x,1,u))d\varepsilon + \int_{[p_M(x^*,u) \wedge p_M(x,u)]}^{p_M(x^*,u)} p_Y(x,0,u)d\varepsilon}{4\mathbb{P}(M=1)} \\
&= \sum_{\substack{x^* \in \{0,1\} \\ u \in \{0,1\}}} \frac{p_Y(x,1,u)[p_M(x^*,u) \wedge p_M(x,u)] + p_Y(x,0,u)\{p_M(x^*,u) - [p_M(x^*,u) \wedge p_M(x,u)]\}}{4\mathbb{P}(M=1)}
\end{aligned}$$

Etant donné que $\mathbb{P}(M=1) = \sum_{x^*,u} p_M(x^*,u)/4$, il est alors possible de calculer $\mathbb{P}(Y_x = 1|M = 1)$ et donc Δ_{CE} ou OR_{CE} pour toutes combinaisons de paramètres $\alpha_0, \alpha_X, \alpha_U, \alpha_{XU}, \beta_0, \beta_X, \beta_U, \beta_M, \beta_{XM}, \beta_{XU}, \beta_{UM}$ et β_{XUM}.

4.2.4 Résultats

L'objectif de cette partie est de présenter différentes configurations numériques où un $OR_{AS} < 1$ peut être obtenu conjointement à des odds ratios causaux OR_{CE}, OR_{CDE}, et $OR_{CDE|M=1}$ supérieur à 1.

Dans un premier temps, nous posons $\beta_{XM} = \beta_{XU} = \beta_{UM} = \beta_{XUM} = 0$. Sous cette configuration, la figure 4.5 illustre les différences entre $OR_{AS}, OR_{CE}, OR_{CDE}$ et $OR_{CDE|X=1,M=1}$ pour différentes combinaisons de paramètres.
Les deux premières lignes de la figure correspondent aux résultats présentés par Sperrin et al. (2016). Lorsque $\beta_M = 0$, cas particulier où M n'est pas un médiateur, le biais de collision n'intervient pas, et $OR_{CE} = OR_{CDE} = OR_{CDE|X=1,M=1} = 1$. Que X ait un effet direct sur Y (i.e. $\beta_X \neq 0$) ou non (i.e. $\beta_X = 0$), cette première configuration décrit une situation où $Y_x \perp\!\!\!\perp X|M$, et donc $OR_{Sp} = OR_{CE} = 1$. OR_{AS} est en général différent de 1 et le biais entre OR_{AS} et OR_{CE} est plutôt faible, et se réduit à un biais de confusion.
Lorsque $\beta_M \neq 0$, c'est-à-dire lorsque M est un médiateur, la situation est très différente (excepté si $\alpha_X = 0$), contrairement à ce que Sperrin et al. (2016) prétendent. Dans ce cas, OR_{CE} diffère des autres quantités OR_{AS}, OR_{Sp}, OR_{CDE}, $OR_{CDE|X=1,M=1}$. Le fait que $\beta_M \neq 0$ illustre la différence entre la véritable mesure causale OR_{CE} et l'expression OR_{Sp} étudiée par Sperrin et al. (2016). Le biais de collision est en fait actif puisque $Y_x \not\perp\!\!\!\perp X|M$, et ce quelle que soit la valeur de β_X. D'autre part, OR_{AS} est généralement différent de 1. Si la quantité d'intérêt reste OR_{CE}, le biais entre OR_{AS} et OR_{CE} peut être relativement important dès lors que $\alpha_X \geq 2$. Les deux dernières lignes de la figure 4.5 montrent notamment qu'il est tout à fait possible d'avoir $OR_{AS} \leq 1$ et $OR_{CE} \geq 1$. Il est donc possible d'obtenir une association négative entre l'obésité et la mortalité précoce chez les diabétiques alors que l'odds ratio causal chez ces mêmes patients est positif.
Si l'on choisit en revanche d'étudier les odds ratios causaux OR_{CDE} ou $OR_{CDE|X=1,M=1}$, le biais entre OR_{AS} et l'une des deux quantités est assez faible. L'association observée OR_{AS} chez les patients diabétiques fournit une approximation légèrement biaisée de l'odds ratio direct contrôlé OR_{CDE}, et de sa version conditionnelle $OR_{CDE|X=1,M=1}$ (cf. figure 4.6).

Dans un deuxième temps, nous présentons des résultats en présence de facteurs de confusion et d'interactions et posons $\alpha_X = \alpha_U = \beta_X = \beta_{UM} = 2$, $\beta_U = 3$ et $\beta_{XM} = -2$. Les résultats sont présentés à travers la figure 4.6. Dans chaque cadran, un des quatre paramètres α_{XU}, β_M, β_{XU} et β_{XUM} varie entre -3 et 3, les autres étant fixés à 1. Nous observons que le biais entre OR_{AS} et OR_{CE} peut être très important lorsqu'un de ces quatre paramètres est grand. D'autre part, lorsqu'on s'intéresse au biais entre OR_{AS} et un des odds ratios directs contrôlés, le biais est moins important. Cependant, dans de nombreuses situations, il est possible d'observer un OR_{AS} sensiblement inférieur à 1, et des odds ratios directs contrôlés OR_{CDE} ou $OR_{CDE|X=1,M=1}$ supérieur à 1.

En résumé, nos résultats numériques montrent que dès lors que $\beta_M \neq 0$ et $\alpha_X \neq 0$, ce qui décrit une situation où M est un médiateur, l'odds ratio causal OR_{CE} et la quantité OR_{Sp} étudiée par Sperrin et al. (2016) diffèrent. Dans de nombreuses configurations, il est ainsi possible d'observer $OR_{AS} \leq 1$ et $OR_{CE} \geq 1$. En présence de facteurs de confusion et d'interactions, obtenir une association négative entre X et Y, parmi les patients diabétiques ($OR_{AS} \leq 1$) n'implique pas forcément $OR_{CE} \leq 1$, $OR_{CDE} \leq 1$ ou $OR_{CDE|X=1,M=1} \leq 1$.

FIGURE 4.5 – Les odds ratios causaux et observés sont présentés pour $(\beta_X = 0, \beta_M = 0)$, $(\beta_X = 1, \beta_M = 0)$, $(\beta_X = 0, \beta_M = 1)$ et $(\beta_X = 1, \beta_M = 1)$, dans les situations où $\beta_{XM} = \beta_{XU} = \beta_{UM} = \beta_{XUM} = 0$, et en faisant varier les autres paramètres $\alpha_X, \alpha_U, \beta_U$ et α_{XU}. Dans chaque colonne, sur l'axe des abscisses, un des paramètres varie entre -3 et 3 (colonne de gauche :α_X, colonne suivante : α_U, colonne suivante β_U, et colonne de droite α_{XU}), et les autres paramètres sont fixés à une valeur par défaut (1 pour $\alpha_X, \alpha_U, \beta_U$ et 0 pour α_{XU}).

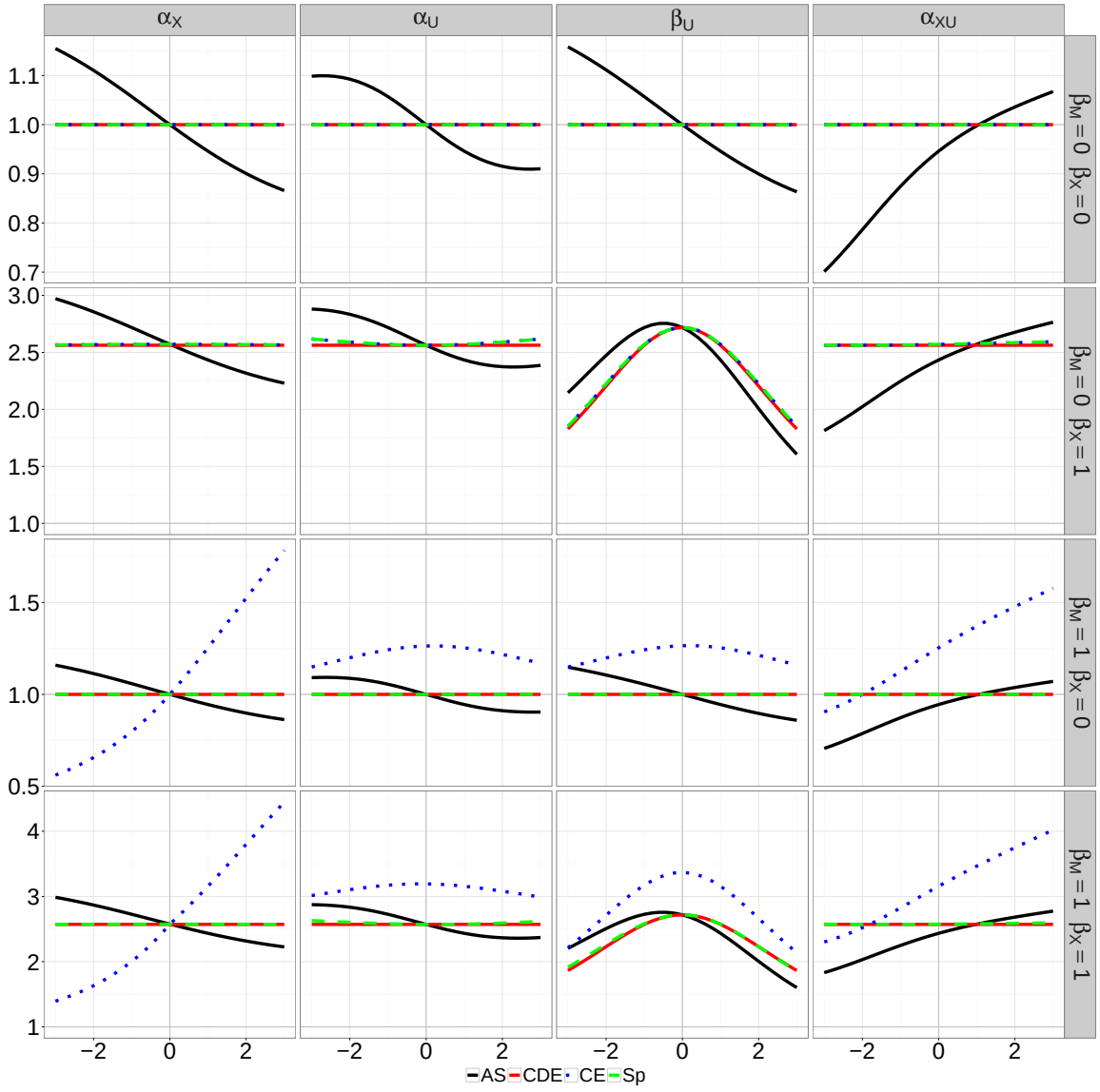
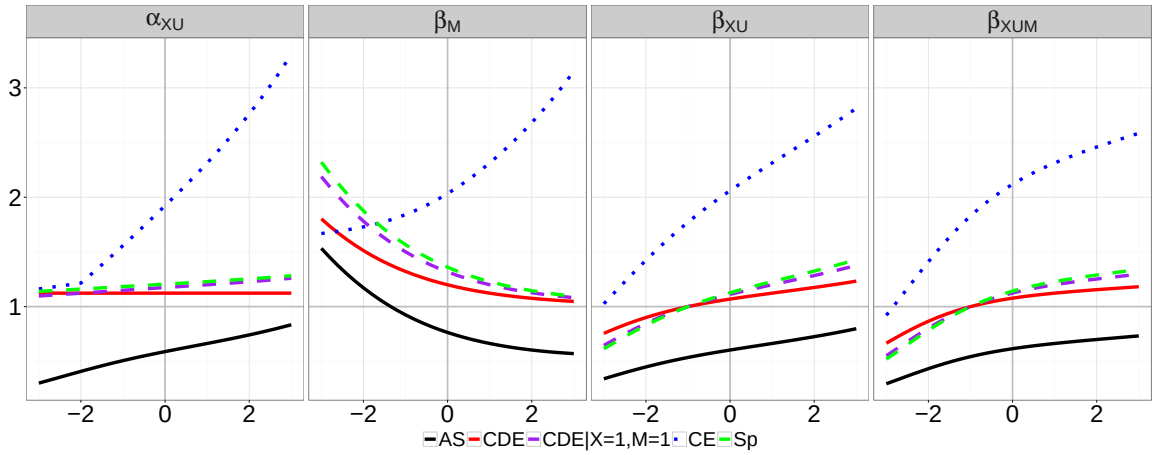


FIGURE 4.6 – Odds ratios causaux et observés dans le cas où $\alpha_X = \alpha_U = 2 = \beta_X = \beta_{UM} = 2$, $\beta_U = 3$ et $\beta_{XM} = -2$, pour des valeurs variables de α_{XU}, β_M, β_{XU} et β_{XUM}. Dans chaque colonne, sur l'axe des abscisses, un des paramètres varie entre -3 et 3 (colonne de gauche :α_{XU}, colonne suivante : β_M, colonne suivante β_{XU}, et colonne de droite β_{XUM}), et les autres paramètres sont fixés à 1 par défaut .



Conclusion

L'obésité est aujourd'hui considérée comme une cause avérée de mortalité prématurée. En utilisant les notations propres à l'inférence causale, cela signifie que l'odds ratios causal,

$$(\mathbb{P}(Y_1 = 1)/\mathbb{P}(Y_1 = 0))/(\mathbb{P}(Y_0 = 1)/\mathbb{P}(Y_0 = 0)),$$

est supérieur à 1. Or, paradoxalement, plusieurs études observationnelles ont reporté des odds ratios inférieurs à 1, en étudiant des individus souffrant de diabète ou d'insuffisance cardiaque (Carnethon et al., 2012, Oreopoulos et al., 2008), suggérant ainsi $OR_{AS} \leq 1$. Ce résultat peut être considéré comme paradoxal si l'association observée parmi les patients atteints de diabète ou d'insuffisance cardiaque OR_{AS} constitue un odds ratio causal parmi ces mêmes patients OR_{CE}. Cependant, OR_{AS} n'est pas un odds ratio causal total de type OR_{CE}, puisque conditionner sur des patients atteints de maladie chronique, i.e. conditionner sur $M = 1$, revient à ajuster sur le médiateur M entre X et Y. Le conditionnement sur le médiateur ruine la condition d'ignorabilité conditionnelle $Y_x \perp\!\!\!\perp X|(M, U)$ nécessaire à l'identifiabilité de l'odds ratio causal OR_{CE}. Contrairement à ce que Sperrin et al. (2016) prétendent, nous montrons par une illustration numérique que ce conditionnement peut être une source de biais importante sous le modèle causal simple considéré par ces mêmes auteurs. Par ailleurs, nous avons montré que d'autres quantités causales pouvaient être plus proches de l'OR estimable OR_{AS}, en particulier l'odds ratio direct contrôlé OR_{CDE} et des versions conditionnelles $OR_{CDE|M=1}$ et $OR_{CDE|X=1,M=1}$. Selon les configurations proposées par Sperrin et al. (2016), mais en considérant M comme médiateur ($\beta_M \neq 0$), la comparaison entre OR_{AS} et les différents odds ratios contrôlés suggère une différence généralement faible. En ajoutant en revanche des termes d'interaction, il est possible d'observer $OR_{AS} \leq 1$ et les odds ratios directs contrôlés supérieurs à 1. Par conséquent, l'estimation de OR_{AS} doit être interprétée avec prudence, et considérée au regard d'une quantité causale pertinente.

On peut ensuite se demander si OR_{CE} et OR_{CDE} sont des quantités pertinentes pour déterminer si la perte de poids est bénéfique en termes d'espérance de vie pour un patient à la fois obèse et atteint d'une maladie chronique(diabète ou insuffisance cardiaque). Dans la lignée de Anker and Haehling (2011) qui se sont posées ces questions chez les patients atteints d'insuffisance cardiaque, on peut se demander si la perte de poids a un effet positif sur la guérison du diabète. A t-elle également un effet positif sur l'espérance de vie chez les diabétiques... ? Le risque de mortalité prématurée chez ces patients est $\mathbb{P}(Y = 1|X = 1, M = 1) = \mathbb{P}(Y_1 = 1|X = 1, M = 1)$, par cohérence. Mais quel serait ce risque après une perte de poids ? S'il s'agit simplement du risque que l'on observerait dans le monde contrefactuel Ω^0 où on aurait fixé l'intervention $do(X = 0)$, alors ce risque correspondrait à $\mathbb{P}(Y_0 = 1|X = 1, M = 1)$. Mais comme $\mathbb{P}(Y_0 = 1|X = 1, M = 1) = \mathbb{P}(Y_{0,M_0} = 1|X = 1, M = 1)$, il n'est pas évident que cette hypothèse implique que ces patients soient guéris du diabète ou d'insuffisance cardiaque après une perte de poids. Sous cette hypothèse, la quantité d'intérêt est $\mathbb{P}(Y = 1|X = 1, M = 1) - \mathbb{P}(Y_0 = 1|X = 1, M = 1) = \mathbb{P}(Y_1 = 1|X = 1, M = 1) - \mathbb{P}(Y_0 = 1|X = 1, M = 1)$. Cette quantité est reliée à Δ_{CE} mais aussi à la fraction attribuable ou à la probabilité d'évitement (« probability of disablement ») (Pearl, 2000, 2009). Si en revanche, la perte de poids ne peut typiquement pas guérir ces patients du diabète ou d'insuffisance cardiaque, alors leur risque après une perte de poids est égal à $\mathbb{P}(Y_{0,1} = 1|X = 1, M = 1)$. Et la quantité causale d'intérêt devient $\mathbb{P}(Y = 1|X = 1, M = 1) - \mathbb{P}(Y_{0,1} = 1|X = 1, M = 1) = \mathbb{P}(Y_{1,1} = 1|X = 1, M = 1) - \mathbb{P}(Y_{0,1} = 1|X = 1, M = 1) = \Delta_{CDE|X=1,M=1}$.

Enfin, rappelons qu'il n'existe pas d'intervention unique et bien défini lorsqu'il s'agit de déterminer l'effet causal de l'obésité sur le décès prématuré. Par conséquent, l'inférence causale est une tâche compliquée dans le cas d'une exposition comme l'obésité (Hernán and Taubman, 2008). Suite à notre lettre, Vansteelandt (2017) ajoute également que l'effet direct contrôlé ne reflète pas un effet d'intérêt clinique. D'après lui, il exprime l'effet qu'une intervention sur l'obésité aurait sur la mortalité, avant l'apparition de la maladie chronique, chez les malades chroniques. Et cet effet a peu d'intérêt pour déterminer l'effet potentiel d'une intervention sur l'obésité sur la mortalité après l'apparition de la maladie. Selon Vansteelandt (2017), le poids avant l'apparition de la maladie est plutôt liée à la génétique, au contexte socio-économique, aux habitudes alimentaires, tandis que le poids après l'apparition de la maladie peut être lié à la maladie elle-même, à des prescriptions médicamenteuses, ou des exercices physiques. Et cette différence fait qu'une intervention sur l'obésité a peu de chance d'avoir le même impact selon les deux cas. Enfin, la perte de poids peut être liée à la maladie chronique, et dans ce cas, nous serions dans un cas de « causalité inversée » (« reverse causation ») (Stokes and Preston, 2015). Or, il n'est pas certain que la perte de poids soit bénéfique chez les patients diabétiques. En général, chez ces patients, celle-ci est signe de mauvaise santé et augmente le risque de mortalité prématurée. Afin de déterminer si la perte de poids est bénéfique en termes de durée de vie chez les patients atteints de diabète ou d'insuffisance cardiaque, un protocole pourrait être envisagé. Dans un premier temps, dans le but de définir les interventions possibles afin de perdre du poids, dans un deuxième temps, dans le but de planifier un essai randomisé.

Discussion générale

Afin d'aider les décideurs publics à réduire le nombre d'accident, il est nécessaire d'identifier les causes intervenant dans la survenue de l'accident, et de mesurer l'effet causal de chacune sur le risque d'accident. Aujourd'hui, les principales causes d'accident sont considérées connues et l'enjeu réside davantage dans la mesure de l'effet de chaque cause, et dans la hiérarchisation des facteurs causaux. L'expérimentation étant impossible pour des raisons éthiques, l'épidémiologiste du risque routier se heurte à plusieurs sources de biais classiques lorsqu'il souhaite estimer une quantité causale : le biais de classement, le biais de confusion et le biais de sélection. C'est sur le biais de sélection que nous avons davantage insisté dans le cadre de cette thèse.
La mesure d'un effet causal en sécurité routière est d'autant plus complexe que les données à disposition concernent uniquement des accidentés. Pour faire face à cette absence de témoins, et donc à cette sélection extrême, il est maintenant classique d'utiliser des analyses en responsabilité, et de comparer les responsables de l'accident aux les non-responsables de l'accident. Les analyses en responsabilité se fondent ainsi sur une hypothèse très forte : les individus identifiés comme non-responsables accidentés sont représentatifs de la population des non-responsables ou des non-impliqués dans un accident. Quelques auteurs et articles se sont questionnés sur la validité et la robustesse des analyses en responsabilité, soit en remettant en question cette hypothèse (Sanghavi, 2013), soit en questionnant son statut d'alternative à une étude cas-témoins (Salmi et al., 2014). Mais faute d'alternatives plus pertinentes, les analyses en responsabilité sont largement utilisées en sécurité routière et rarement remises en question.

L'objectif de cette thèse était donc de déterminer si les données disponibles en sécurité routière permettaient d'obtenir des mesures causales, via les analyses en responsabilité, sans biais de sélection résiduel.

La réponse que nous fournissons dans le cadre de cette thèse est la suivante. L'estimation d'une quantité causale, sans biais de sélection résiduel, via les analyses en responsabilité dépend du respect d'une de ces deux conditions :

- les données disponibles ne dépendent pas de la sévérité de l'accident, et donc de la vitesse
- l'exposition considérée n'a pas d'effet sur la vitesse

Si au moins l'une de ces deux conditions est vérifiée, sous des hypothèses de prévalences faibles, l'odds ratio causal de l'exposition X sur la responsabilité de l'accident R, $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$, est en fait approximable par un autre odds ratio causal identifiable, l'odds ratio causal de X sur la faute de conduite F, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$.
Or, le plus fréquemment, les données disponibles en sécurité routière dépendent de la sévérité de l'accident. En effet, seuls les accidents donnant lieu à un procès-verbal sont récoltés, ce qui correspond aux accidents mortels et aux accidents corporels les plus graves. Ainsi, l'estimation

d'une quantité causale via les analyses en responsabilité est la plupart du temps compromise. Lorsque les données dépendent de la sévérité de l'accident et donc de la vitesse, l'estimation d'une quantité causale sans biais de sélection résiduel est possible si l'exposition X n'a pas d'effet sur la vitesse.
Cela signifie que lorsque l'on souhaite estimer l'effet causal de l'alcool ou du cannabis sur la responsabilité de l'accident mortel ou grave, l'estimation obtenue restera toujours une mesure d'association entachée de biais. L'alcool ayant pour effet d'augmenter la vitesse, le cannabis de la diminuer. En souhaitant évaluer l'effet du cannabis sur le risque d'être responsable d'un accident mortel, les auteurs de l'étude SAM avaient pressenti l'existence d'un biais dans les analyses en responsabilité (Laumon et al., 2005).
L'originalité de notre travail réside dans la démonstration formelle, mathématique de l'existence de ce biais. Pour cela, nous nous sommes appuyés sur une théorie novatrice, encore peu utilisée en épidémiologie : la théorie de l'inférence causale, et en particulier les modèles causaux structuraux. Cette théorie met à disposition des outils et définitions permettant de déterminer l'identifiabilité et l'estimation d'une quantité causale à partir de données observationnelles. Une quantité causale est définie à l'aide de variables contrefactuelles, partiellement observées. Et l'identifiabilité de cet effet causal est possible s'il peut s'exprimer en fonction des probabilités observées. Cette identifiabilité repose sur la structure d'un modèle graphique, le DAG, représentant les différentes relations entre les variables composant le mécanisme causal aboutissant à la survenue de l'événement. Si la quantité causale est identifiable, d'autres critères permettent de déterminer quelles variables sont suffisantes pour son estimation. En présence de mécanisme de sélection, les modèles causaux structuraux sont également utiles. Une variable S représentant la sélection préférentielle d'individus est ajoutée au DAG, et selon sa localisation la quantité causale est identifiable ou non. C'est grâce aux récents résultats de Bareinboim and Pearl (2012), Bareinboim and Tian (2015) que nous avons pu obtenir une approche formalisée de l'analyse en responsabilité.
Grâce à cette théorie, nous avons pu également proposer une définition formelle de la validité externe et interne de la mesure d'association estimable sur les individus pour lesquels $S = 1$. La mesure d'association estimable $OR(X, Y|W = w, S = 1)$ est valide d'un point de vue externe si $OR(X, Y|W = w, S = 1) = COR(X, Y|W = w)$, et donc en d'autres termes si $COR(X, Y|W = w)$ est identifiable. $OR(X, Y|W = w, S = 1)$ est valide d'un point de vue interne si $OR(X, Y|W = w, S = 1) = COR(X, Y|W = w, S = 1)$. Nous avons montré que dès lors que les données d'accidentés dépendaient de la sévérité de l'accident, l'odds ratio causal de l'exposition X sur la responsabilité $\mathcal{R}$ n'était pas identifiable et donc pas valide d'un point de vue externe. Dans ce cas, la question de la validité interne de la mesure se posait. Et contrairement à ce que la majorité des épidémiologistes prétendent, la mesure estimée n'a pas de validité interne non plus. Conditionner les estimations sur les accidents graves conduit à ajuster sur un collider et descendant de X, et à rompre la condition d'ignorabilité conditionnelle, une des conditions nécessaires à l'identifiabilité d'une quantité causale. En utilisant des arguments causaux, nous avons également pu fournir une description formelle de la mesure d'association estimable dans les analyses en responsabilité, lorsque les données d'inclusion dépendent de la sévérité de l'accident. La mesure estimable est alors à interpréter avec prudence car elle n'est pas fondée sur des populations identiques et stables.

Nous avons pu ensuite illustrer nos résultats théoriques avec des résultats numériques, afin de fournir un premier ordre de grandeur du biais induit par les analyses en responsabilité. Nos ré-

sultats théoriques montraient notamment que l'odds ratio causal de l'exposition X sur $\mathcal{R}$ était approximable dès lors que $X \perp\!\!\!\perp \mathcal{A}|(F, W)$, où $\mathcal{A}$ représente l'accident grave. En considérant un modèle génératif relativement simple, nous avons montré que l'ampleur du biais entre la mesure estimable $OR(X, \mathcal{R}|W = w, \mathcal{A} = 1)$ et les odds ratios causaux $COR(X, \mathcal{R}|W = w)$ et $COR(X, F|W = w)$ dépendait de cet écart à l'indépendance $X \perp\!\!\!\perp A|(F, W)$. Ce sont donc l'intensité de la corrélation entre l'exposition X et la vitesse élevée V, et celle entre V et $\mathcal{A}$ qui détermine en premier l'ampleur du biais entre les trois quantités. Ensuite, d'autres paramètres modifient légèrement le sens et l'ampleur de biais : l'intensité de la relation entre un facteur de confusion W et V et l'interaction de X et W sur V. La variation d'autres paramètres ne modifie que très marginalement les biais entre les trois quantités. Dans les configurations étudiées dans le cadre de cette thèse, l'ampleur des biais entre ces quantités varie dans un rapport de 1 à 5. Notons que $OR(X, \mathcal{R}|W = w, \mathcal{A} = 1)$ semble mieux estimer $COR(X, F|W = w)$ que $COR(X, \mathcal{R}|W = w)$. Dans un premier cas où l'exposition X étudiée augmenterait le fait de rouler à une vitesse élevée V, la mesure estimable dans les analyses en responsabilité $OR(X, \mathcal{R}|W = w, \mathcal{A} = 1)$ sous-estime les odds ratios causaux d'intérêt. En revanche, lorsque X contribue à diminuer la probabilité de rouler à une vitesse élevée, la mesure estimable $OR(X, \mathcal{R}|W = w, \mathcal{A} = 1)$ sur-estime les odds ratios causaux d'intérêt. Dans ce dernier cas, il est même possible de conclure à un effet délétère de l'exposition alors qu'elle aurait un effet protecteur.

Ces derniers résultats sont particulièrement utiles pour donner un premier ordre de grandeur du sens et de l'ampleur du biais dans le cas de l'estimation du risque causal de l'alcool ou du cannabis sur la responsabilité de l'accident grave. Cela signifie notamment que les effets de l'alcool sont globalement sous-estimés dans les analyses en responsabilité alors que ceux du cannabis sont surestimés. Ces résultats sont également intéressants dans la perspective d'une hiérarchisation des facteurs causaux. La hiérarchisation des facteurs se réalise la plupart du temps en estimant des risques attribuables, c'est-à-dire en calculant le pourcentage d'accident attribuable à un facteur causal donné. Dans l'étude SAM (Laumon et al., 2011), étude focalisée sur l'effet de l'alcool et des stupéfiants sur l'insécurité routière, les auteurs ont notamment conclu que 31 % des accidents seraient attribuables à l'alcool, et que 4 % seraient attribuables au cannabis. Nos résultats allant dans le sens d'une sous-estimation des mesures causales, et donc des risques attribuables, pour l'alcool, et une sur-estimation pour le cannabis, cela signifie que la hiérarchisation de ces deux facteurs reste dans l'ordre présenté à l'issue de l'étude SAM.

Dans l'optique de hiérarchiser les facteurs, les analyses causales en médiation peuvent être intéressantes à considérer en vue de décomposer l'effet total d'une cause en effet direct et effet indirect. En effet, lorsque la relation entre l'exposition X et l'événement Y est médiée par un facteur intermédiaire M, la médiation permet notamment de calculer la part médiée par le médiateur dans la relation de X à Y. Or, dans le cadre de la sécurité routière, il serait intéressant de décomposer par exemple l'effet total de l'alcool sur la responsabilité de l'accident en considérant la vitesse comme médiateur. On peut en effet considérer que l'alcool à deux effets sur la responsabilité de l'accident. D'une part, l'alcool à un premier effet excitant qui se traduit par l'augmentation de la vitesse de conduite. La vitesse ayant elle même un impact sur la responsabilité de l'accident. Ce premier effet de l'alcool est donc un effet médié par la vitesse. Le deuxième effet de l'alcool sera qualifié d'effet direct, et correspond à la baisse des capacités psychomotrices. L'analyse causale en médiation permettrait d'estimer des effets total, direct et indirect et d'estimer la proportion médiée par la vitesse, dans la relation entre l'alcool et la responsabilité de l'accident. Néanmoins, même en négligeant les biais de sélection identifiés dans le

chapitre 2 et en considérant nos données comme des données cas/témoins, l'analyse en médiation ne peut se faire de manière classique : un modèle de régression sur l'événement d'intérêt Y et un modèle de régression sur le médiateur M, tous deux réalisés sur l'ensemble de l'échantillon. Dans cette configuration, le biais de collision est en jeu puisque l'association entre X et M est biaisée si l'inclusion dépend de Y. Une des solutions est de réaliser le modèle régressant le médiateur M uniquement sur les témoins de l'étude (VanderWeele and Vansteelandt, 2010), et non sur l'ensemble de l'échantillon. Cela pose cependant un problème sur nos données. En effet, il existe par exemple trop peu d'alcoolisés non responsables pour obtenir des estimations précises.

Dans le cadre de cette thèse, nous avons insisté sur le biais de sélection, mais d'autres limites peuvent être soulevées quant à l'identifiabilité de quantités en sécurité routière.
Au delà du biais de sélection résiduel qui perdure dans les analyses en responsabilité, se pose tout d'abord la fiabilité de la détermination de la responsabilité, et donc la possible présence d'un biais de classement. Pour un conducteur impliqué dans un accident, la responsabilité est déterminée par la police, un algorithme (Robertson and Drummer, 1994, Terhune, 1986) ou par des experts. Même si responsabilité assignée est dans tous les cas entachée d'une erreur de mesure, il est admis que la responsabilité est assignée en ne prenant pas en considération toute information concernant l'exposition X (Brubacher et al., 2014). L'identifiabilité d'odds ratios causaux a été étudiée en insistant sur le biais de sélection induit par les analyses en responsabilité. Une réflexion plus approfondie mérite d'être menée sur les facteurs de confusion également nécessaires pour l'identifiabilité d'odds ratios causaux. En effet, une démarche à suivre serait de considérer un DAG bien plus complexe que celui que nous avons considéré avec l'ensemble des facteurs de confusion qu'il est possible de récolter, mais également ceux dont nous ne disposerons probablement jamais (par exemple la propension au risque). Car, selon la localisation de certains facteurs inobservés dans le DAG, l'identifiabilité de quantité causale est parfois impossible.
Concernant la construction du DAG, elle a jusqu'à maintenant été déterminée par des experts. Il existe également des algorithmes permettant de construire le DAG à partir des données observées. Plusieurs méthodes sont à disposition : le PC-algorithm, le score based method, les GES-algorithm ou encore le FCI algorithm (Kalisch and Bühlmann, 2013). Par exemple, le PC-algorithm propose de construire le DAG en testant les indépendances conditionnelles (Spirtes et al., 2000). Dans un premier temps, un squelette, composé uniquement de toutes les flèches non dirigées, est constitué. Si une indépendance est trouvée, la flèche non dirigée est supprimée. A l'aide du critère de d-séparation, les flèches sont ensuite dirigées.
Enfin, notons que les données de sécurité routière sont composées des accidents à un et plusieurs véhicules. Pour les accidents à plusieurs véhicules, les variables sont en général disponibles pour tous les conducteurs impliqués dans le même accident. Par conséquent, les conducteurs non-responsables peuvent être considérés comme des témoins appariés. Une réflexion pourrait être alors menée sur la combinaison d'analyse standard, et d'analyse sur données appariées (Mansournia et al., 2013).

Même si nos recherches comportent certaines limites, elles ne manqueront pas d'alimenter les discussions des études fondées sur l'analyse en responsabilité. Une perspective possible serait notamment de réfléchir à quel type d'information additionnelle permettrait de contourner le biais de sélection. En effet, la connaissance de la prévalence $\mathbb{P}(X)$ permet l'identifiabilité de $COR(X, \mathcal{R}|W = w)$, $COR(X, \mathcal{R})$, à partir des données disponibles, même si la structure du

DAG ne permet pas l'identifiabilité de quantité causale (Hernán et al., 2004). D'autre part, nos réflexions sont également utiles en dehors du champ de la sécurité routière. En effet, de nombreux champs de recherche en épidémiologie sont confrontés à des problèmes de sélection d'individus, et donc de nombreux chercheurs se questionnent sur la validité interne de leurs mesures. Citons par exemple le paradoxe de l'obésité qui suscite l'intérêt de nombreux épidémiologistes et biostatisticiens. Plusieurs études observationnelles ont conclu que lorsque l'inférence est réalisée chez des patients souffrant d'une maladie chronique (diabète (Carnethon et al., 2012) ou insuffisance cardiaque (Oreopoulos et al., 2008)), l'obésité protègerait d'une mortalité précoce. Or, l'OR estimé parmi les patients atteints de maladie chronique n'est pas un effet total mais un effet direct. Le conditionnement sur un médiateur, descendant de l'exposition X, ruine une des conditions nécessaire à l'identifiabilité de l'effet causal, et l'OR estimé n'a ainsi pas de valeur causale. Ainsi, contrairement à ce que certains auteurs ont prétendu (Sperrin et al., 2016), le biais de collision suffit à expliquer le paradoxe de l'obésité (Viallon and Dufournet, 2017a). Et la réponse que nous avons apportée à ces travaux a permis de mettre en avant d'autres mesures d'intéret, issues de l'analyse en médiation, comme les effets directs contrôlés, et également de susciter l'intérêt de la communauté scientifique (Sperrin et al., 2017, Vansteelandt, 2017).

Bibliographie

Aarts, L. and I. van Schagen (2006) : "Driving speed and the risk of road crashes : A review," *Accident Analysis & Prevention*, 38, 215–224.

Amoros, E., J.-L. Martin, and B. Laumon (2008) : "Estimation de la morbidité routière, France, 1996-2004," *Bulletin épidémiologique hebdomadaire*, 19, 157–160.

Anker, S. and S. Haehling (2011) : "The obesity paradox in heart failure : accepting reality and making rational decisions," *Clinical Pharmacology & Therapeutics*, 90, 188–190.

Aronson, J. L. (1971) : "On the Grammar of 'Cause'," *Synthese*, 22, 414–430.

Asbridge, M., J. R. Brubacher, and H. Chan (2013) : "Cell phone use and traffic crash risk : a culpability analysis," *International Journal of Epidemiology*, 42, 259–267.

Balke, A. and J. Pearl (2011) : "Probabilistic evaluation of counterfactual queries," *Department of Statistics, UCLA*.

Bareinboim, E. and J. Pearl (2012) : "Controlling selection bias in causal inference," *Proceedings of The Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2012) ; JMLR*, 22, 100–108.

Bareinboim, E. and J. Tian (2015) : "Recovering causal effects from selection bias," in *ICAIL '15 : Proceedings of the 15th International Conference on Artificial Intelligence and Law*, New York, NY, USA : ACMI, 3475–3481.

Bareinboim, E., J. Tian, and J. Pearl (2014) : "Recovering from selection bias in causal and statistical inference," in *Proceedings of the 28th National Conference on Artificial Intelligence (AAAI)*, Menlo Park, CA : CE Brodley and P. Stone, eds, AAAI Press, 2410–2416.

Baron, R. M. and D. A. Kenny (1986) : "The moderator–mediator variable distinction in social psychological research : Conceptual, strategic, and statistical considerations." *Journal of Personality and Social Psychology*, 51, 1173.

Benichou, J. (2001) : "A review of adjusted estimators of attributable risk," *Statistical Methods in Medical Research*, 10, 195–216.

Berkson, J. (1946) : "Limitations of the application of fourfold table analysis to hospital data," *Biometrics Bulletin*, 2, 47–53.

Bickel, P. J., E. A. Hammel, J. W. O'Connell, et al. (1975) : "Sex bias in graduate admissions : Data from berkeley," *Science*, 187, 398–404.

Blyth, C. R. (1972) : "On Simpson's Paradox and the Sure-Thing Principle," *Journal of the American Statistical Association*, 67, 364–366.

Bouaoun, L., M. M. Haddak, and E. Amoros (2015) : "Road crash fatality rates in france : A comparison of road user types, taking account of travel practices," *Accident Analysis & Prevention*, 75, 217–225.

Brubacher, J., H. Chan, and M. Asbridge (2014) : "Culpability analysis is still a valuable technique," *International Journal of Epidemiology*, 43, 270–272.

Bruzzi, P., S. Green, D. Byar, L. Brinton, and C. Schairer (1985) : "Estimating the population attributable risk for multiple risk factors using case-control data," *American Journal of Epidemiology*, 122, 904–914.

Carnethon, M. R., P. J. D. De Chavez, M. L. Biggs, C. E. Lewis, J. S. Pankow, A. G. Bertoni, S. H. Golden, K. Liu, K. J. Mukamal, B. Campbell-Jenkins, et al. (2012) : "Association of weight status with mortality in adults with incident diabetes," *JAMA*, 308, 581–590.

Cartwright, N. (1979) : "Causal laws and effective strategies," *Noûs*, 13, 419–437.

Cartwright, N. (1989) : "Nature's capacities and their measurement," *Oxford Scholarship*.

Chambaz, A., I. Drouet, and J. C. Thalabard (2013) : "Causalité, trilogue," *MAP5 2013-10*, URL `https://hal.archives-ouvertes.fr/hal-00807337v1`.

Cleveland Clinic (22 avril 2017) : "Cleveland clinic study finds obesity as top cause of preventable life-years lost," URL `https://newsroom.clevelandclinic.org/2017/04/22/cleveland-clinic-study-finds-obesity-top-cause-preventable-life-years-lost/`.

Davis, G. A. and Y. Gao (1993) : "Statistical methods to support induced exposure analyses of traffic accident data," *Transportation Research Record (1401)*, 43–49, URL `https://trid.trb.org/view.aspx?id=383957`.

Didelez, V., S. Kreiner, and N. Keiding (2010) : "Graphical models for inference under outcome-dependent sampling," *Statistical Science*, 25, 368–387.

Drouet, I. and M. Kistler (2012) : *Causes, probabilités, inférences*, Ph.D. thesis, Vuibert, Paris.

DRUID (2008) : "Deliverable 1.1.2a : Meta-analysis of empirical studies concerning the effects of alcohol on safe driving. Driving under the influence of drugs, alcohol and medicines," Technical report.

Dufournet, M., E. Lanoy, J.-L. Martin, and V. Viallon (2017) : "Causal inference to formalize responsibility analyses," *Journal of causal inference (under review)*, URL `https://arxiv.org/abs/1607.03775`.

Duncan, O. D. (1975) : *Introduction to Structural Equation Models (Studies in population)*, Elsevier, iSBN 978-1-4832-9532-9.

Elvik, R. (2009) : *The Power Model of the relationship between speed and road safety : update and new analyses*, TOI 1034/2009, Institute of Transport Economics.

Elvik, R., P. Christensen, and A. Amundsen (2004) : "Speed and road accidents," *An evaluation of the Power Model. TØI report*, 740, 2004.

Fair, D. (1979) : "Causation and the flow of energy," *Erkenntnis*, 14(3), 219–250.

France, d. l. Ministère de l'écologie, du développement durable et de l'aménagement du territoire, Direction de la recherche et de l'animation scientifique et technique, France, and Délégation interministérielle à la sécurité routière (2008) : *Gisements de sécurité routière les deux-roues motorisés*, Paris - La Défense : PREDIT.

Glennan, S. S. (1996) : "Mechanisms and the nature of causation," *Erkenntnis*, 44, 49–71.

Glymour, M. and S. Greenland (2008) : "Causal diagrams," in *Modern epidemiology*, Philadelphia, 3rd ed. lippincott williams & wilkins edition, 183–209.

Greenland, S. (2003) : "Quantifying biases in causal models : classical confounding vs collider-stratification bias," *Epidemiology*, 14, 300–306.

Greenland, S., J. Pearl, and J. M. Robins (1999) : "Causal Diagrams for Epidemiologic Research," *Epidemiology*, 10, 37–48.

Haavelmo, T. (1943) : "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, 1–12.

Haddon Jr, W. (1972) : "A logical framework for categorizing highway safety phenomena and activity." *Journal of Trauma and Acute Care Surgery*, 12, 193–207.

Hernán, M. A., S. Hernández-Díaz, and J. M. Robins (2004) : "A structural approach to selection bias," *Epidemiology*, 15, 615–625.

Hernán, M. A. and S. L. Taubman (2008) : "Does obesity shorten life ? The importance of well-defined interventions to answer causal questions," *International Journal of Obesity*, 32, S8–S14.

Hill, A. B. (1965) : "The environment and disease : association or causation ?" *Proceedings of the Royal Society of Medicine*, 58(5), 295–300.

Hirano, K. and G. W. Imbens (2004) : "The propensity score with continuous treatments," *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164, 73–84.

Holland, P. W. (1986) : "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81(396), 945–960.

Hume, D. (1739) : *Traité de la nature humaine. Livre 1 : l'entendement*, Paris, gf-flammarion edition.

Imai, K., L. Keele, and D. Tingley (2010) : "A general approach to causal mediation analysis." *Psychological Methods*, 15, 309–334.

Imbens, G. W. (2000) : "The role of the propensity score in estimating dose-response functions," *Biometrika*, 87, 706–710.

INRETS (2005) : *Sur la route, la sécurité : paroles de chercheurs : 20 ans de recherche à l'INRETS : 1985-2005*, TOI 1034/2009, Collections de l'INRETS.

Jones, A. W., A. Holmgren, and F. C. Kugelberg (2007) : "Concentrations of scheduled prescription drugs in blood of impaired drivers : considerations for interpreting the results," *Therapeutic Drug Monitoring*, 29, 248–260.

Kalisch, M. and P. Bühlmann (2013) : "Causal structure learning and inference : A selective review." *Quality Technology & Quantitative Management*, 11(1), 3–21.

Kelly, E., S. Darke, and J. Ross (2004) : "A review of drug use and driving : epidemiology, impairment, risk factors and risk perceptions," *Drug and Alcohol Review*, 23(3), 319–344.

Lajous, M., A. Bijon, G. Fagherazzi, M.-C. Boutron-Ruault, B. Balkau, F. Clavel-Chapelon, and M. A. Hernán (2014) : "Body mass index, diabetes, and mortality in french women : explaining away a "paradox"," *Epidemiology (Cambridge, Mass.)*, 25(1), 10–14.

Laumon, B., B. Gadegbeku, J.-L. Martin, and M.-B. Biecheler (2005) : "Cannabis intoxication and fatal road crashes in france : population based case-control study," *BMJ*, 331(7529), 1371.

Laumon, B., B. Gadegbeku, J.-L. Martin, and le groupe SAM (2011) : "Stupéfiants et accidents mortels : l'étude SAM, 2ème partie : analyse épidémiologique," Technical report, Paris : OFDT.

Lenguerrand, E., J.-L. Martin, A. Moskal, B. Gadegbeku, B. Laumon, and SAM group (2008) : "Limits of the quasi-induced exposure method when compared with the standard case-control design. Application to the estimation of risks associated with driving under the influence of cannabis or alcohol," *Accident ; Analysis and Prevention*, 40(3), 861–868.

Lewis, D. (1973) : "Causation," *The Journal of Philosophy*, 70(17), 556–567.

Lewis, D. (2000) : "Causation as Influence," *The Journal of Philosophy*, 97(4), 182–197.

Machamer, P., L. Darden, and C. F. Craver (2000) : "Thinking about Mechanisms," *Philosophy of Science*, 67(1), 1–25.

Mackie, J. L. (1974) : *The Ciment of the universe : a study of causation*, Oxford, oxford university press edition.

Mansournia, M. A., M. A. Hernán, and S. Greenland (2013) : "Matched designs and causal diagrams," *International Journal of Epidemiology*, 42, 860–869.

Martin, J.-L., B. Gadegbeku, D. Wu, V. Vivian, and B. Laumon (2017) : "Cannabis, alcohol and fatal road accidents," *PLoS One (en révision)*.

Meek, C. and C. Glymour (1994) : "Conditioning and Intervening," *The British Journal for the Philosophy of Science*, 45, 1001–1021.

Michael Walsh, J., R. Flegel, R. Atkins, L. A. Cangianelli, C. Cooper, C. Welsh, and T. J. Kerns (2005) : "Drug and alcohol use among drivers admitted to a Level-1 trauma center," *Accident Analysis & Prevention*, 37, 894–901.

Mignot, D. and J. P. Mizzi (2011) : "L'insécurité routière, facteurs et mesures : des enseignements pour la france. revue de la littérature scientifique réalisée par l'ifsttar à la demande de la "mission parlementaire d'information relative à l'analyse des causes des accidents de la circulation et à la prévention routière"." Technical report.

Morabia, A. (1991) : "On the origin of Hill's causal criteria," *Epidemiology (Cambridge, Mass.)*, 2, 367–369.

Nguyen, Q. C., T. L. Osypuk, N. M. Schmidt, M. M. Glymour, and E. J. Tchetgen Tchetgen (2015) : "Practical Guidance for Conducting Mediation Analysis With Multiple Mediators Using Inverse Odds Ratio Weighting," *American Journal of Epidemiology*, 181, 349–356.

Nilsson, G. (2004) : "Traffic safety dimensions and the power model to describe the effect of speed on safety," *Bulletin-Lunds Tekniska Högskola, Inst för Teknik och Samhälle, Lunds Universitet*, 221.

OECD/ECMT Transport Research Centre, European Conference of Ministers of Transport, Organisation for Economic Co-operation and Development, and United Nations, eds. (2006) : *Young drivers : the road to safety*, Paris : Organisation for Economic Cooperation and Development ; European Conference of Ministers of Transport.

ONISR (2014) : "Les infractions au code de la route et l'impact sur le permis à points - bilan 2014 (onisr)," .

ONISR (2015) : "La sécurité routière en france. bilan de l'accidentalité de l'année 2015," , 80,84,86.

ONU (2009) : "Déclaration de moscou adoptée à l'issue de la première conférence ministérielle mondiale sur la sécurité routière," URL `http://www.un.org/fr/roadsafety/documents.shtml`.

Oreopoulos, A., R. Padwal, K. Kalantar-Zadeh, G. C. Fonarow, C. M. Norris, and F. A. McAlister (2008) : "Body mass index and mortality in heart failure : a meta-analysis," *American Heart Journal*, 156, 13–22.

Pearl, J. (1995) : "Causal diagrams for empirical research," *Biometrika*, 82, 669–688.

Pearl, J. (2000) : *Causality : models, reasoning, and inference*, Cambridge, U.K. ; New York : Cambridge University Press.

Pearl, J. (2001) : "Direct and Indirect Effects," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 411–420, 978-1-55860-800-9.

Pearl, J. (2009) : "Causal inference in statistics : An overview," *Statistics Surveys*, 3, 96–146.

Perchonok, K. (1978) : "Identification of specific problems and countermeasures targets for reducing alcohol related casualties," Washington, DC : National Highway Traffic Safety Administration, U.S. Department of Transportation.

Price, H. (1991) : "Agency and Probabilistic Causality," *The British Journal for the Philosophy of Science*, 42, 157–176.

Richardson, T. S. and J. M. Robins (2013) : "Single world intervention graphs (swigs) : A unification of the counterfactual and graphical approaches to causality," *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128, 2013.

Robertson, M. D. and O. H. Drummer (1994) : "Responsibility analysis : A methodology to study the effects of drugs in driving," *Accident Analysis & Prevention*, 26, 243–247.

Robins, J. M. (2001) : "Data, design, and background knowledge in etiologic inference," *Epidemiology*, 12, 313–320.

Robins, J. M. and S. Greenland (1992) : "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 3, 143–155.

Robins, J. M., M. A. Hernan, and B. Brumback (2000) : "Marginal structural models and causal inference in epidemiology," *Epidemiology (Cambridge, Mass.)*, 11(5), 550–560.

Ronen, A., H. S. Chassidim, P. Gershon, Y. Parmet, A. Rabinovich, R. Bar-Hamburger, Y. Cassuto, and D. Shinar (2010) : "The effect of alcohol, THC and their combination on perceived effects, willingness to drive and performance of driving and non-driving tasks," *Accident Analysis & Prevention*, 42, 1855–1865.

Rosenbaum, P. R. and D. B. Rubin (1983) : "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rothman, K. J. (1976) : "Causes," *American Journal of Epidemiology*, 104, 587–592.

Rothman, K. J. and S. Greenland (2005) : "Causation and causal inference in epidemiology," *American Journal of Public Health*, 95, S144–S150.

Rothman, K. J., S. Greenland, and T. L. Lash (2008) : *Modern Epidemiology*, Lippincott Williams & Wilkins.

Rubin, D. B. (1974) : "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66, 688–701.

Rubin, D. B. (2004) : "On principles for modeling propensity scores in medical research," *Pharmacoepidemiology and drug safety*, 13, 855–857.

Russell, B. (1948) : *Human Knowledge : its scope and limits*, Londres, routledge edition.

Salmi, L. R., L. Orriols, and E. Lagarde (2014) : "Comparing responsible and non-responsible drivers to assess determinants of road traffic collisions : time to standardise and revisit," *Injury Prevention*, 20, 380–386.

Sanghavi, P. (2013) : "Commentary : Culpability analysis won't help us understand crash risk due to cell phones," *International Journal of Epidemiology*, 42, 267–269.

Schwilke, E. W., M. I. Sampaio dos Santos, and B. K. Logan (2006) : "Changing Patterns of Drug and Alcohol Use in Fatally Injured Drivers in Washington State," *Journal of Forensic Sciences*, 51, 1191–1198.

Simpson, E. H. (1951) : "The Interpretation of Interaction in Contingency Tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, 13, 238–241.

Skyrms, B. (1980) : *Causal Necessity : A Pragmatic Investigation of the Necessity of Laws*, volume 18, Yale University Press.

Smith, H. and R. Popham (1951) : "Blood alcohol levels in relation to driving." *Canadian Medical Association Journal*, 65, 325–328.

Sperrin, M., J. Candlish, E. Badrick, A. Renehan, and I. Buchan (2016) : "Collider bias is only a partial explanation for the obesity paradox :," *Epidemiology*, 27, 525–530.

Sperrin, M., J. Candlish, E. Badrick, A. Renehan, and I. Buchan (2017) : "The authors respond," *Epidemiology*, 28(5), e46.

Spirtes, P., C. N. Glymour, and R. Scheines (2000) : *Causation, prediction, and search*, Cambridge MA, MIT press.

Splawa-Neyman, J. (1990) : "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Statistical Science*, 5, 465–472.

Stamatiadis, N. and J. A. Deacon (1997) : "Quasi-induced exposure : methodology and insight," *Accident Analysis & Prevention*, 29(1), 37–52.

Steinmeyer, S., H. Ohr, H. J. Maurer, and M. R. Moeller (2001) : "Practical aspects of roadside tests for administrative traffic offences in Germany," *Forensic Science International*, 121(1-2), 33–36.

Stokes, A. and S. H. Preston (2015) : "Smoking and reverse causation create an obesity paradox in cardiovascular disease," *Obesity*, 23(12), 2485–2490.

Suppes, P. (1970) : *A probabilistic theory of causality*, Amsterdam, north holland publishing company edition.

Tchetgen Tchetgen, E. J. (2013) : "Inverse odds ratio-weighted estimation for causal mediation analysis," *Statistics in Medicine*, 32(26), 4567–4580.

Terhune, K. (1986) : "Problems and methods in studying drug crash effects," *Alcohol, Drugs, and Driving*, 2(3-4), 1–14.

Tian, J. and J. Pearl (2002) : "A general identification condition for causal effects," in *Proceedings of the Eighteenth National Conference of Artificial Intelligence*, AAAI Press/ The MIT Press, 567–573.

Vaa, T. (2003) : "Impairments, diseases, age and their relative risk of accident involvement : results from meta-analysis 2003," Technical report, Transportokonomisk institutt.

VanderWeele, T. and S. Vansteelandt (2009) : "Conceptual issues concerning mediation, interventions and composition," *Statistics and Its Interface*, 2, 457–468.

VanderWeele, T. J. (2009) : "Marginal Structural Models for the Estimation of Direct and Indirect Effects :," *Epidemiology*, 20(1), 18–26.

VanderWeele, T. J. and J. M. Robins (2009) : "Minimal sufficient causation and directed acyclic graphs," *The Annals of Statistics*, 37(3), 1437–1465.

VanderWeele, T. J. and I. Shpitser (2011) : "A new criterion for confounder selection," *Biometrics*, 67(4), 1406–1413.

VanderWeele, T. J. and S. Vansteelandt (2010) : "Odds Ratios for Mediation Analysis for a Dichotomous Outcome," *American Journal of Epidemiology*, 172(12), 1339–1348.

Vansteelandt, S. (2017) : "Asking too much of epidemiologic studies : The problem of collider bias and the obesity paradox," *Epidemiology*, 28(5), e47–e49.

Viallon, V. and M. Dufournet (2017a) : "Can collider bias fully explain the obesity paradox ?" *Epidemiology*, URL `https://arxiv.org/abs/1612.06547`.

Viallon, V. and M. Dufournet (2017b) : "Re : Collider bias is only a partial explanation for the obesity paradox," *Epidemiology*, 28(5), e43–e45.

Viallon, V. and B. Laumon (2013) : "Fractions of fatal crashes attributable to speeding : Evolution for the period 2001-2010 in france," *Accident Analysis and Prevention*, 52, 250–256.

Wahlberg, A. E. (2009) : "The determination of fault in collisions," in G. Dorn, L Matthews and I. Glendon, eds., *Driver Behaviour and Accident Research Methodology : Unresolved Problems.*, Farnham, Surrey, IK : Ashgate Publishing, 101–120.

Wahlberg, A. E. and L. Dorn (2007) : "Culpable versus non-culpable traffic accidents ; what is wrong with this picture ?" *Journal of Safety Research*, 38(4), 453–459.

WHO (2015) : "Rapport de situation sur la sécurité routière dans le monde 2015," , p.12.

Williamson, E., R. Morley, A. Lucas, and J. Carpenter (2012) : "Propensity scores : From naive enthusiasm to intuitive understanding," *Statistical Methods in Medical Research*, 21(3), 273–293.

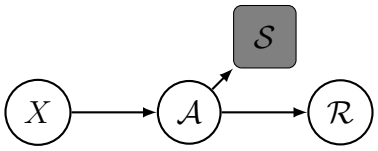
Woodward, J. (2003) : *Making Things Happen : A Theory of Causal Explanation*, Oxford University Press.

Wright, S. (1921) : "Correlation and causation," *Journal of agricultura research*, 20(7), 557–585.

Annexes

.1 Preuve que $\mathcal{R}_x$ n'est pas indépendant de X sachant W et $\mathcal{S}$ dans le cas où seuls les accidents sévères sont disponibles

FIGURE 7 – Cas simplifié décrivant le cas (ii) des analyses en responsabilité



Considérons le DAG simplifié de la figure 7 où $\mathcal{S} = \mathcal{A}$. Ce DAG est une version simplifié du DAG de droite de la figure 2.4 décrivant le mécanisme causal lorsque seuls les données d'accident sévères sont disponibles. Dans ce DAG, $\mathcal{R}_x \not\perp\!\!\!\perp X|\mathcal{S}$ et par conséquent, rien n'assure que $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{S})$ d'après le DAG de la figure 2.4. Sous le modèle causal représenté par la figure 7, nous avons $\mathcal{R} \perp\!\!\!\perp X|\mathcal{S}$, but $\mathcal{R}_x \not\perp\!\!\!\perp X|\mathcal{S}$. Cela peut également se démontrer par une preuve par l'absurde. En effet, si nous avions $\mathcal{R}_x \perp\!\!\!\perp X|\mathcal{S}$, alors

$$
\begin{aligned}
\mathbb{P}(\mathcal{R}_x = 1) &= \sum_s \mathbb{P}(\mathcal{R}_x = 1|\mathcal{S} = s)\mathbb{P}(\mathcal{S} = s) \\
&= \sum_s \mathbb{P}(\mathcal{R}_x = 1|\mathcal{S} = s, X = x)\mathbb{P}(\mathcal{S} = s) \quad \text{si } \mathcal{R}_x \perp\!\!\!\perp X|S \\
&= \sum_s \mathbb{P}(\mathcal{R} = 1|\mathcal{S} = s, X = x)\mathbb{P}(\mathcal{S} = s) \quad \text{sous l'hypothèse de cohérence} \\
&= \sum_s \mathbb{P}(\mathcal{R} = 1|\mathcal{S} = s)\mathbb{P}(\mathcal{S} = s) \quad \text{because } \mathcal{R} \perp\!\!\!\perp X|\mathcal{A} \text{ et } \mathcal{A} = \mathcal{S} \\
&= \mathbb{P}(\mathcal{R} = 1).
\end{aligned}
$$

Enfin, si $\mathcal{R}_x \perp\!\!\!\perp X|\mathcal{S}$, on obtiendrait $ACE = \mathbb{P}(\mathcal{R}_1 = 1) - \mathbb{P}(\mathcal{R}_0 = 1) = 0$, ce qui est généralement faux sous le DAG représenté par la figure 7.

.2 Illustration numérique lorsque X est catégorielle

FIGURE 8 – Effets causaux et association pour la modalité $j = 2$ de X sur l'échelle logarithmique dans le cas où $\beta_V = \beta_{X_1} = \beta_W = \gamma_W = 1$, $\gamma_F = 4$ et pour des valeurs variables des autres paramètres $\alpha_{X_1} et \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_{X_1} varie de 0 à 3.

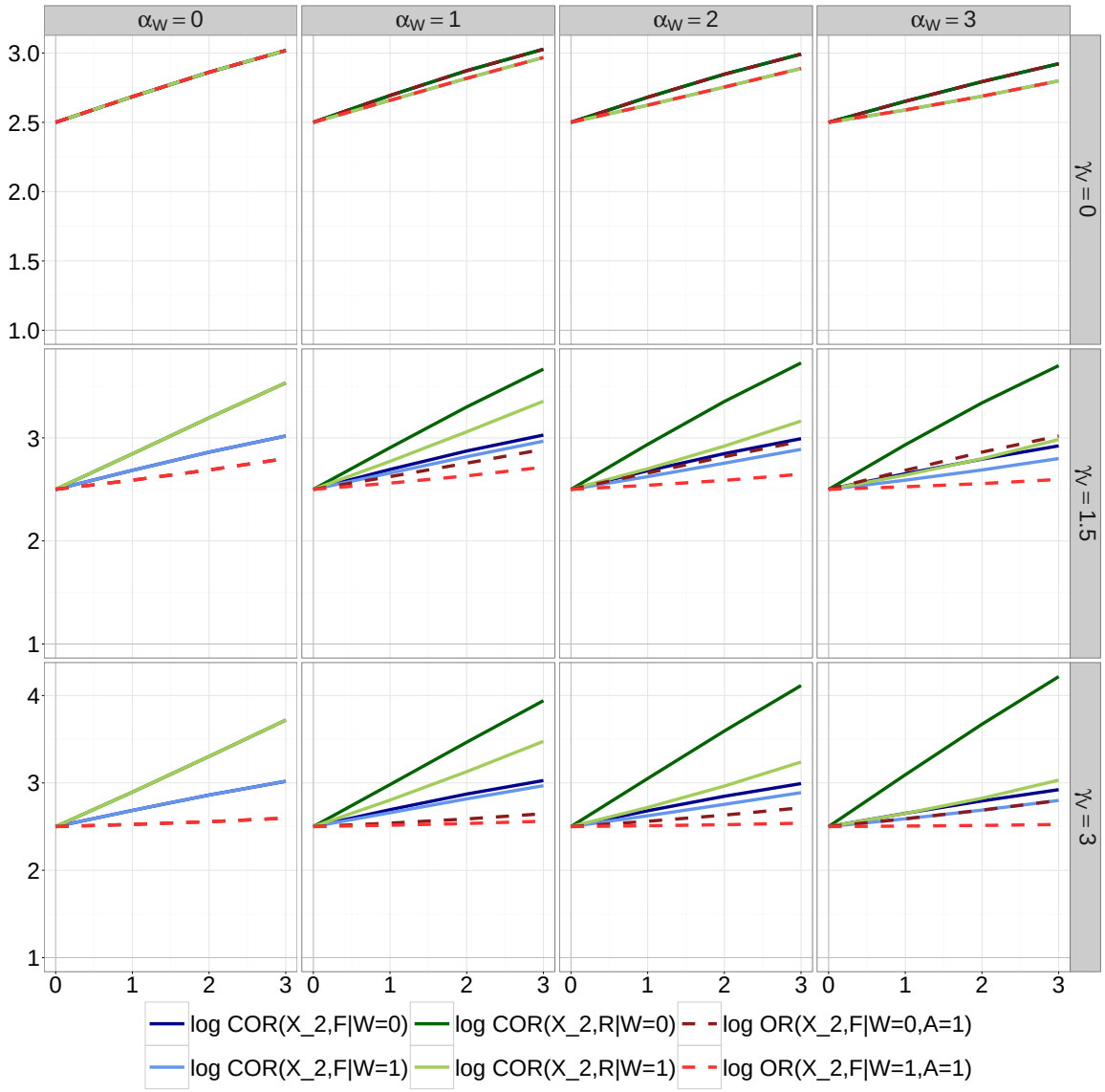


FIGURE 9 – Effets causaux et association pour chaque modalité $j = 3$ de X sur l'échelle logarithmique dans le cas où $\beta_V = \beta_{X_1} = \beta_W = \gamma_W = 1$, $\gamma_F = 4$ et pour des valeurs variables des autres paramètres $\alpha_{X_1} et \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_{X_1} varie de 0 à 3.

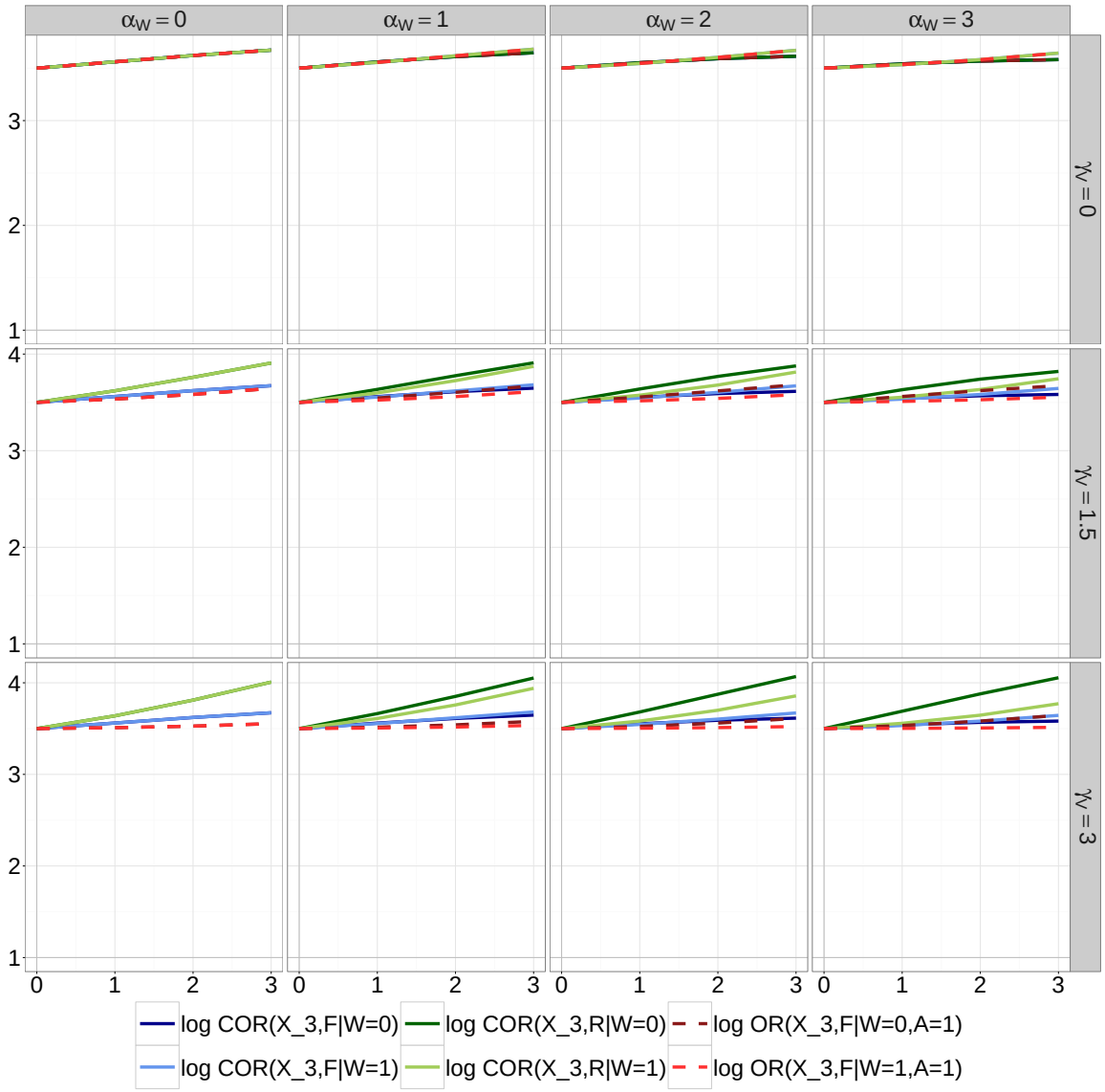
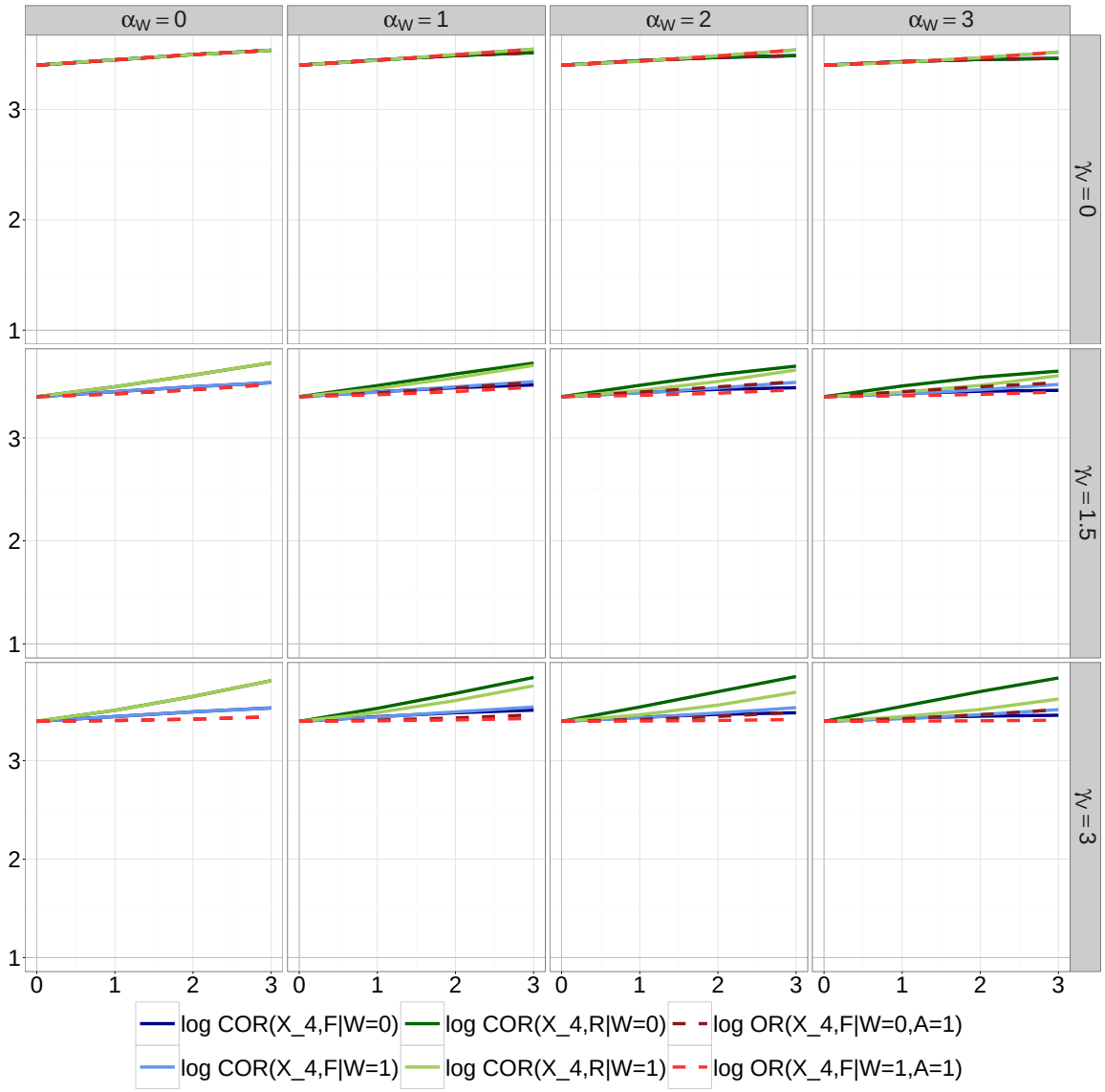


FIGURE 10 – Effets causaux et association pour chaque modalité $j = 4$ de X sur l'échelle logarithmique dans le cas où $\beta_V = \beta_{X_1} = \beta_W = \gamma_W = 1$, $\gamma_F = 4$ et pour des valeurs variables des autres paramètres $\alpha_{X_1} et \gamma_V$. Dans chaque cadran, le long de l'axe des abscisses, α_{X_1} varie de 0 à 3.



.

.3 Article soumis dans Journal of causal inference

arXiv:1607.03775v2 [stat.ME] 20 Apr 2017

Causal inference to formalize responsibility analyses in road safety epidemiology

M. Dufournet[(1)], E. Lanoy[(2),(3)], J.L. Martin[(1)], V. Viallon[(1)]

[(1)] Univ Lyon, Université Claude Bernard Lyon 1, Ifsttar, UMRESTTE, UMR T_9405, F-69373 LYON

[(2)] Gustave Roussy, Université Paris-Saclay, Service de biostatistique et d'épidémiologie, Villejuif, F-94805

[(3)] Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, F-94085

Abstract

The last few decades have seen the Structural Causal Model framework provide valuable tools to assess causal effects from observational data. In this article, we briefly review recent results regarding the recoverability of causal effects from selection biased data, and apply them to the case of responsibility analyses in road safety epidemiology. Our objective is to formally determine which effects can be unbiasedly estimated through this type of analyses, depending on available data. We consider the estimation of the causal effect of alcohol in two situations. First, when data are available for all types of crashes, we show that the causal odds-ratio of alcohol on the responsibility to trigger a car accident can be approximated under some assumptions. However, available data most often depends on the accident severity, which is related to speed. In this case, the causal odds-ratio of alcohol is not recoverable. We present numerical results to illustrate our arguments and the magnitude of the bias. We finally show that this bias makes association measures not only externally, but also internally invalid.

Keywords: causal inference; recoverability ; selection bias; responsibility analysis; road safety.

1 Introduction

Research into road safety must recognise causality. Preventing road crashes, in the interest of public health and safety, implies identifying the underlying causes of crashes. In the field, main causes of the occurrence and severity of road crashes due to human behavior have been established: driving under the influence of alcohol or drugs, inappropriate speed, distracted driving, etc. Questions asked by decision-makers now mostly concern the magnitude of their causal effects (on the excess risk, relative risk or odds-ratio scale), as well as the burden of deaths or victims attributable to these various causes of accident ([1]).

In order to measure causal effects, the ideal designs are interventional studies such as the randomized

clinical trial ([2]). However, interventional studies are not always possible in epidemiology or clinical research, either for ethical, costs or other technical reasons ([3]). Consequently, it is common to only have access to observational data. In this case, two main sources of bias often arise: confounding and selection bias ([4, 5, 6, 7]).

In the case of road safety epidemiology, no interventional study can be performed for clear ethical reasons. In addition, data are usually only available when the outcome of interest, the accident, occurred. Indeed, data are restricted to drivers and vehicles involved in road accidents only, and often to severe road accidents only (e.g., injury or fatal accidents). This extreme selection bias precludes the estimation of the effect of any exposure on the risk of car accident. Two closely related approaches have been adopted to circumvent this issue of "selected" drivers: quasi-induced exposure ([8]) and responsibility analysis ([9, 10, 11, 12]). They first rely on the assessment of the responsibility of each driver involved in crashes, usually from police reports. They then consist in comparing responsible drivers with non-responsible drivers among involved drivers. The underlying assumption is that non-responsible drivers represent a random sample of the general driving population that was "selected" to crash by circumstances beyond their control and therefore have the same risk factor profile as other drivers on the road at the same time ([12, 13]). A standard claim is that "if this randomness assumption is met, then the risk estimate derived from a responsibility analysis would be expected to be similar to that from a standard case-control study" ([12, 13]). Even if some authors have raised questions about the validity of these approaches ([14]), induced exposure and responsibility analysis are now widely adopted in the field ([15, 16, 17]).

However, the randomness assumption is questionable, because the inclusion in the dataset depends on the occurence of the accident. So in this article, we insist on the estimation of causal effect in the presence of selection bias (see discussion for confounding bias). We adress these questions by applying general concepts of causal inference theory under the Structural Causal Model (SCM) framework ([18, 19]). We mostly consider two cases: (i) when data are available for all types of crashes and the causal effect of interest is the effect of alcohol on the risk of being responsible for a road accident and (ii) when data are available for severe crashes only and the causal effect of interest is the effect of alcohol on the risk of being responsible for a severe road accident. These two types of available data correspond to different schemes of selection bias, which may alter the randomness assumption.

The article is organized as follows. In Section 2 below, we briefly recall some basics of the SCM framework and recent results regarding *the recoverability* of causal effects in the presence of selection bias. In Section 3, the application to responsibility analyses is presented, under both cases (i) and

(*ii*). In Section 4, we present results from a simple numerical analysis to illustrate the magnitude of bias under both cases. In Section 5, we conclude with some leads for future research.

2 Causal inference in the presence of selection bias

In this section, we present the Structural Causal Model (SCM) framework and study the recoverability of causal effects in the general context of selection biased data.

We consider the scenario where Y denotes a dichotomous outcome, X a binary exposure of interest, and $\boldsymbol{W}$ is a vector of additional categorical variables. We will denote by $\boldsymbol{\mathcal{V}}$ the set of observable variables $(X, \boldsymbol{W}, Y)$. The causal model $\mathcal{M}$ leading to Y can be graphically represented by a DAG G (Directed Acyclic Graph) ([20, 21, 22]). In the SCM framework, the DAG is associated to a set of structural funtions, each corresponding to one of the covariates in the DAG. See Appendix A for more details. This set of equations allows the definition of Y_x, the counterfactual outcome that would have been observed in the counterfactual world where exposure would have been set to $X = x$, for $x \in \{0, 1\}$. Then, causal effects can be precisely defined. In particular, in the simple case considered here where both X and Y are binary, the average causal total effect (ACE) of X on Y is defined as:

$$\text{ACE} = \mathbb{E}(Y_1 - Y_0) = \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1). \tag{1}$$

Here, the causal effect is defined on the excess risk scale, but causal risk ratio or odds-ratio can be defined similarly.

Causal inference is mainly concerned with the identification of $\mathbb{P}(Y_x = y)$ for $x \in \{0, 1\}$ and then of the causal effects, from the distribution of the observed variables $\mathbb{P}(\boldsymbol{\mathcal{V}} = \boldsymbol{v})$ ([23]). When there is no selection bias, criteria such as the back-door and the front-door criteria ensure the identifiability of causal effects ([20, 19]). When selection bias is present, the binary variable S indicating inclusion in the study has to be added to G, leading to a new DAG G_s. It is standard to represent S in a specific way in G_s to emphasize that S does not play any role in the causal model described by G but plays a role in the selection process since data are only available for individuals for whom $S = 1$ (see Figure 1). Then, the question is the recoverability of $\mathbb{P}(Y_x = y)$ for $x \in \{0, 1\}$ in terms of the observable distribution $\mathbb{P}(\boldsymbol{\mathcal{V}} = \boldsymbol{v}|S = 1)$ ([24, 23]). Recoverability can be seen as a generalization of identifiability in the presence of selection bias. Whether this selection bias may result in biased estimation of the causal effects of interest depends on the structure of G_s, and in particular on the arrows pointing to

or emanating from S in G_s ([25]). Two types of results allow us to determine whether the causal effect is recoverable depending on the structure of the DAG G_s. First, [23] define necessary and sufficient conditions ensuring the recoverability of the quantity $\mathbb{P}(Y_x = y)$, and as a byproduct the recoverability of the causal effects (causal excess risk, causal relative risk, causal odds-ratio), in the presence of selection bias. Second, under some assumptions, causal odds-ratios are recoverable from selection biased data in some situations where the distribution $\mathbb{P}(Y_x = y)$ and other causal effects are not ([24]). For illustration, and motivated by the context of responsibility analyses, we mostly focus on situations where inclusion in the study depends on the outcome Y, as in DAGs A and B of Figure 1. We also consider the case where S is a descendant of a mediator, as in DAG C of Figure 1.

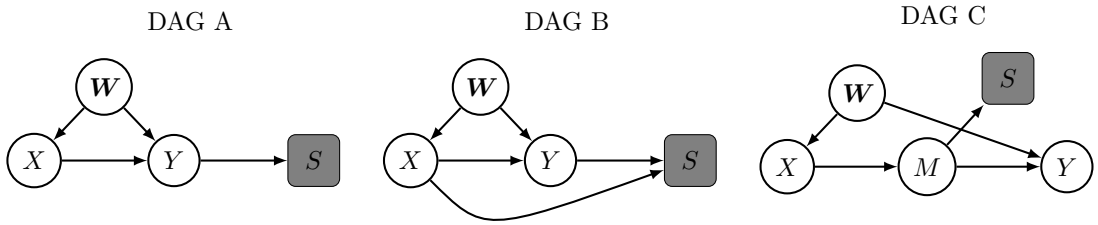


Figure 1: Examples of DAGs in the presence of selection bias

2.1 Recoverability of $\mathbb{P}(Y_x = y)$ in the presence of selection bias

A central result for the recoverability of $\mathbb{P}(Y_x = y)$ in the presence of selection bias is stated by [23]. As above, denote by G the DAG of interest, composed of the observable variables $\boldsymbol{\mathcal{V}}$, and by G_s the DAG obtained after the addition of the selection variable S in G. For any $\boldsymbol{C} \subseteq \boldsymbol{\mathcal{V}}$, further define $G_{\boldsymbol{C}}$ the subgraph of DAG G composed of the variables in $\boldsymbol{C}$ only. For any $V_i \in \boldsymbol{\mathcal{V}}$, further denote by $An(V_i)_G$ the union of V_i and the ancestors of V_i in DAG G. Then, Theorem 2 in [23] states that $\mathbb{P}(Y_x = y)$ is recoverable from selection biased data if and only if

$$\text{(C.1)} \qquad An(Y)_{G_{\boldsymbol{\mathcal{V}}\setminus X}} \cap An(S)_{G_s} = \emptyset.$$

Condition (C.1) above is violated in the three examples depicted in DAGs A, B and C of Figure 1. For instance, we have $Y \in An(Y)_{G_{\boldsymbol{\mathcal{V}}\setminus X}} \cap An(S)_{G_s}$ in DAGs A and B, and $M \in An(Y)_{G_{\boldsymbol{\mathcal{V}}\setminus X}} \cap An(S)_{G_s}$ in DAG C. Therefore, $\mathbb{P}(Y_x = y)$ is not recoverable under these types of selection bias. Examples of DAGs ensuring that $\mathbb{P}(Y_x = y)$ is recoverable from selection biased data can be found in [23].

2.2 Recoverability of causal odds-ratios in the presence of selection bias

Assume that $\boldsymbol{W}$ is a vector of confounders between X and Y, as in DAGs A, B and C. Here, we consider the recoverability of $\boldsymbol{w}$-specific causal odds-ratio

$$COR(X,Y|\boldsymbol{W}=\boldsymbol{w}) = \frac{\mathbb{P}(Y_1=1|\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(Y_1=0|\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(Y_0=1|\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(Y_0=0|\boldsymbol{W}=\boldsymbol{w})}.$$

The $\boldsymbol{w}$-specific causal odds-ratio is the causal odds-ratio in the stratum of the population defined by $\boldsymbol{W}=\boldsymbol{w}$. When $\boldsymbol{W}$ contains a set of sufficient confounders, conditional ignorability $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$ holds ([18, 26, 27]), so that $\mathbb{P}(Y_x=1|\boldsymbol{W}=\boldsymbol{w}) = \mathbb{P}(Y=1|X=x,\boldsymbol{W}=\boldsymbol{w})$. Then, $\boldsymbol{w}$-specific causal odds-ratio equals the adjusted odds-ratio

$$OR(X,Y|\boldsymbol{W}=\boldsymbol{w}) = \frac{\mathbb{P}(Y=1|X=1,\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(Y=0|X=1,\boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(Y=1|X=0,\boldsymbol{W}=\boldsymbol{w})/\mathbb{P}(Y=0|X=0,\boldsymbol{W}=\boldsymbol{w})}.$$

Following Definition 2 of [24], $OR(X,Y|\boldsymbol{W}=\boldsymbol{w})$, and then in our case $COR(X,Y|\boldsymbol{W}=\boldsymbol{w})$, are recoverable from selection biased data if the assumptions embedded in the DAG renders it expressible in terms of the observable distribution $\mathbb{P}(\boldsymbol{\mathcal{V}}=\boldsymbol{v}|S=1)$. The symmetry of the odds-ratio $OR(X,Y|\boldsymbol{W}=\boldsymbol{w}) = OR(Y,X|\boldsymbol{W}=\boldsymbol{w})$ makes it recoverable in some situations where $\mathbb{P}(Y_x=y)$ is not. More precisely, Theorem 1 in [24], or Corollary 4 in [28], states that $OR(X,Y|\boldsymbol{W}=\boldsymbol{w})$ is recoverable from selection biased data if and only if

$$\text{(C.2)} \qquad X \perp\!\!\!\perp S|(Y,\boldsymbol{W}) \text{ or } Y \perp\!\!\!\perp S|(X,\boldsymbol{W}).$$

This condition holds under DAG A, but is not guaranteed under DAGs B and C. Consequently, $OR(X,Y|\boldsymbol{W}=\boldsymbol{w})$ and then $COR(X,Y|\boldsymbol{W}=\boldsymbol{w})$ are recoverable under DAG A, while they are not under DAGs B and C.

3 Application to responsibility analyses

We can now check whether the conditions to recover causal total effects in the presence of selection bias hold in the particular setting of responsibility analyses. As a leading example, we will consider the estimation of the causal total effect of blood concentration of alcohol on "responsibility" among drivers involved in a car crash.

3.1 Two possible surrogate outcomes in responsibility analyses

First, let us formalize what "responsibility" means in responsibility analyses. In this context, the "responsibility" of a driver is not entirely driven by the determination of a legal fault. A driver

involved in a crash is considered as responsible of this crash if he committed a driving fault (lane departure, failure to obey traffic signs, driving against traffic, etc), which should trigger the crash. The responsibility, which is assigned after the crash from police reports, is a measure of the driving fault, which takes place before the crash. We will denote by F the binary variable indicating whether the driver commits a driving fault. Observe that this variable is defined for all drivers, not only those involved in a crash. Further denote by A the binary variable indicating whether this driver is involved in an accident. We also consider the binary variable R indicating whether the driver is responsible for an accident. So, $R = F \times A$, and we have:

$$\begin{cases} R = 1 & \text{if and only if } A = 1 \text{ and } F = 1 \\ R = 0 & \text{if } A = 1 \text{ and } F = 0 \\ R = 0 & \text{if } A = 0 \text{ even if } F = 1. \end{cases}$$

Most often, available data depend on the accident severity, which is related to speed. Define $\mathcal{A}$ the binary variable indicating the occurrence of a severe accident. Then, we can analogously define the binary variable $\mathcal{R}$ indicating whether the driver is responsible for a severe accident. So, $\mathcal{R} = F \times \mathcal{A}$.
The DAG G below represents a simplified causal mechanism which leads to an accident or a severe accident. We denote by $\boldsymbol{W}$ the set of all confounders (*e.g.*, age, sex, daytime, period of the year, speed limit etc.). Here, X is the binary exposure indicating whether a given driver has a blood concentration of alcohol above the legal limit. It is commonly accepted that alcohol X is related to speed V, so there is an arrow from X to V. V is the binary variable indicating whether a given driver drives at high speed. The fact that there is no arrow pointing from V to A, but there is an arrow pointing to V to $\mathcal{A}$ has to be clarified. This difference is related to the definition of F. Indeed, if speed was considered as a cause of an accident, it would mean that speed was inappropriate, and therefore that a fault has been committed. Therefore, V is a cause of A only through F. However, V has an effect on $\mathcal{A}$, because speed has an effect on the severity of an accident.
In Sections 3.2 and 3.3 below, we apply the principles presented in Section 2 on two particular situations: (i) the ideal one, when data are available for all types of crashes ($A = 1$) and the causal total effect of interest is that of X on the risk of being responsible of a road accident R and (ii) the common one, when data are available for severe crashes only ($\mathcal{A} = 1$) and the causal total effect of interest is that of X on the risk of being responsible of a severe road accident $\mathcal{R}$. Two simplified versions of these two cases are given in Figure 3. In case (i), because V has no effect on A, the effect of X on A is entirely mediated by F, while, in case (ii), the effect of X on $\mathcal{A}$ is mediated by both F and V.

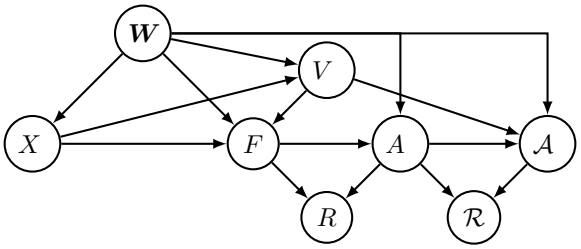


Figure 2: DAG G representing the causal mechanism leading to the occurence of an accident A or a severe accident $\mathcal{A}$.

(i) with data from all types of accidents

(ii) with data from severe accidents only

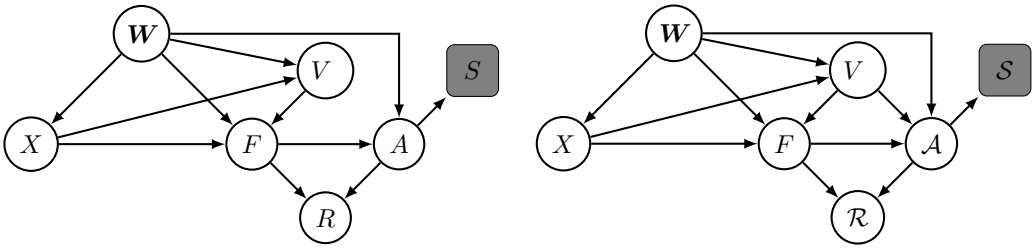


Figure 3: DAGs G_s in responsibility analyses

3.2 Responsibility analyses when data from all types of accidents are available

A version of the DAG G_s is represented in the left panel of Figure 3 when data from all types of accidents are available. As before, A represents the occurrence of a crash, on which selection depends in the sense that $S = 1 \Rightarrow A = 1$. In this case (i), the effect of X on A is entirely mediated by F and $X \perp\!\!\!\perp A|(F, \boldsymbol{W})$.

Because $A \in An(R)_{G_{\boldsymbol{V}\setminus X}} \cap An(S)_{G_s}$, results presented in Section 2 indicate that $\mathbb{P}(R_x = r)$ is not recoverable in this context, for $(x, r) \in \{0, 1\}$. Regarding $\boldsymbol{w}$-specific odds-ratios, $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ is not recoverable since neither $X \perp\!\!\!\perp S|(R, \boldsymbol{W})$ nor $R \perp\!\!\!\perp S|(X, \boldsymbol{W})$ is guaranteed. However, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ is recoverable because $X \perp\!\!\!\perp S|(F, \boldsymbol{W})$, which is implied by $X \perp\!\!\!\perp A|(F, \boldsymbol{W})$. In

other words, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ equals

$$= \frac{\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}, A = 1)/\mathbb{P}(F = 0|X = 1, \boldsymbol{W} = \boldsymbol{w}, A = 1)}{\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w}, A = 1)/\mathbb{P}(F = 0|X = 0, \boldsymbol{W} = \boldsymbol{w}, A = 1)} = \frac{\mathbb{P}(R = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}, A = 1)/\mathbb{P}(R = 0|X = 1, \boldsymbol{W} = \boldsymbol{w}, A = 1)}{\mathbb{P}(R = 1|X = 0, \boldsymbol{W} = \boldsymbol{w}, A = 1)/\mathbb{P}(R = 0|X = 0, \boldsymbol{W} = \boldsymbol{w}, A = 1)},$$

where the last equality comes from the fact that $F = R$ in the subpopulation $\{A = 1\}$. In addition, if $\mathbb{P}(F_x = 1|\boldsymbol{W} = \boldsymbol{w})$ and $\mathbb{P}(R_x = 1|\boldsymbol{W} = \boldsymbol{w})$ are both small for $x \in \{0, 1\}$, it can be shown that

$$COR(X, R|\boldsymbol{W} = \boldsymbol{w}) \approx COR(X, F|\boldsymbol{W} = \boldsymbol{w}). \tag{2}$$

See Appendix B for the proof of this result, which especially uses the fact that $X \perp\!\!\!\perp A|(F, \boldsymbol{W})$. Therefore, if $\mathbb{P}(F_x = 1|\boldsymbol{W} = \boldsymbol{w})$ and $\mathbb{P}(R_x = 1|\boldsymbol{W} = \boldsymbol{w})$ are small, $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ is approximately recoverable in case (i).

Note that the approximate recoverability of the $\boldsymbol{w}$-specific causal odds-ratios $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ is not sufficient for the approximate recoverability of the population causal odds-ratio $COR(X, R)$. However, the recoverability of adjusted causal odds-ratio $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ is still useful for instance to derive estimates of population attributable fractions (PAF). See Appendix C for more details.

3.3 Responsibility analyses when data from severe accidents only are available

In road safety epidemiology, available data are mainly composed of severe accidents. So inclusion in the study depends on the severity of the accident. A version of the DAG G_s is represented in the right panel of Figure 3. As before, $\mathcal{A}$ represents the occurrence of a severe crash, on which selection depends in the sense that $\mathcal{S} = 1 \Rightarrow \mathcal{A} = 1$. In this case (ii), the effect of V on A is not entirely mediated by F anymore and $X \perp\!\!\!\perp \mathcal{A}|(F, \boldsymbol{W})$ does not hold.

As in Section 3.2, neither $\mathbb{P}(\mathcal{R}_x = r)$ nor $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$ is recoverable. But now, it is not straightforward to approximate $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$ by a recoverable causal effect. First, Equation (2) is not valid here because $X \perp\!\!\!\perp \mathcal{A}|(F, \boldsymbol{W})$ is not guaranteed. Second, and above all, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ is not recoverable since neither $X \perp\!\!\!\perp \mathcal{A}|(F, \boldsymbol{W})$ nor $F \perp\!\!\!\perp \mathcal{A}|(X, \boldsymbol{W})$ generally holds. Note that, if there were no arrow between X and V, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ would be recoverable, and so $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$ could be approximated.

In our situation where there is an arrow between X and V, the validity of association measures conditioned on $\mathcal{S} = 1$ is therefore questionable. According to [29], "internal validity implies validity of

inference for the source population of study subjects", while a result is externally valid if it is generalizable "to people outside that population", that is the target population. A more formal definition can be proposed, based on causal considerations. In our case, the sample consists of individuals such that $\{\mathcal{S}=1\}$, while the source and target populations are made of individuals such that $\{\mathcal{A}=1\}$ and $\{\mathcal{A}=1\}\cup\{\mathcal{A}=0\}$ respectively. Consider the observable adjusted odds-ratio $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)$, which equals

$$\frac{\mathbb{P}(\mathcal{R}=1|X=1,\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)/\mathbb{P}(\mathcal{R}=0|X=1,\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)}{\mathbb{P}(\mathcal{R}=1|X=0,\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)/\mathbb{P}(\mathcal{R}=0|X=0,\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)}.$$

We will say $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)$ is externally valid if it equals the adjusted causal odds-ratio in the target population $COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w})$. We will say it is internally valid if it equals the adjusted causal odds-ratio in the source population, $COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$, with

$$COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)=\frac{\mathbb{P}(\mathcal{R}_1=1|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)/\mathbb{P}(\mathcal{R}_1=0|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)}{\mathbb{P}(\mathcal{R}_0=1|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)/\mathbb{P}(\mathcal{R}_0=0|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)}.$$

According to the results stated above, $COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w})$ is not recoverable. In particular, $COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w})$ is not equal to $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)$, which is therefore not externally valid. Note that in case (i), $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)$ is externally valid.

In order to study whether $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)$ is internally valid, let us consider the case where $\{\mathcal{S}=\mathcal{A}\}$, for simplicity. Then the question reduces to assessing whether $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)=COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$. This equality holds if $\mathbb{P}(\mathcal{R}=1|X=x,\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)=\mathbb{P}(\mathcal{R}_x=1|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$, which is guaranteed if $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W},\mathcal{A})$. But conditioning on $\mathcal{A}$ breaks the ignorability condition. Indeed, $\boldsymbol{W}$ satisfies the back-door criterion and $\mathcal{R}_x \perp\!\!\!\perp X|\boldsymbol{W}$, but $(\boldsymbol{W},\mathcal{A})$ does not satisfy the back-door criterion and we generally do not have $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W},\mathcal{A})$; see Appendix D for a simple proof by a reduction to absurdity argument. As a result, $\mathbb{P}(\mathcal{R}=1|X=x,\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$ generally differs from $\mathbb{P}(\mathcal{R}_x=1|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$, and $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$ is not internally valid. In other words, even in the most favorable situation where $\{\mathcal{S}=\mathcal{A}\}$, the observable adjusted odds-ratio $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{S}=1)=OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$ can not be interpreted as the causal odds-ratio in the source population $COR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$. As mentioned in [24], $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$ "does not stand for a causal effect in a stable subset of individuals", because $\mathcal{A}$ is a descendant of X. More precisely, using the fact that $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W},\mathcal{A}_x)$, it can be shown (See Appendix E) that the adjusted odds-ratio conditioned on $\mathcal{A}=1$ is

$$OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)=\frac{\mathbb{P}(\mathcal{R}_1=1|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}_1=1)/\mathbb{P}(\mathcal{R}_1=0|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}_1=1)}{\mathbb{P}(\mathcal{R}_0=1|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}_0=1)/\mathbb{P}(\mathcal{R}_0=0|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}_0=1)}.$$

Because subsets $\{\mathcal{A}_0=1,\boldsymbol{W}=\boldsymbol{w}\}$ and $\{\mathcal{A}_1=1,\boldsymbol{W}=\boldsymbol{w}\}$ correspond to different subsets of individuals, $OR(X,\mathcal{R}|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}=1)$ has to be interpreted with caution.

This issue is not restricted to responsibility analyses. It is known that selection on the outcome, as well as selection on a variable affected by the outcome, can lead to bias ([30]). Moreover, selectionning on a descendant of the exposure in the DAG is known as a more general phenomenon called collider bias ([5, 29]), that is for instance at play in the obesity paradox explained by [31, 32].

4 Numerical illustration

Here, we present simple numerical results in the absence of confounders to check the absence of bias in case (i), and to illustrate the magnitude of bias in case (ii). A more thorough numerical analysis will be considered in future research. The generative model described below is in the most general case, the case (ii), where we consider $\mathcal{A}$ and $\mathcal{R}$. Case (i) will be illustrated by setting some parameters to zero in this generative model.

4.1 Data generation mechanism

We consider a generative model consistent with the right panel of Figure 3. More precisely, our data generation mechanism is obtained by specifying the structural functions f_F, $f_{\mathcal{A}}$, f_X, and f_V as well as the distributions of the disturbances ε_F, $\varepsilon_{\mathcal{A}}$, ε_X, and ε_V.

Denote the indicator function by $\mathbb{I}[\cdot]$. Define four independent random variables ε_F, $\varepsilon_{\mathcal{A}}$, ε_X, and ε_V distributed according to a uniform distribution over the interval $[0, 1]$. For any given $p_X \in (0, 1)$, define $X = \mathbb{I}[\varepsilon_X \leq p_X]$ so that $X \sim B(p_X)$ is a Bernoulli variable. We consider the special case where $p_X = 0.5$. Now, introduce the sigmoid function $\mathrm{h}(x) = (1 + \exp(-x))^{-1}$, and set, for any $(x, v, f) \in \{0, 1\}^3$ and for some real parameters $\alpha_0, \alpha_X, \beta_0, \beta_X, \beta_V, \gamma_0, \gamma_F, \gamma_V$.

$$
\begin{aligned}
p_V(x) &= \mathrm{h}(\alpha_0 + \alpha_X x) \\
p_F(x, v) &= \mathrm{h}(\beta_0 + \beta_X x + \beta_V v) \\
p_{\mathcal{A}}(f, v) &= \mathrm{h}(\gamma_0 + \gamma_F f + \gamma_V v).
\end{aligned}
$$

Finally, variables V, F, $\mathcal{A}$, $\mathcal{R}$ are defined as

$$
\begin{aligned}
V &= f_V(X, \varepsilon_V) = \mathbb{I}[\varepsilon_V \leq p_V(X)], \\
F &= f_F(X, V, \varepsilon_F) = \mathbb{I}[\varepsilon_F \leq p_F(X, V)], \\
\mathcal{A} &= f_{\mathcal{A}}(F, V, \varepsilon_{\mathcal{A}}) = \mathbb{I}[\varepsilon_{\mathcal{A}} \leq p_{\mathcal{A}}(F, V)], \\
\mathcal{R} &= F \times \mathcal{A}.
\end{aligned}
$$

Moreover, for $\nu \geq 0$, we set

$$\begin{aligned}
\alpha_0 &= -\frac{1}{2}\alpha_X \\
\beta_0 &= -\frac{1}{2}\left(\beta_X + \beta_V - \nu\right) \\
\gamma_0 &= -\frac{1}{2}\left(\gamma_F + \gamma_V - \nu\right)
\end{aligned}$$

Then, we compute causal effects and the measure of association on the odds-ratio scale defined below :

$$\begin{aligned}
COR(X, F|\boldsymbol{W} = \boldsymbol{w}) &= \frac{\mathbb{P}(F_1 = 1|W = w)/\mathbb{P}(F_1 = 0|W = w)}{\mathbb{P}(F_0 = 1|W = w)/\mathbb{P}(F_0 = 0|W = w)}, \\
COR(X, \mathcal{R}|W = w) &= \frac{\mathbb{P}(\mathcal{R}_1 = 1|W = w)/\mathbb{P}(\mathcal{R}_1 = 0|W = w)}{\mathbb{P}(\mathcal{R}_0 = 1|W = w)/\mathbb{P}(\mathcal{R}_0 = 0|W = w)}, \\
OR(X, F|W = w, \mathcal{A} = 1) &= \frac{\mathbb{P}(F = 1|X = 1, W = w, \mathcal{A} = 1)/\mathbb{P}(F = 0|X = 1, W = w, \mathcal{A} = 1)}{\mathbb{P}(F = 1|X = 0, W = w, \mathcal{A} = 1)/\mathbb{P}(F = 0|X = 0, W = w, \mathcal{A} = 1)}.
\end{aligned}$$

Under our generative model, we get, for $x \in \{0, 1\}$,

$$\begin{aligned}
\mathbb{P}(F_x = 1|W = w) &= \mathbb{P}(F = 1|X = x, W = w) \\
&= \sum_{v \in (0,1)} \{\mathbb{P}(F = 1|X = x, V = v, W = w)\mathbb{P}(V = v|X = x, W = w)\} \\
&= \mathrm{h}(\beta_0 + \beta_X x + \beta_V)\mathrm{h}(\alpha_0 + \alpha_X x) + \mathrm{h}(\beta_0 + \beta_X x)\mathrm{h}(-(\alpha_0 + \alpha_X x))
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_x = 1|W = w) &= \mathbb{P}(\mathcal{A} = 1, F = 1|W = w) \\
&= \sum_{v \in (0,1)} \{\mathbb{P}(\mathcal{A} = 1|F = 1, V = v, W = w)\mathbb{P}(F = 1|X = x, V = v, W = w) \\
&\qquad \mathbb{P}(V = v|X = x, W = w)\} \\
&= \mathrm{h}(\gamma_0 + \gamma_F + \gamma_V)\mathrm{h}(\beta_0 + \beta_X x + \beta_V)\mathrm{h}(\alpha_0 + \alpha_X x) + \\
&\quad \mathrm{h}(\gamma_0 + \gamma_F)\mathrm{h}(\beta_0 + \beta_X x)\mathrm{h}(-(\alpha_0 + \alpha_X x))
\end{aligned}$$

$$
\begin{aligned}
&\mathbb{P}(F=1|X=x,W=w,\mathcal{A}=1)\\
&=\frac{\mathbb{P}(F=1,X=x,W=w,\mathcal{A}=1)}{\mathbb{P}(X=x,W=w,\mathcal{A}=1)}\\
&=\frac{\mathbb{P}(\mathcal{A}=1|F=1,X=x,W=w)\mathbb{P}(F=1X=x,W=w)}{\mathbb{P}(\mathcal{A}=1|X=x,W=w)\mathbb{P}(X=x|W=w)\mathbb{P}(W=w)}\\
&=\frac{\sum\limits_{v\in(0,1)}\{\mathbb{P}(\mathcal{A}=1|F=1,V=v,W=w)\mathbb{P}(F=1|X=x,V=v,W=w)\mathbb{P}(V=v|X=x,W=w)\}}{\sum\limits_{v,f\in(0,1)}\{\mathbb{P}(\mathcal{A}=1|F=f,V=v,W=w)\mathbb{P}(F=1|X=x,V=v,W=w)\mathbb{P}(V=v|X=x,W=w)\}}\\
&=\mathrm{h}(\gamma_0+\gamma_F+\gamma_V)\mathrm{h}(\beta_0+\beta_X x+\beta_V)\mathrm{h}(\alpha_0+\alpha_X x)+\mathrm{h}(\gamma_0+\gamma_F)\mathrm{h}(\beta_0+\beta_X x)\mathrm{h}(-(\alpha_0+\alpha_X x))\times\\
&\quad[\mathrm{h}(\gamma_0+\gamma_F+\gamma_V)\mathrm{h}(\beta_0+\beta_X x+\beta_V)\mathrm{h}(\alpha_0+\alpha_X x)+\mathrm{h}(\gamma_0+\gamma_F)\mathrm{h}(\beta_0+\beta_X x)\mathrm{h}(-(\alpha_0+\alpha_X x))+\\
&\quad\mathrm{h}(\gamma_0+\gamma_V)\mathrm{h}(-(\beta_0+\beta_X x+\beta_V))\mathrm{h}(\alpha_0+\alpha_X x)+\mathrm{h}(\gamma_0)\mathrm{h}(-(\beta_0+\beta_X x))\mathrm{h}(-(\alpha_0+\alpha_X x))]^{-1}
\end{aligned}
$$

4.2 Results

We present results under configurations where $\beta_V = 1$ and $\gamma_F = 4$, because speed V increases the risk of comitting a fault and because F largely increases the risk of having an accident. To choose $\nu = 13$ leads to have prevalences of A, R, and F inferior than $10^{-6}\%$, which can be considered as realistic. Then, we make the tree remaining parameters, α_X, β_X and γ_V, vary between 0 and 3, because alcohol X increases speed V and the risk of committing a fault F, and V increases the severity of the accident. Results are presented in Figure 4.

First, in the situation where $\gamma_V = 0$, which mimics case (i), our results confirm that there is no bias between $COR(X, F|W = w)$ and $OR(X, F|W = w, A = 1)$, whatever the value of α_X or β_X. It can also be seen that $COR(X, R|W = w) \simeq COR(X, F|W = w)$, so that $COR(X, R|W = w)$ can be well-approximated by $COR(X, F|W = w)$.

In the situation where $\gamma_V \neq 0$, we observe bias between $COR(X, F|W = w)$ and $OR(X, F|W = w, \mathcal{A} = 1)$ as soon as $\alpha_X \neq 0$, and it increases with α_X or β_X. So, in the case where X increases V, $OR(X, F|W = w, \mathcal{A} = 1)$ underestimates the real causal effect $COR(X, F|W = w)$. We can also observe that $COR(X, R|W = w)$ cannot be approximated by $COR(X, F|W = w)$ anymore. Finally, the higher the value of γ_V, the higher the magnitude of the bias. Note that there is also no bias when $\alpha_X = 0$, whatever the value of γ_V, which corresponds to the absence of arrow between X and V in case (ii).

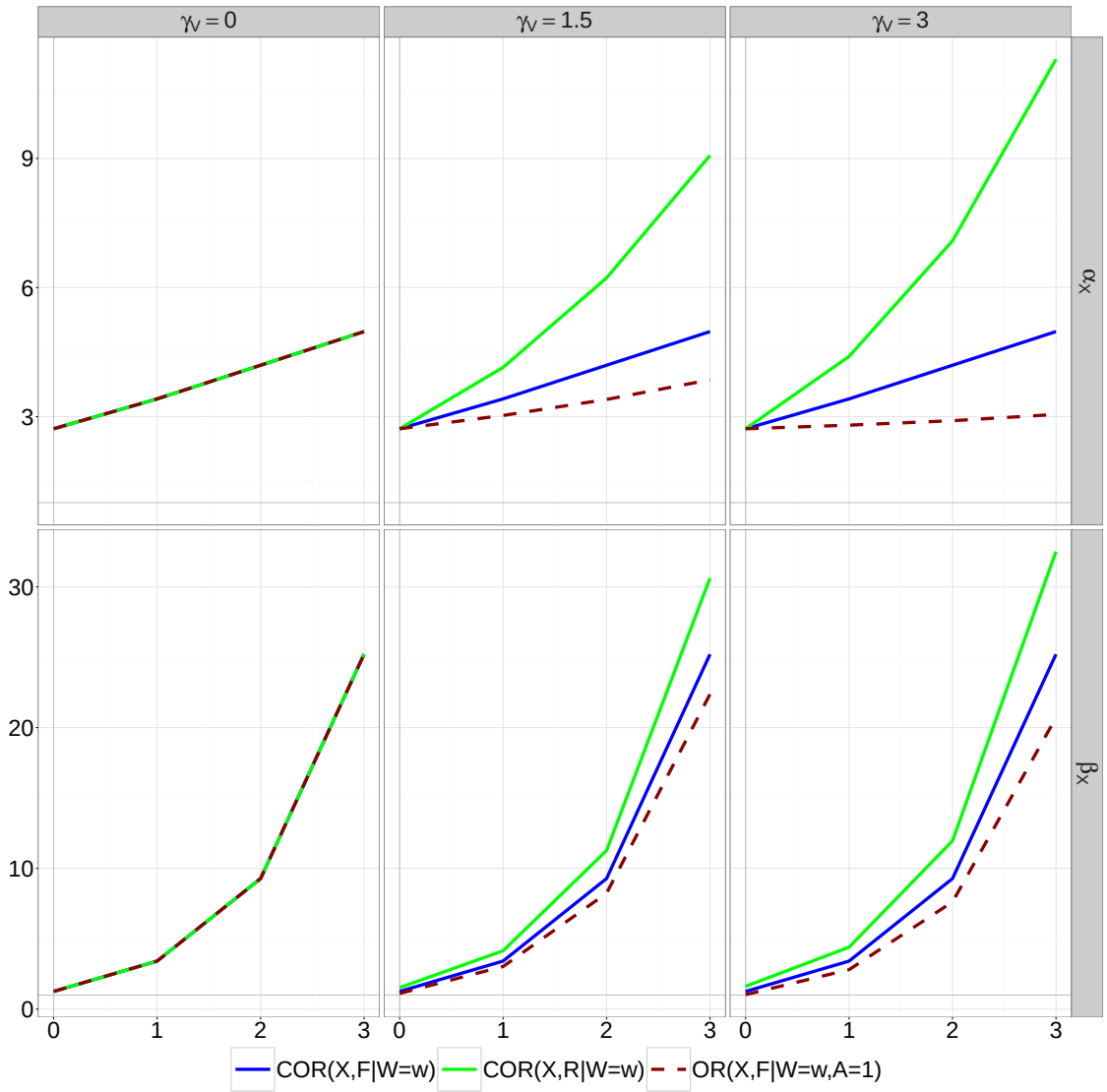


Figure 4: Causal and associational odds-ratios in the case where $\beta_V = 1$ and $\gamma_F = 4$, and for varying values of the other parameters $\alpha_X, \beta_X, \gamma_V$. In each panel, along the x axis, α_X or β_X are varied from 0 to 3, and the other parameter is set to 1.

To recap, even under the simple generative model considered here, we can observe that the observable associational effect $OR(X, F|W = w, A = 1)$ unbiasedly estimates $COR(X, F|W = w)$ when $\gamma_V = 0$ or $\alpha_X = 0$. Otherwise, $OR(X, F|W = w, \mathcal{A} = 1)$ underestimates the real causal effect under the settings considered here.

5 Discussion

In this article, we lay the emphasis on the conditions ensuring the recoverability of causal effects in the presence of selection bias. We apply these conditions to provide a formal description of responsibility analyses, which is commonly used in the field of road safety epidemiology.

Focusing on an exposure X defined as blood alcohol concentration above the legal limit, and its effect on R or $\mathcal{R}$, we show that $\boldsymbol{w}$-specific causal odds-ratios $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ is not recoverable *(i)* when data from all types of accident are available and *(ii)* when data from the most severe accidents only are available. Under some additional assumptions, $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ can be approximated by $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ in case *(i)*. However, case *(i)* is not the most common situation in practice: even when data for property damaged accidents are available, the least severe ones are generally not reported. And in this case, $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$ is not recoverable and can not be approximated. We use numerical examples to illustrate our arguments. In particular, we show that, in case (ii), the observable associational effect $OR(X, F|W = w, \mathcal{A} = 1)$ underestimates causal effects $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ or $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$ in the situations where X and V are positively correlated.

As a matter of fact, available controls in case *(ii)* are not representative of the non-responsible drivers ($\mathcal{R} = 0$). Indeed, this population is composed by three types of drivers: the ones for whom $(F = 0, \mathcal{A} = 0)$, the ones for whom $(F = 0, \mathcal{A} = 1)$, and the ones for whom $(F = 1, \mathcal{A} = 0)$. However, because data describes drivers involved in a crash only, the control group in case (ii) is only composed of drivers who did not commit a driving fault but had a severe accident $(F = 0, \mathcal{A} = 1)$. There is no reason why this subgroup of non-responsible drivers should be similar to the two other subgroups. In particular, drivers involved in a severe crash are likely to drive faster than those who are not involved in an accident, since crash severity is partly caused by speed. Because inappropriate speed is related to alcohol, the subpopulations $(F = 0, \mathcal{A} = 0)$ and $(F = 0, \mathcal{A} = 1)$ are different regarding exposure X. Some authors have sensed this non-representativeness issue and have proposed a heuristic transformation of the control group to make the final sample resemble that from a case-control study ([33]). However, there is no way to assess whether the final sample is indeed representative

without additional information regarding $\mathbb{P}(X)$ for instance. This type of additional information could also make adjusted causal odds-ratio $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$, causal odds-ratio $COR(X, \mathcal{R})$ or even $\mathbb{P}(\mathcal{R}_x = r)$, for $(x, r) \in \{0, 1\}$, recoverable from the available data, even if the structure of the DAG alone does not make these quantities recoverable. One interesting lead for future research would be to use our formalisation of responsibility analyses to determine which additional information would be sufficient to recover causal effects such as $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$. Note that if we choose an exposure X, which has no effect on V, $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ is recoverable and $COR(X, \mathcal{R}|\boldsymbol{W} = \boldsymbol{w})$ can be approximated.

In other respect, selection bias is not the only issue in responsibility analyses. First, for drivers involved in a crash, responsibility is usually determined from police reports, from an algorithm ([34, 11]), or by experts. This assigned responsibility is a noisy measure of the true responsibility. In particular, when responsibility is assigned by experts, a safe rule would be to remove all information regarding X from police reports so that experts are blinded to exposure status ([12]). Otherwise, it is possible that experts use the information regarding X when assigning drivers' responsibility, which is particularly problematic ([16]). Second, unmeasured confounders are sources of confounding bias. For instance, risk proneness is likely to be a common cause of alcohol consumption and inappropriate speed, and is therefore a confounder regarding the causal relationship between X and A (or $\mathcal{A}$). It is generally unobserved and only loosely related to age, gender, etc, so that adjusting for these covariates is not sufficient and confounding bias arises.

Finally, when the $\boldsymbol{w}$-specific causal odds-ratio $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ is not recoverable or can not be approximated, we can only estimate quantities such as $OR(X, R|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1)$, where $\mathcal{A}$ denote the occurence of a severe accident. From a causal point of view, this quantity has to be interpreted carefully, since it equals $\frac{\mathbb{P}(\mathcal{R}_1=1|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}_1=1)/\mathbb{P}(\mathcal{R}_1=0|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}_1=1)}{\mathbb{P}(\mathcal{R}_0=1|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}_0=1)/\mathbb{P}(\mathcal{R}_0=0|\boldsymbol{W}=\boldsymbol{w},\mathcal{A}_0=1)}$, and subpopulations $\{\mathcal{A}_1 = 1, \boldsymbol{W} = \boldsymbol{w}\}$ and $\{\mathcal{A}_0 = 1, \boldsymbol{W} = \boldsymbol{w}\}$ can be quite different. This formal interpretation of the estimable quantity in the presence of selection bias is not only useful in the context of responsibility analyses. It is for instance useful to explain the obesity paradox ([32]).

References

[1] P. Cummings, F. P. Rivara, C. M. Olson, and K. Smith, "Changes in traffic crash mortality rates attributed to use of alcohol, or lack of a seat belt, air bag, motorcycle helmet, or bicycle helmet, united states, 1982–2001," *Injury Prevention*, vol. 12, no. 3, pp. 148–154, 2006.

[2] J. S. Mill, *A System of logic, ratiocinative and inductive, being a connected view of the principles of evidence, and the methods of scientific investigation.* Harper and Brothers, New York, eighth edition, 1843.

[3] M. Baiocchi, J. Cheng, and D. S. Small, "Instrumental variable methods for causal inference," *Statistics in Medicine*, vol. 33, pp. 2297–2340, June 2014. WOS:000335772800011.

[4] M. A. Hernán, S. Hernández-Díaz, M. M. Werler, and A. A. Mitchell, "Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology," *American Journal of Epidemiology*, vol. 155, no. 2, pp. 176–184, 2002.

[5] S. Greenland, "Quantifying biases in causal models: classical confounding vs collider-stratification bias," *Epidemiology*, vol. 14, no. 3, pp. 300–306, 2003.

[6] F. Elwert, "Graphical causal models," in *Handbook of causal analysis for social research*, pp. 245–273, Springer, 2013.

[7] F. Elwert and C. Winship, "Endogenous selection bias: the problem of conditioning on a collider variable," *Annual Review of Sociology*, vol. 40, 2014.

[8] N. Stamatiadis and J. A. Deacon, "Quasi-induced exposure: methodology and insight," *Accident Analysis & Prevention*, vol. 29, no. 1, pp. 37–52, 1997.

[9] H. Smith and R. Popham, "Blood alcohol levels in relation to driving.," *Canadian Medical Association Journal*, vol. 65, no. 4, pp. 325–328, 1951.

[10] K. Perchonok, "Identification of specific problems and countermeasures targets for reducing alcohol related casualties." Washington, DC: National Highway Traffic Safety Administration, U.S. Department of Transportation, 1978.

[11] K. Terhune, "Problems and methods in studying drug crash effects," *Alcohol, Drugs, and Driving*, vol. 2, no. 3-4, 1986.

[12] J. Brubacher, H. Chan, and M. Asbridge, "Culpability analysis is still a valuable technique," *International Journal of Epidemiology*, vol. 43, no. 1, pp. 270–272, 2014.

[13] A. E. Wahlberg and L. Dorn, "Culpable versus non-culpable traffic accidents; what is wrong with this picture?," *Journal of Safety Research*, vol. 38, no. 4, pp. 453–459, 2007.

[14] P. Sanghavi, "Commentary: Culpability analysis won't help us understand crash risk due to cell phones," *International Journal of Epidemiology*, vol. 42, no. 1, pp. 267–269, 2013.

[15] M. Asbridge, J. R. Brubacher, and H. Chan, "Cell phone use and traffic crash risk: a culpability analysis," *International Journal of Epidemiology*, vol. 42, no. 1, pp. 259–267, 2013.

[16] L. R. Salmi, L. Orriols, and E. Lagarde, "Comparing responsible and non-responsible drivers to assess determinants of road traffic collisions: time to standardise and revisit," *Injury Prevention*, vol. 20, pp. 380–386, Jan. 2014.

[17] A. E. Wahlberg, "The determination of fault in collisions," in *Driver Behaviour and Accident Research Methodology: Unresolved Problems.* (G. Dorn, L Matthews and I. Glendon, eds.), pp. 101–120, Farnham, Surrey, IK: Ashgate Publishing, 2009.

[18] J. Pearl, *Causality: models, reasoning, and inference.* Cambridge, U.K. ; New York: Cambridge University Press, 2000.

[19] J. Pearl, "Causal inference in statistics: An overview," *Statistics Surveys*, vol. 3, no. 0, pp. 96–146, 2009.

[20] J. Pearl, "Causal diagrams for empirical research," *Biometrika*, vol. 82, pp. 669–688, Jan. 1995.

[21] S. Greenland, J. Pearl, and J. M. Robins, "Causal Diagrams for Epidemiologic Research," *Epidemiology*, vol. 10, pp. 37–48, Jan. 1999.

[22] M. Glymour, "Causal diagrams. modern epidemiology. edited by: Rothman kj, greenland s, lash tl. 2008," 2008.

[23] E. Bareinboim and J. Tian, "Recovering causal effects from selection bias.," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence, AAAI*, pp. 3475–3481, 2015.

[24] E. Bareinboim and J. Pearl, "Controlling selection bias in causal inference," *Proceedings of The Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2012); JMLR*, vol. 22, pp. 100–108, 2012.

[25] M. A. Hernán, S. Hernández-Díaz, and J. M. Robins, "A structural approach to selection bias," *Epidemiology*, vol. 15, no. 5, pp. 615–625, 2004.

[26] T. J. VanderWeele and J. M. Robins, "Minimal sufficient causation and directed acyclic graphs," *The Annals of Statistics*, pp. 1437–1465, 2009.

[27] T. J. VanderWeele and I. Shpitser, "A new criterion for confounder selection," *Biometrics*, vol. 67, no. 4, pp. 1406–1413, 2011.

[28] V. Didelez, S. Kreiner, and N. Keiding, "Graphical models for inference under outcome-dependent sampling," *Statistical Science*, vol. 25, no. 3, pp. 368–387, 2010.

[29] K. J. Rothman, S. Greenland, and T. L. Lash, *Modern Epidemiology*. Lippincott Williams & Wilkins, 2008.

[30] J. Berkson, "Limitations of the application of fourfold table analysis to hospital data," *Biometrics Bulletin*, vol. 2, pp. 47–53, 1946.

[31] M. Lajous, A. Bijon, G. Fagherazzi, M.-C. Boutron-Ruault, B. Balkau, F. Clavel-Chapelon, and M. A. Hernán, "Body mass index, diabetes, and mortality in french women: explaining away a "paradox"," *Epidemiology (Cambridge, Mass.)*, vol. 25, no. 1, p. 10, 2014.

[32] V. Viallon and M. Dufournet, "Can collider bias fully explain the obesity paradox?," *Epidemiology*, 2017.

[33] B. Laumon, B. Gadegbeku, J.-L. Martin, and M.-B. Biecheler, "Cannabis intoxication and fatal road crashes in france: population based case-control study," *BMJ*, vol. 331, no. 7529, p. 1371, 2005.

[34] M. D. Robertson and O. H. Drummer, "Responsibility analysis: A methodology to study the effects of drugs in driving," *Accident Analysis & Prevention*, vol. 26, pp. 243–247, Apr. 1994.

[35] J. Splawa-Neyman, "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Statistical Science*, vol. 5, pp. 465–472, Nov. 1990.

[36] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.

[37] P. W. Holland, "Statistics and Causal Inference," *Journal of the American Statistical Association*, vol. 81, pp. 945–960, Dec. 1986.

[38] J. M. Robins, M. A. Hernan, and B. Brumback, "Marginal structural models and causal inference in epidemiology," *Epidemiology*, pp. 550–560, 2000.

[39] J. Robins, "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect," *Mathematical Modelling*, vol. 7, no. 9, pp. 1393–1512, 1986.

[40] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.

[41] K. L. Moore, R. Neugebauer, M. J. Laan, and I. B. Tager, "Causal inference in epidemiological studies with strong confounding," *Statistics in medicine*, vol. 31, no. 13, pp. 1380–1404, 2012.

[42] J. Benichou, "A review of adjusted estimators of attributable risk," *Statistical Methods in Medical Research*, vol. 10, no. 3, pp. 195–216, 2001.

[43] P. Bruzzi, S. Green, D. Byar, L. Brinton, and C. Schairer, "Estimating the population attributable risk for multiple risk factors using case-control data," *American Journal of Epidemiology*, vol. 122, no. 5, pp. 904–914, 1985.

[44] T. S. Richardson and J. M. Robins, "Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality," *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, vol. 128, no. 30, p. 2013, 2013.

Appendix A Causal inference in the Structural Causal Model framework

This Appendix presents the basis of causal inference, with a special emphasis on the SCM framework ([18]). We consider the same simple setting as that of Section 2 where cause X and effect Y are binary, and vector $\boldsymbol{W}$ is a set of categorical variables. A first useful tool to describe causal assumptions underlying a causal model $\mathcal{M}$ is the DAG. See the three examples considered in Figure 5. Under each of these examples, the DAG can be translated into a Structural Equation Model (SEM), that is a set of three structural equations, involving three autonomous functions $f_X, f_{\boldsymbol{W}}$ and f_Y, as well as three exogenous random variables $U_X, U_{\boldsymbol{W}}$ and U_Y, sometimes called disturbances ([18, 19]). We will assume that $U_X, U_{\boldsymbol{W}}$ and U_Y are jointly independent. These sets of structural equations govern the distributions of the variables $\boldsymbol{\mathcal{V}} = (X, \boldsymbol{W}, Y)$ observed in the actual world and available in observational studies. But they also allow the description of what we would have observed in counterfactual worlds, if we could have intervened and forced the exposure to be set to $X = x_0$ (with $x_0 \in \{0, 1\}$ in the simple scenario considered here). This physical intervention can be modeled by the *do*-operator introduced by [20, 18]. In model $\mathcal{M}$, this operator modifies the value of X: instead of being the result of the autonomous function f_X, it is set to $X = x_0$. The *do*-operator further removes any arrow pointing to X in the DAG. For instance, in Figure 5, DAG a' and DAG c' are identical while DAG a and DAG c are different: the two causal models are identical in the counterfactual world that we would observe after the intervention $do(X = x_0)$, but are different in the actual world.

Causal effects represent a general ability to transfer changes among covariates ([19]), and the *do*-operator is a key tool for their definition in the SCMs framework. The intervention $do(X = x_0)$ does not affect functions f_{V_1} (nor disturbances U_{V_1}) for $V_1 \in \boldsymbol{V} \setminus \{X\}$, but it does affect the distribution of descendants of X in the DAG (its children, the children of its children, etc.). More precisely, in the *counterfactual* world following the intervention $do(X = x_0)$, model $\mathcal{M}$ would be model $\mathcal{M}_{x_0}$ leading to observations of the variables $\boldsymbol{\mathcal{V}}(x_0) = (x_0, \boldsymbol{W}_{x_0}, Y_{x_0})$ instead of $\boldsymbol{\mathcal{V}} = (X, \boldsymbol{W}, Y)$. In case (a) for instance, we have $\boldsymbol{W}_{x_0} = f_{\boldsymbol{W}}(U_{\boldsymbol{W}}) = \boldsymbol{W}$ but $Y_{x_0} = f_Y(x_0, \boldsymbol{W}_{x_0}, U_Y) = f_Y(x_0, \boldsymbol{W}, U_Y)$, which is typically different from Y, unless $X = x_0$. The random variables $(Y_{x_0})_{x_0 \in \{0,1\}}$ are not (fully) observed: they are counterfactual variables, and correspond to potential outcomes ([35, 36, 37]). In the Marginal Structural Models (MSMs) framework ([38]), these counterfactual variables $(Y_{x_0})_{x_0 \in \{0,1\}}$ are connected to the variable Y observed in the actual world, through consistency constraints like the coherence assumption ([39]):

$$\text{(Coh)} \qquad (X = x) \Rightarrow (Y = Y_x), \quad \text{for all potential values } x \text{ of } X.$$

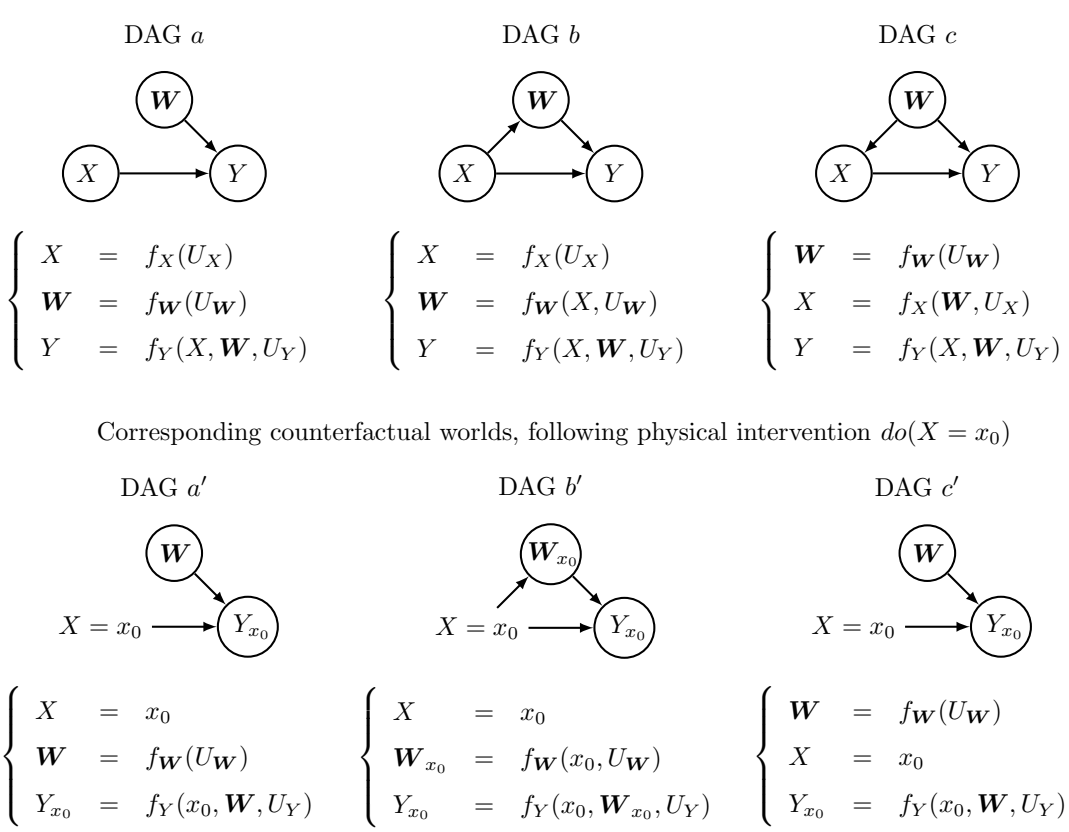


Figure 5: Examples of DAGs and SEMs: actual and counterfactual worlds.

In words, this assumption states that the actual outcome for an individual whose actual level of exposure is x equals the outcome we would observe for this same individual in the counterfactual world following $do(X = x)$, *i.e.*, where exposure would be "physically" set to value x. This assumption can be violated in randomized clinical trial if full adherence is not guaranteed for instance. In the simple case of binary exposure considered here, the coherence assumption entails that $Y = XY_1 + (1 - X)Y_0$. Under this coherence assumption, Y_1 and Y_0 are therefore partially observed and causal inference from observational data can be seen as a missing data problem. One first advantage of SCMs over MSMs rests in the precise definition of the counterfactual variables from the model, thanks to the DAG and the corresponding SEM. For instance, it is easy to show that the coherence assumption is automatically fulfilled in the structural interpretation of counterfactual variables considered here, since $Y = f_Y(X, \boldsymbol{W}, U_Y)$ and $Y_x = f_Y(x, \boldsymbol{W}, U_Y)$.

Once potential outcomes have been introduced (and precisely defined in the SCM framework),

causal effects can be precisely defined. In particular, in the simple example considered here where both X and Y are binary, the average causal effect of X on Y is defined as

$$\text{ACE} = \mathbb{E}(Y_1 - Y_0) = \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1). \tag{3}$$

In the SCM framework, $\mathbb{P}(Y_x = 1)$ is the probability that the outcome variable Y would equal 1 in the counterfactual world following the intervention $do(X = x)$: using Pearl's notations, we have $\mathbb{P}(Y_x = 1) = \mathbb{P}(Y = 1|do(X = x))$ (see, *e.g.*, Equation (7) in [19]). The average causal effect measures the difference between the risk of observing the outcome in the counterfactual world where all individuals would be exposed, $\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y = 1|do(X = 1))$, and the risk of observing the outcome in the counterfactual world where exposure would be eliminated, $\mathbb{P}(Y_0 = 1) = \mathbb{P}(Y = 1|do(X = 0))$. Here, causal effect is defined on the excess risk scale, but causal risk ratio or odds-ratio can of course be defined similarly: e.g., for the causal risk ratio, it is defined as $\mathbb{P}(Y_1 = 1)/\mathbb{P}(Y_0 = 1)$.

So far, we have shown that the introduction of the *do*-operator and the counterfactual variables Y_1 and Y_0 allow a precise definition of causal effects. However, because such causal effects rely on quantities that are not (fully) observed in the actual world, a natural question arises whether these effects can be estimated from observational data.

Consider a general DAG G composed of observed variables $\boldsymbol{\mathcal{V}}$, with $(X, Y) \in \boldsymbol{\mathcal{V}}$. In the SCMs framework, $\mathbb{P}(Y = y|do(X = x)) = \mathbb{P}(Y_x = y)$ is said to be identifiable if the assumptions embedded in DAG G ensure that this quantity is expressible in terms of the observable distribution $\mathbb{P}(\boldsymbol{\mathcal{V}} = \boldsymbol{v})$; see Definition 1 in [23] for instance.

Here we start by recalling some standard conditions introduced in the literature, that are sufficient for identifiability of $\mathbb{P}(Y = y|do(X = x)) = \mathbb{P}(Y_x = y)$, when combined with the coherence assumption (Coh). These conditions ensure a certain level of "comparability" between exposed and non-exposed individuals in the actual world. In particular, the ignorability assumption reads as follows ([21]):

$$\text{(Ign)} \qquad Y_x \perp\!\!\!\perp X, \quad \text{for all potential values } x \text{ of } X,$$

where $V_1 \perp\!\!\!\perp V_2$ stands for "V_1 and V_2 are independent". This assumption states that the outcome Y_x we would observe in the counterfactual world where exposure would be set to value x is independent of the actual level of exposure X. Under (Ign) and (Coh), the causal effect of X can be estimated from observations of the actual world, as long as $0 < \mathbb{P}(X = 1) < 1$. Indeed, by successively applying assumption (Ign) and (Coh), it follows that

$$\mathbb{P}(Y_x = 1) = \mathbb{P}(Y_x = 1|X = x) = \mathbb{P}(Y = 1|X = x) \tag{4}$$

Under (Coh) and (Ign) the association measure between X and Y is said to be unconfounded and equals the causation measure. For instance, under (Coh) and (Ign), $\mathbb{P}(Y = 1|X = 1) - \mathbb{P}(Y = 1|X = 0) = \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1)$. Of course, (Ign) is a strong assumption and, as will be made clearer below, it does not hold in the presence of confounders for instance. In such cases, a conditional version of (Ign) can also been considered ([40]). In particular, assume that, for some variable $\boldsymbol{W}$, the following assumption holds:

$$\text{(C.Ign)} \qquad Y_x \perp\!\!\!\perp X|\boldsymbol{W}, \quad \text{for all potential values } x \text{ of } X,$$

where $V_1 \perp\!\!\!\perp V_2|V_3$ stands for "V_1 and V_2 are conditionally independent given V_3". Then the so-called adjustment formula, or back-door formula ([20]), holds too:

$$\begin{aligned}\mathbb{P}(Y_x = 1) &= \sum_{\boldsymbol{w}} \mathbb{P}(Y_x = 1|\boldsymbol{W} = \boldsymbol{w})\mathbb{P}(\boldsymbol{W} = \boldsymbol{w}) \\ &= \sum_{\boldsymbol{w}} \mathbb{P}(Y_x = 1|X = x, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(\boldsymbol{W} = \boldsymbol{w}) \\ &= \sum_{\boldsymbol{w}} \mathbb{P}(Y = 1|X = x, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(\boldsymbol{W} = \boldsymbol{w}).\end{aligned}$$

Therefore, under (Coh) and (C.Ign), we have

$$\text{ACE} = \sum_{\boldsymbol{w}} \{\mathbb{P}(Y = 1|X = 1, \boldsymbol{W} = \boldsymbol{w}) - \mathbb{P}(Y = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})\}\mathbb{P}(\boldsymbol{W} = \boldsymbol{w}), \tag{5}$$

and ACE corresponds to a marginalized version of the adjusted excess risk. Under the Experimental Treatment Assignment (ETA) assumption, that is $0 < \mathbb{P}(X = x|\boldsymbol{W}) < 1$ almost everywhere, ACE can be estimated from observations of the actual world. Methods adapted to situations where the ETA assumption fails to hold have also been proposed (see [41] for a review and further extensions).

Appendix B Proof of approximation (2)

Under the assumption of no unmeasured confounder, $OR(X, F|\boldsymbol{W} = \boldsymbol{w}) = COR(X, F|\boldsymbol{W} = \boldsymbol{w})$. Moreover, if $\mathbb{P}(F_x = 1|\boldsymbol{W} = \boldsymbol{w})$ is small for $x \in \{0, 1\}$, then $COR(X, F|\boldsymbol{W} = \boldsymbol{w}) \approx CRR(X, F|\boldsymbol{W} = \boldsymbol{w})$, while if $\mathbb{P}(R_x = 1|\boldsymbol{W} = \boldsymbol{w})$ is small for $x \in \{0, 1\}$, then $COR(X, R|\boldsymbol{W} = \boldsymbol{w}) \approx CRR(X, R|\boldsymbol{W} = \boldsymbol{w})$. Therefore, under these assumptions, and if $X \perp\!\!\!\perp A|(F, \boldsymbol{W})$, we have

$$
\begin{aligned}
COR(X, R|\boldsymbol{W} = \boldsymbol{w}) &\approx CRR(X, R|\boldsymbol{W} = \boldsymbol{w}) \\
&= \frac{\mathbb{P}(R_1 = 1|\boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(R_0 = 1|\boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F_1 = 1, A_1 = 1|\boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F_0 = 1, A_0 = 1|\boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F_1 = 1, A_1 = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F_0 = 1, A_0 = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F = 1, A = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F = 1, A = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(A = 1|F = 1, X = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(A = 1|F = 1, X = 0, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(A = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(A = 1|F = 1, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})} \\
&= \frac{\mathbb{P}(F_1 = 1|\boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(F_0 = 1|\boldsymbol{W} = \boldsymbol{w})} \\
&= CRR(X, F|\boldsymbol{W} = \boldsymbol{w}) \\
&\approx COR(X, F|\boldsymbol{W} = \boldsymbol{w})
\end{aligned}
$$

Appendix C Population attributable fraction

The PAF ([42, 43]) is usually defined as

$$
\begin{aligned}
\text{PAF} = \frac{\mathbb{P}(R = 1) - \mathbb{P}(R_0 = 1)}{\mathbb{P}(R = 1)} &= \frac{\mathbb{P}(R = 1) - \sum_{\boldsymbol{w}} \mathbb{P}(R = 1|X = 0, \boldsymbol{W} = \boldsymbol{w})\mathbb{P}(\boldsymbol{W} = \boldsymbol{w})}{\mathbb{P}(R = 1)} \\
&= 1 - \sum_{i=0}^{1} \sum_{\boldsymbol{w}} \frac{p_{i,\boldsymbol{w}}}{RR_{i\boldsymbol{w}}} \\
&= \mathbb{P}(X = 1|R = 1) - \sum_{\boldsymbol{w}} \frac{p_{1\boldsymbol{w}}}{RR_{1\boldsymbol{w}}}. \qquad (6)
\end{aligned}
$$

with $p_{i,\boldsymbol{w}} = \mathbb{P}(X = i, \boldsymbol{W} = \boldsymbol{w}|R = 1)$ and $RR_{i\boldsymbol{w}} = \frac{\mathbb{P}(R=1|X=i, \boldsymbol{W}=\boldsymbol{w})}{\mathbb{P}(R=1|X=0, \boldsymbol{W}=\boldsymbol{w})}$. Because $R = 1 \Rightarrow A = 1$, we have $p_{1,\boldsymbol{w}} = \mathbb{P}(X = 1, \boldsymbol{W} = \boldsymbol{w}|R = 1) = \mathbb{P}(X = 1, \boldsymbol{W} = \boldsymbol{w}|R = 1, A = 1)$: $p_{1,\boldsymbol{w}}$ and $\mathbb{P}(X = 1|R = 1)$

can therefore unbiasedly estimated from data describing the subpopulation $\{A = 1\}$ only. On the other hand, the $\boldsymbol{w}$-specific adjusted relative risks $RR_{1\boldsymbol{w}}$ can be approximated by the $\boldsymbol{w}$-specific adjusted odds-ratios $OR(X, R|\boldsymbol{W} = \boldsymbol{w})$ provided $\mathbb{P}(R = 1|X = 1, \boldsymbol{W} = \boldsymbol{w})$ is small for $x \in \{0, 1\}$ and $\boldsymbol{w}$. Therefore, under this assumption, all the quantities in (6) can be unbiasedly estimated or well approximated in case (i) and the PAF can therefore be approximately unbiasedly estimated in case (i).

Appendix D Why is there no guarantee that $\mathcal{R}_x \perp\!\!\!\perp X|(\mathbf{W}, \mathcal{S})$ in case (ii)?

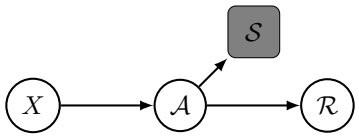


Figure 6: A special case of the DAG describing case (ii)

Consider the simplified DAG of Figure 6 and further assume that $\mathcal{S} = \mathcal{A}$. Because this DAG is a special case of the DAG of the right panel of Figure 3 describing our case (ii), it is enough to show that $\mathcal{R}_x \perp\!\!\!\perp X|\mathcal{S}$ does not hold in this special case to ensure that there is no guarantee that $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{S})$ under the DAG of Figure 3. Under the model depicted by the DAG of Figure 6, we have $\mathcal{R} \perp\!\!\!\perp X|\mathcal{S}$ of course, but $\mathcal{R}_x \perp\!\!\!\perp X|\mathcal{S}$ is not guaranteed. This result can be shown using graph-based arguments but it can also easily be shown by a reduction to absurdity argument. Indeed, if $\mathcal{R}_x \perp\!\!\!\perp X|\mathcal{S}$ then we have

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_x = 1) &= \sum_s \mathbb{P}(\mathcal{R}_x = 1|\mathcal{S} = s)\mathbb{P}(\mathcal{S} = s) \\
&= \sum_s \mathbb{P}(\mathcal{R}_x = 1|\mathcal{S} = s, X = x)\mathbb{P}(\mathcal{S} = s) \quad \text{if } \mathcal{R}_x \perp\!\!\!\perp X|S \\
&= \sum_s \mathbb{P}(\mathcal{R} = 1|\mathcal{S} = s, X = x)\mathbb{P}(\mathcal{S} = s) \quad \text{under the coherence assumption} \\
&= \sum_s \mathbb{P}(\mathcal{R} = 1|\mathcal{S} = s)\mathbb{P}(\mathcal{S} = s) \quad \text{because } \mathcal{R} \perp\!\!\!\perp X|\mathcal{A} \text{ and } \mathcal{A} = \mathcal{S} \\
&= \mathbb{P}(\mathcal{R} = 1).
\end{aligned}$$

Therefore, if $\mathcal{R}_x \perp\!\!\!\perp X|\mathcal{S}$ we especially obtain $ACE = \mathbb{P}(\mathcal{R}_1 = 1) - \mathbb{P}(\mathcal{R}_0 = 1) = 0$, which is generally false under the DAG of Figure 6.

Appendix E Proof of independence $\mathcal{R}_x \perp\!\!\!\perp X|(\mathbf{W}, \mathcal{A}_x)$ by the SWIG Approach

Denote by $\mathcal{R}_x$, F_x, V_x, $\mathcal{A}_x$ the counterfactual outcome variables we would observe in the counterfactual world following the intervention $X = x$.

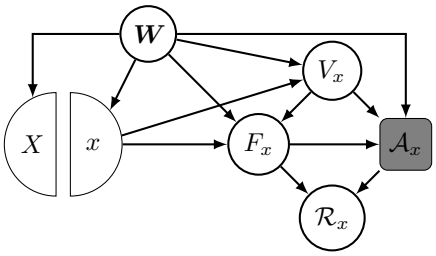


Figure 7: The SWIT resulting from the intervention $do(X = x)$ in case *(ii)*

The fact that $\mathcal{R}_x \perp\!\!\!\perp X|(W, \mathcal{A}_x)$ can be shown using the SWIG approach ([44]). Figure 7 presents the SWIT corresponding to the intervention $do(X = x)$. From this representation, it directly follows that $\mathcal{R}_x \perp\!\!\!\perp X|W$, and therefore that $\mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A}_x)$. Then, the following holds

$$\begin{aligned}\mathbb{P}(\mathcal{R} = 1|X = x, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A} = 1) &= \mathbb{P}(\mathcal{R}_x = 1|X = x, \boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_x = 1) \quad \text{by consistency} \\ &= \mathbb{P}(\mathcal{R}_x = 1|\boldsymbol{W} = \boldsymbol{w}, \mathcal{A}_x = 1) \quad \text{since } \mathcal{R}_x \perp\!\!\!\perp X|(\boldsymbol{W}, \mathcal{A}_x). \qquad (7)\end{aligned}$$

.4 Poster présenté à UK-Causal inference meeting

Causal inference and selection bias: The case of responsibility analyses in road safety epidemiology

M. Dufournet [1], E. Lanoy [2], JL. Martin [1], V.Viallon [1]

[1] University of Lyon, F-69622, Lyon, France; University Lyon 1, UMRESTTE, F-69373, Lyon, France; IFSTTAR, UMRESTTE, F-69675, Bron, France.

[2] Institut Gustave Roussy, Villejuif, France.

marine.dufournet@ifsttar.fr

Introduction

– road accidents : major public health issue (1,24 million of deaths in the world [1]
– most causes are known (alcohol, etc) ⇒ how to measure their causal effects?
– only data for drivers involved in an accident is available: extreme selection bias! (no data for the control group)
⇒ standard to perform *responsibility analyses* [2, 3, 4]:
◦ assessment of the responsibility, or culpability, of each driver involved in crashes
◦ comparison of responsible drivers/vehicles with non-responsible drivers/vehicles.
⇒ non-responsible drivers (control group) supposed to represent a random sample of the general driving population ⇒ *responsibility analyses* ≃ case-control study.
≠ but is it really the case ?

Our Objective: To formalize which causal effects can be estimated through responsibility analysis, depending on the available data.
For instance, we consider the estimation of causal effects of blood alcohol concentration (BAC).

Principles of causal inference

Notations

– Y, a dichotomous outcome, X a binary exposure of interest, and $\boldsymbol{W}$, a set of other categorical factors, composed the DAG G. All observed variables composed G denoted by $\boldsymbol{\mathcal{V}}$.
If X is a direct cause of Y, in the SCM framework, $Y = f_Y(X, ..., u_Y)$ and the potential outcome or the conterfactual are defined by $Y_x = f_Y(x, ..., u_Y)$.

– S, a binary variable indicating inclusion in the study, and a new DAG G_s
– $OR(X, Y|\boldsymbol{W})$, the conditional odds-ratio where $\boldsymbol{W}$ is the vector of confounders.
– $COR(X, Y|\boldsymbol{W} = \boldsymbol{w})$, the $\boldsymbol{w}$-specific causal odds-ratio

Definitions of causal effects and identifiability

In the simplest case, the average causal effect is defined by (1). And under coherence and ignorability assumptions, it is given by (2).

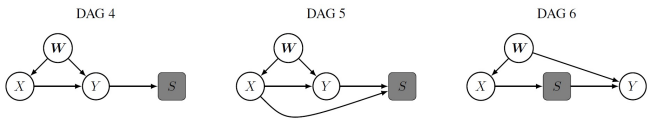
$$\begin{aligned} \text{ACE} = \mathbb{E}(Y_1 - Y_0) &= \mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1) && (1) \\ &= \mathbb{P}(Y = 1|X = 1) - \mathbb{P}(Y = 1|X = 0) && (2) \end{aligned}$$

In the presence of confounders W, and if $Y_x \perp\!\!\!\perp X|\boldsymbol{W}$, the ACE is given by:

$$\text{ACE} = \sum_{\boldsymbol{w}} \{\mathbb{P}(\text{Y} = 1|\text{X} = 1, \boldsymbol{W} = \boldsymbol{w}) - \mathbb{P}(\text{Y} = 1|\text{X} = 0, \boldsymbol{W} = \boldsymbol{w})\}\mathbb{P}(\boldsymbol{W} = \boldsymbol{w})$$

Recoverability of causal effects in the presence of selection bias

Figure 1: Exemples of DAGs in the presence of selection bias



$\mathbb{P}(Y_x = y)$ recoverable from selection biased data when the assumptions embedded in G_s renders it expressible in terms of the distribution under selection bias $\mathbb{P}(\boldsymbol{\mathcal{V}} = \boldsymbol{v}|S = 1)$ [5].

• Recoverability of $\mathbb{P}(Y_x = y)$ in the presence of selection bias [5]
For any $\boldsymbol{C} \subseteq \boldsymbol{\mathcal{V}}$ define $G_{\boldsymbol{C}}$ the subgraph of DAG G composed only of variables in $\boldsymbol{C}$. For any $V_i \in \boldsymbol{\mathcal{V}}$, further denote by $An(V_i)_G$ the union of V_i and the ancestors of V_i the DAG G. $\mathbb{P}(Y_x = y)$ is recoverable from selection biased data if and only if

$$An(Y)_{G_{\boldsymbol{\mathcal{V}}\setminus X}} \cap An(S)_{G_s} = \emptyset.$$

⟶ Theorem violated in DAGs of Figure 2 because of Y is an ancestor of S or S is an ancestor of Y.

• Recoverability of causal odds-ratios in the presence of selection bias [6]

– $OR(X, Y|\boldsymbol{W})$ recoverable iff $X \perp\!\!\!\perp S|(Y, \boldsymbol{W})$ or $Y \perp\!\!\!\perp S|(X, \boldsymbol{W})$.
⟶ condition fulfilled under DAG 4, but violated under DAGs 5 and 6.

• Identifiability of other causal effects in the presence of selection bias

– Is $OR(X, Y|\boldsymbol{W} = \boldsymbol{w}, S = 1)$ internally valid under DAGs 5 and 6 ?
– No ! because $\mathbb{P}(Y = 1|X = x, \boldsymbol{W} = \boldsymbol{w}, S = 1) = \mathbb{P}(Y_x = 1|S_x = 1, \boldsymbol{W} = \boldsymbol{w})$
– illustration by obesity paradox [7] : $\mathbb{P}(Y_1 = 1|\boldsymbol{W} = \boldsymbol{w}, S_1 = 1)/\mathbb{P}(Y_0 = 1|\boldsymbol{W} = \boldsymbol{w}, S_0 = 1)$ estimable but difficult to interpret

Application to responsibility analyses

– X, binary variable indicating whether a driver has a BAC above the legal limit.
– A, binary variable, indicating whether this same driver is involved in a accident.
⇒ selection bias because $A = 1$ for all observations

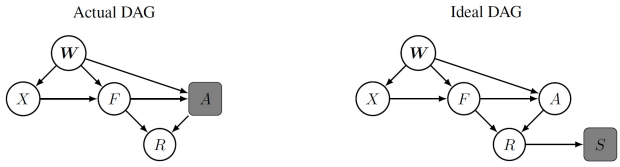
Two possible outcomes in responsibility analysis : fault (F) or responsibility (R)

– F, binary variable indicating whether the driver commits a fault (lane departure, failure to obey traffic signs, driving against traffic, etc.).
⟶ defined for all drivers, not only those involved in a crash.
– R, binary variable indicating whether the driver is responsible for an accident.
⇒ $R = F \times A$. $R = 1$ if $A = 1$ and $F = 1$ ($R = 0$ if $A = 0$ even if $F = 1$).
⇒ neither R nor F is an ideal outcome when assessing causal effects of risk factors in road safety: e.g. $F = 1 \not\Rightarrow A = 1$, $R = 0 \not\Rightarrow A = 0$ and $R = 0 \not\Rightarrow F = 0$.

Responsibility analyses when data from all types of accidents available

⟶ left panel of Figure 2.

Figure 2: Actual and ideal DAGs in responsibility analyses with data from all accidents

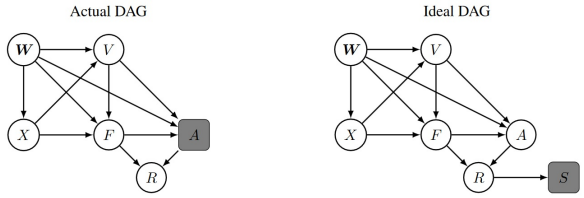


– $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ recoverable since $X \perp\!\!\!\perp A|(F, \boldsymbol{W})$
– $COR(X, R|\boldsymbol{W} = \boldsymbol{w})$ not recoverable since $X \not\perp\!\!\!\perp A|(R, \boldsymbol{W})$ and $R \not\perp\!\!\!\perp A|(X, \boldsymbol{W})$
But $COR(X, R|\boldsymbol{W} = \boldsymbol{w}) \simeq COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ if $\mathbb{P}(R_x = 1|\boldsymbol{W} = \boldsymbol{w})$ and $\mathbb{P}(F_x = 1|\boldsymbol{W} = \boldsymbol{w})$ are small

Responsibility analyses when data from severe accidents only is available

⟶ effect of X not entirely mediated by F because of speed V.

Figure 3: Actual and ideal DAGs in responsibility analyses with data from severe accidents accidents



– none of causal quantities is directly recoverable, except quantities like $\mathbb{P}(R_1 = 1|\boldsymbol{W} = \boldsymbol{w}, A_1 = 1)/\mathbb{P}(R_0 = 1|\boldsymbol{W} = \boldsymbol{w}, A_0 = 1)$.
– because $X \not\perp\!\!\!\perp A|(F, \boldsymbol{W})$: X is a cause of speed V and V has an effect of A in addition to the one "mediated" by F.

Discussion-Conclusion

– we provide a formal description of responsibility analysis.
– $\boldsymbol{w}$-specific causal odds-ratios $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ recoverable if data from all types of accidents are available.
– none of causal quantities recoverable when data from severe accidents only is available (the most common situation).

⇒ we propose two strategies to conduct the causal analysis:
– transformations of the original sample in order to make the final sample resemble that from a case-control study and compute sensitivity analysis
– to refrain from estimating causal effects of the form $COR(X, F|\boldsymbol{W} = \boldsymbol{w})$ and limit ourselves to quantities like $\mathbb{P}(R_1 = 1|\boldsymbol{W} = \boldsymbol{w}, A_1 = 1)/\mathbb{P}(R_0 = 1|\boldsymbol{W} = \boldsymbol{w}, A_0 = 1)$.

References

[1] World Health Organization (2013). Rapport de situation sur la securite routiere dans le monde, p12.

[2] Perchonok, K. (1978). Identification of specific problems and countermeasures targets for reducing alcohol related casualties. Washington, DC: National Highway Traffic Safety Administration, US Department of Transportation

[3] Terhune, K. (1986). Problems and methods in studying drug crash effects. *Alcohol, Drugs, and Driving*, 2(3-4).

[4] Brubacher, J., Chan, H., and Asbridge, M. (2014). Culpability analysis is still a valuable technique. *International Journal of Epidemiology*, **43**(1): 270-272

[5] Bareinboim, E. and Tian, J. (2015). Recovering causal effects from selection bias. *In AAAI*, pages 3475-3481

[6] Bareinboim, E. and Pearl, J. (2012). Controlling selection bias in causal inference. *Proceedings of the fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2012*, JMLR(22): 100-108

[7] Lajous, M., Bijon, A., Fagherazzi, G., Boutron-Ruault, M-C., Balkau, B., Clavel-Chapelon, F., and Hernan, M.A. (2014). Body mass index, diabetes, and mortality in french women: explaining away a "paradox". *Epidemiology (Cambridge, Mass*, **25**(1):10

[8] Pearl, J. (2000). *Causality, models, reasoning. International Journal of Epidemiology*, **43**(1): 270-272

.5 Lettre de recherche publiée dans Epidemiology

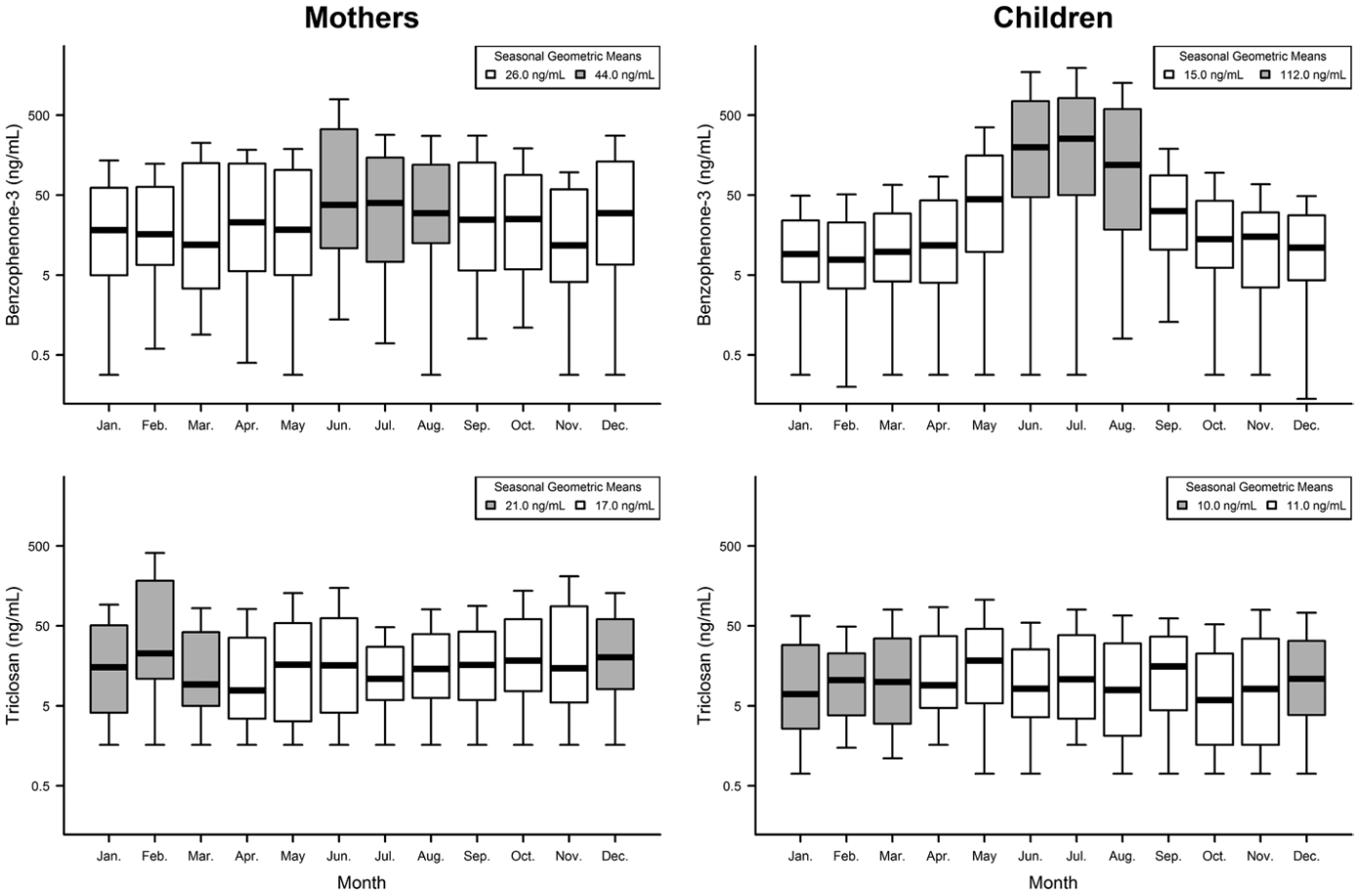


FIGURE. Maternal and childhood urinary benzophenone-3 and triclosan concentrations by month of collection among mothers and children in the HOME Study (2003–2006).

Megan E. Romano
Department of Epidemiology
Geisel School of Medicine at Dartmouth
Lebanon, New Hampshire
Department of Epidemiology
Brown University School of Public Health
Providence, Rhode Island
Megan.E.Romano@Dartmouth.edu

Geetika Kalloo
Taylor Etzel
Joseph M. Braun
Department of Epidemiology
Brown University School of Public Health
Providence, Rhode Island

REFERENCES

1. Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman J. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics*. 2009;21:606–631.
2. Smith MN, Workman T, McDonald KM, et al. Seasonal and occupational trends of five organophosphate pesticides in house dust. *J Expo Sci Environ Epidemiol*. 2016;27:372–378.
3. Weiss L, Arbuckle TE, Fisher M, et al. Temporal variability and sources of triclosan exposure in pregnancy. *Int J Hyg Environ Health*. 2015;218:507–513.
4. Calafat AM, Wong LY, Ye X, Reidy JA, Needham LL. Concentrations of the sunscreen agent benzophenone-3 in residents of the United States: National Health and Nutrition Examination Survey 2003–2004. *Environ Health Perspect*. 2008;116:893–897.
5. Food and Drug Administration HHS. Safety and effectiveness of consumer antiseptics; topical antimicrobial drug products for over-the-counter human use (81 FR 42911). *Fed Regist*. 2016;81:42911–42937.
6. Braun JM, Kalloo G, Chen A, et al. Cohort profile: The Health Outcomes and Measures of the Environment (HOME) study. *Int J Epidemiol*. 2017;46:24.
7. Ye X, Kuklenyik Z, Needham LL, Calafat AM. Automated on-line column-switching HPLC-MS/MS method with peak focusing for the determination of nine environmental phenols in urine. *Anal Chem*. 2005;77:5407–5413.
8. Zamoiski RD, Cahoon EK, Michal Freedman D, Linet MS. Self-reported sunscreen use and urinary benzophenone-3 concentrations in the United States: NHANES 2003-2006 and 2009-2012. *Environ Res*. 2015;142:563–567.

Re: Collider Bias Is Only a Partial Explanation for the Obesity Paradox

To the Editor:

Recently, Sperrin et al[1] claim that collider bias is unlikely to be the main explanation for the "obesity paradox." However, the authors neglected the fact that $Y^{A=a} \not\perp A \mid (M, U)$ in their causal model and their conclusions are question-

The authors report no conflicts of interest.
SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Copyright © 2017 Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 1044-3983/17/2805-0e43
DOI: 10.1097/EDE.0000000000000691

Copyright © 2017 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

able. In particular, the bias they report for $\Delta_{AS}-\Delta_{CE}$ sensibly underestimates its actual value, and collider bias can fully explain the obesity paradox under the simple generative model they consider.

Sperrin et al[1] consider the causal system depicted in Figure 1, where A, M, Y, and U correspond to obesity, diabetes status, early death, and some typically unobserved binary confounder, respectively. See Section 1 in the eAppendix (http://links.lww.com/EDE/B216) for details. They first introduce the association measure:

$$\begin{aligned}\Delta_{AS}&=E[Y|M=m,A=a]-E[Y|M=m,A=a^*]\\&=\sum_u\left\{\begin{array}{l}E[Y|M=m,A=a,U=u]\\P[U=u|M=m,A=a]\\-E[Y|M=m,A=a^*,U=u]\\P[U=u|M=m,A=a^*]\end{array}\right\}.\end{aligned}\quad(1)$$

Several epidemiologic studies reported a negative estimate for Δ_{AS}, suggesting that obesity was associated with a decreased mortality among diabetic patients, hence the obesity paradox. However, Δ_{AS} has to be related to a causal effect to support this paradox. Because the focus is on diabetic patients, i.e., patients with $M=m$, Sperrin et al[1] consider the causal effect of A on Y, conditioned on M being at level m, Δ_{CE}, with

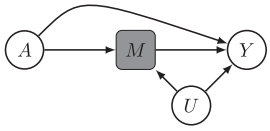


FIGURE 1. The directed acyclic graph considered in Sperrin et al.[1] Here, *A*, *M*, *Y*, and *U* correspond to obesity, diabetes status, early death, and some typically unobserved binary confounder, respectively.

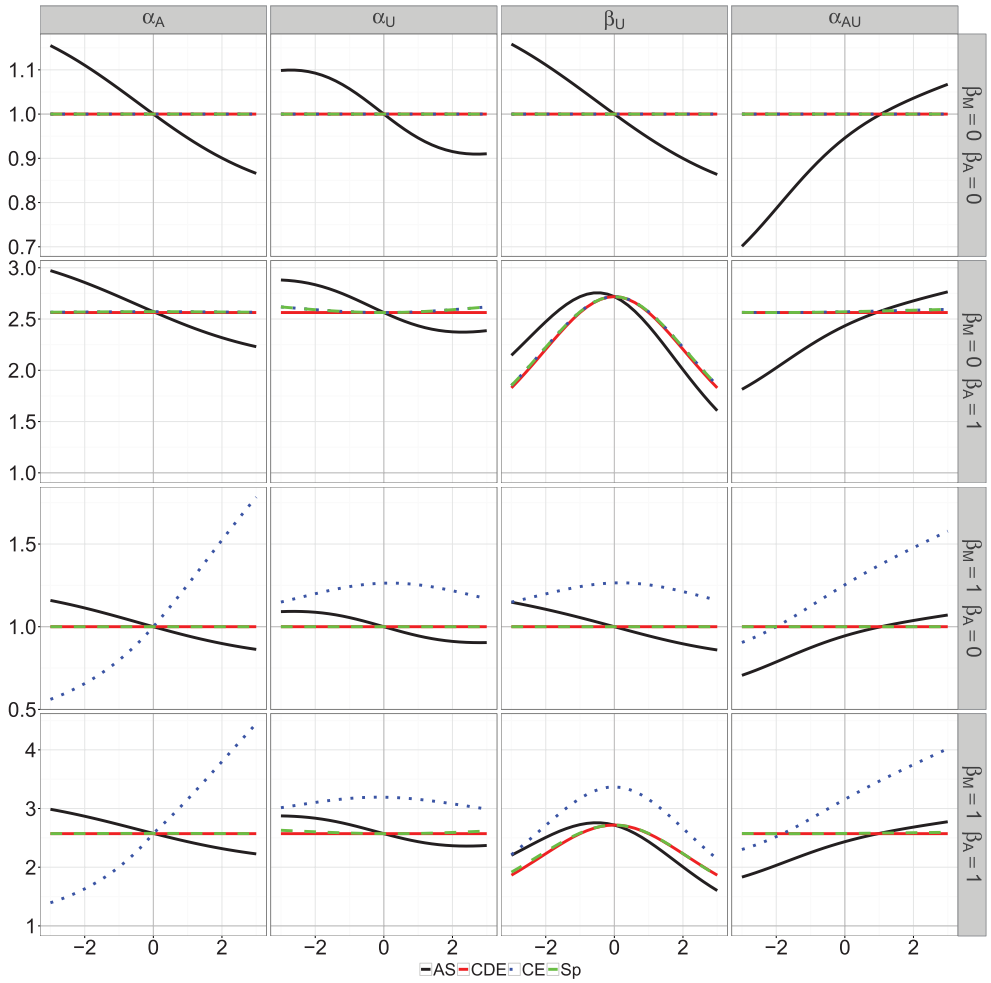


FIGURE 2. Results from the numerical study under the generative causal model introduced by Sperrin et al[1] (see Section 3.1 in the eAppendix; http://links.lww.com/EDE/B216). Causal and "associational" odds ratios are presented in the case where (β_A=0, β_M=0), (β_A=1, β_M=0), (β_A=0, β_M=0), and (β_A=0, β_M=0), and for varying values of the other parameters α_A, α_U, β_U and α_{AU}. In each column, along the *x* axis, one of these parameters is varied from −3 to 3 (left column: α_A, mid-left column: α_U, mid-right column: β_U , and right column: α_{AU}), and the other parameters are set to a default value (1 for α_A, α_U, β_U and 0 for α_{AU}). The first two rows exactly correspond to the configurations considered by Sperrin et al[1] in Figures 2 and 4 of their article, respectively. It is noteworthy that when β_M=0, then *M* is not on a causal path between *A* and *Y*, and there is no collider bias. However, considering for instance the case where β_M=1, *M* is on a causal path between *A* and *Y* and collider bias is at play, whatever the value of β_A: in this case, Δ_{AS} can be severely biased compared to Δ_{CE}.

© 2017 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2017 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

$$\Delta_{CE} = E\left[Y^{A=a} \middle| M=m\right] - E\left[Y^{A=a^*} \middle| M=m\right].$$

The authors incorrectly establish that Δ_{CE} equals

$$\Delta_{Sp} = \sum_u \{E[Y|M=m, A=a, U=u] - E[Y|M=m, A=a^*, U=u]\} P[U=u|M=m],$$

after claiming, on the left column of page 526, that

$$\begin{aligned} E\left[Y^{A=a} \middle| M=m\right] &= \sum_u E\left[Y^{A=a} \middle| M=m, U=u\right] P[U=u|M=m] \\ &= \sum_u E[Y|M=m, A=a, U=u] P[U=u|M=m]. \end{aligned} \quad (2)$$

But equation (2) is only guaranteed if $E\left[Y^{A=a} \middle| M=m, U=u\right] = E[Y|M=m, A=a, U=u]$ which, on turn, is only guaranteed if $Y^{A=a} \perp A | (M, U)$. However, this conditional independence does generally not hold under the model depicted in Figure 1: M being on a causal path between A and Y, the set (M, U) does not satisfy the back-door criterion; see 2, 3 as well as Fine point 7.2 in 4 where the authors use the single world intervention graph approach[5] in a related causal model and Section 4.2 in the eAppendix (http://links.lww.com/EDE/B216). As a result, Δ_{Sp} is generally different from Δ_{CE} under the causal model of Figure 1. The reason for this difference is precisely collider bias, which makes $E\left[Y^{A=a} \middle| M=m, U=u\right]$ and $E[Y|M=m, A=a, U=u]$ different. See Section 4 in the eAppendix (http://links.lww.com/EDE/B216) for more insights on the difference between Δ_{Sp} and Δ_{CE}.

Actually, Δ_{CE} can not be expressed in terms of the distribution of the variables (A, M, U, Y) without further assumptions on the causal model. By specifying the structural functions and the distributions of the disturbances in the generative model considered by Sperrin et al,[1] an analytic formula for Δ_{CE} can be derived and the bias $\Delta_{AS} - \Delta_{CE}$ can be evaluated (Section 3.2 in the eAppendix; http://links.lww.com/EDE/B216). As shown in Figure 2, this bias can be severe under this model, which invalidates the conclusions by Sperrin et al.[1]

Moreover, the premise that we are interested in Δ_{CE} is actually questionable, and the controlled direct effect of A on Y at $M = m$, $\Delta_{CDE} = E\left[Y^{A=a, M=m}\right] - E\left[Y^{A=a^*, M=m}\right]$, may be better suited[6,7] (Sections 2 and 5 in the eAppendix; http://links.lww.com/EDE/B216). Under the model of Figure 1, it can be shown that

$$\Delta_{CDE} = \sum_u \{E[Y|A=a, M=m, U=u] - E[Y|A=a^*, M=m, U=u]\} P(U=u). \quad (3)$$

Comparing equations (2) and (3), the difference between Δ_{AS} and Δ_{CDE} is only due to the discrepancy between P ($U = u$) and P ($U = u|M = m, A = a$). In the absence of the confounder U, Δ_{AS} and Δ_{CDE} are equal. The bias between Δ_{AS} and Δ_{CDE} can then be interpreted as a confounding bias and is generally limited under the generative model considered by Sperrin et al[1] (Figure 2). Note, however, that we can still have a negative association measure $\Delta_{AS} < 0$ and a positive controlled direct effect $\Delta_{CDE} > 0$, as illustrated in eFigure 3 (http://links.lww.com/EDE/B216) after considering additional interaction terms in this model.

To recap, contrary to what Sperrin et al[1] claim, collider bias is likely to make the bias between Δ_{AS} and Δ_{CE} substantial and can fully explain the obesity paradox. The bias between Δ_{AS} and Δ_{CE} is only due to confounding and is likely to be weaker.

Vivian Viallon
Marine Dufournet
Univ Lyon, Université Lyon 1
IFSTTAR, UMRESTTE
F-69373 Lyon, France
vivian.viallon@univ-lyon1.fr

REFERENCES

1. Sperrin M, Candlish J, Badrick E, Renehan A, Buchan I. Collider bias is only a partial explanation for the obesity paradox. *Epidemiology*. 2016;27:525–530.
2. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge, United Kingdom; New York: Cambridge University Press; 2000.
3. Pearl J. An introduction to causal inference. *Int J Biostat*. 2010;6:Article 7.
4. Hernán MA, Robins JM. *Causal Inference*. Boca Raton, Florida: Chapman & Hall/CRC, 2017.
5. Richardson T, Robins JM. (2013) Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. Working Paper 128. Center for the Statistics and the Social Sciences, University of Washington, Seattle.
6. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15:309–334.
7. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172:1339–1348.

© 2017 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2017 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

.6 Réponse de Sperrin et al.

The Authors Respond

Sperrin, Matthew[a]; Candlish, Jane[a,b]; Badrick, Ellena[a]; Renehan, Andrew[c]; Buchan, Iain[a]

Epidemiology: September 2017 - Volume 28 - Issue 5 - p e46
doi: 10.1097/EDE.0000000000000692

Rejoinder

CONTENT NOT FOR REUSE

The Authors Respond

Matthew Sperrin,[a] Jane Candlish,[a,b] Ellena Badrick,[a] Andrew Renehan,[c] and Iain Buchan[a]

We thank Viallon and Dufournet[1] for their letter, and we agree with the concerns raised regarding our paper.[2] As the letter points out, the effect we calculated was not equal to Δ_{CE}, because it is "not" true that $Y^a \coprod A \mid (M, U)$. Thus, the bias in estimating Δ_{CE} may be much larger than we stated. Fortuitously, this has given rise to an interesting discussion about the various causal effects that can be calculated in this framework.

In fact, the effect we have calculated can be shown to be the controlled direct effect, conditioned on the level of M, $\Delta_{CDE\mid M=1}$; we give a proof in the eAppendix (http://links.lww.com/EDE/B217). This is similar to the controlled direct effect derived by Viallon and Durfournet[1] (the red line in their Figure 2), the difference being that we condition on those patients who had diabetes (and intervene to "maintain" their diabetes).

We agree with Viallon and Dufournet[1] that controlled direct effects are more likely to be the effects of interest in discussions around the obesity paradox; this has been noted by others as well.[3] In this setting, the effect of collider stratification is to induce a (or strengthen an existing[4]) relationship between U and A, and hence induce (or strengthen) confounding. Viallon and Dufournet[1] have also shown how the extent of the bias between controlled direct effects and observed associations can become larger than in the scenarios we considered, when interactions are present.

To conclude, we feel our results remain useful as an illustration of the extent of bias in using an observed association to estimate the "controlled direct effect," $\Delta_{CDE\mid M=1}$, under the scenarios we considered. Collider bias is one possible explanation for the obesity paradox, among many. It seems most likely to us that a combination of these explanations act together, possibly in synergy, to generate an observed obesity paradox.

REFERENCES

1. Viallon V, Dufournet M. Re. Collider bias is only a partial explanation for the obesity paradox. *Epidemiology*. 2017;28:e43–e45.
2. Sperrin M, Candlish J, Badrick E, Renehan A, Buchan I. Collider bias is only a partial explanation for the obesity paradox. *Epidemiology*. 2016;27:525–530.
3. Banack HR, Kaufman JS. Does selection bias explain the obesity paradox among individuals with cardiovascular disease? *Ann Epidemiol*. 2015;25:342–349.
4. Banack HR, Kaufman JS. From bad to worse: collider stratification amplifies confounding bias in the "obesity paradox." *Eur J Epidemiol*. 2015;30:1111–1114.

Editor's note: *A related Commentary appears on p. e47.*

From the [a]Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom; [b]School of Health Related Research, University of Sheffield, Sheffield, United Kingdom; and [c]Division of Cancer, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom.

Supported by the University of Manchester's Health eResearch Centre (HeRC) funded by the Medical Research Council Grant MR/K006665/1.
The authors report no conflicts of interest.
SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Matthew Sperrin, Health eResearch Centre, Farr Institute, University of Manchester, Manchester M13 9PL, UK. E-mail: matthew.sperrin@manchester.ac.uk.

Copyright © 2017 Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 1044-3983/17/2805-0e46
DOI: 10.1097/EDE.0000000000000692

Copyright © XXX Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.