

Année 3 - Semestre 6
2022/2023

UC-0324
**Communication professionnelle, scientifique
et interpersonnelle (CoPSI)**

Epidémiologie Clinique

**« Ce n'est pas parce qu'une association statistique est significative ($p \leq 0,05$)
qu'elle traduit une relation de cause à effet »**

Auteur : Pr Loïc Desquilbet
Version : Janvier 2023



INDEX

I.	Introduction.....	6
A.	Commentaires préliminaires	6
B.	Pré-requis	6
C.	Définitions fondamentales	7
1.	Définition de « maladie ».....	7
2.	Définition d'« exposition ».....	7
3.	Définition d'« exposition d'intérêt ».....	7
4.	Illustration des définitions d'exposition et de maladie sur un exemple	7
II.	Causalité, facteur de risque, et inférences	8
A.	La relation causale	8
B.	Facteur de risque et facteur protecteur.....	8
1.	Définition théorique.....	8
2.	Comment suggérer qu'une exposition semble être un facteur de risque (ou protecteur) d'une maladie ?	9
3.	Point d'attention sur le critère d'antériorité	9
C.	Association statistique versus association causale	10
D.	Inférence statistique versus inférence causale	11
1.	L'inférence statistique dans le cas d'une association	11
2.	L'inférence causale.....	11
3.	Signification du mot « réel » ou « réellement »	11
III.	Etudes épidémiologiques pour tenter d'établir une association causale.....	12
A.	Comment distinguer une étude interventionnelle d'une étude observationnelle.....	12
B.	Les études interventionnelles	12
1.	Caractéristiques communes de toutes les études interventionnelles (essais cliniques)	12
2.	L'essai clinique non randomisé	13
3.	L'essai clinique randomisé	13
4.	L'essai clinique en simple aveugle et double aveugle	14
C.	Les études observationnelles	14
1.	Caractéristiques communes de toutes les études observationnelles.....	14
2.	L'étude transversale.....	15
3.	L'étude cas-témoins	15
4.	L'étude de cohorte.....	15

IV.	Mesures d'association entre E et M.....	16
A.	Introduction.....	16
B.	Risque relatif (RR)	16
1.	Calcul d'un RR	16
2.	Interprétation de la valeur d'un RR dans l'échantillon	17
3.	Calcul de l'intervalle de confiance et interprétation	18
C.	Odds Ratio (OR)	18
1.	Calcul d'un OR	18
2.	Interprétation de la valeur d'un OR dans l'échantillon	19
3.	Calcul de l'intervalle de confiance et interprétation	19
D.	Largeur d'un intervalle de confiance à 95% et précision	20
E.	Tester un OR ou un RR.....	20
1.	Au sein de l'UC-0324.....	20
2.	Définition du terme « significativité »	21
3.	Différence avec l'UC-0322	21
F.	Association « positive, « négative », et « forte » : signification et illustration avec un OR ou un RR	22
V.	Généralité sur les biais d'association	25
A.	Remarques préliminaires	25
B.	Définition d'un biais d'association	25
C.	Présentation des biais d'association	25
D.	OR (ou RR) estimé <i>versus</i> OR (ou RR) causal	26
1.	Aspect théorique.....	26
2.	Illustration sur un exemple	26
VI.	Le biais de classement.....	26
A.	Origine du biais de classement et commentaires	26
B.	Evaluation objective ou subjective de l'exposition ou de la maladie étudiée	27
C.	Le biais de classement non différentiel.....	27
1.	Origine du biais de classement non différentiel	27
2.	Illustration sur un exemple	28
3.	Impact du biais de classement non différentiel sur la valeur de l'OR ou du RR....	28
4.	Impact du biais de classement non différentiel sur la significativité d'un OR ou d'un RR	28

D.	Le biais de classement différentiel	31
1.	Origine du biais de classement différentiel	31
2.	Illustration sur un exemple	31
3.	Essai clinique en double aveugle et biais de classement différentiel.....	32
4.	Impact du biais de classement différentiel sur la valeur de l'OR ou du RR.....	32
5.	Impact du biais de classement différentiel sur la significativité d'un OR ou d'un RR	32
E.	Discussion de l'importance des biais de classement.....	34
F.	Biais de classement différentiel ou non différentiel ?	34
1.	Dangerosité du biais de classement différentiel	34
2.	Logigramme pour évaluer la présence et le type de biais de classement.....	34
3.	Priorité au biais de classement différentiel	35
VII.	Le biais de confusion et le facteur de confusion.....	35
A.	Théorie.....	35
1.	Définition et origine du biais de confusion	35
2.	Définition et identification d'un facteur de confusion	35
B.	Biais de confusion et comparabilité clinique.....	36
1.	Rappel sur « l'égalité des indicateurs est synonyme d'absence d'association » ..	36
2.	De « égalité des indicateurs » à « comparabilité clinique »	36
3.	De « comparabilité clinique » à « absence d'association »	37
4.	De « comparabilité clinique » à « absence de biais de confusion »	37
C.	Vérification des deux critères pour identifier un facteur de confusion	38
D.	Sens du biais de confusion (« positif » ou « négatif »).....	39
E.	Commentaires	40
F.	Définition d'association brute	40
VIII.	Comment empêcher ou réduire les biais de confusion	41
A.	Introduction.....	41
B.	Empêcher ou réduire les biais de confusion grâce au protocole d'étude	42
1.	Attribution aléatoire (randomisation) de l'exposition d'intérêt	42
2.	Autres protocoles d'études empêchant ou réduisant les biais de confusion	44
C.	Empêcher ou réduire les biais de confusion grâce aux analyses statistiques : stratification et ajustement.....	45
1.	La stratification	45
2.	L'ajustement	47

D.	Introduction aux modèles de régression multivariés.....	48
1.	Quand utiliser un modèle de régression multivarié ?	49
2.	Ecriture d'un modèle de régression.....	49
3.	Trois modèles de régression	49
4.	Interprétation des résultats issus d'un modèle de régression univarié	50
5.	Interprétation des résultats issus d'un modèle de régression multivarié.....	50
6.	Tester statistiquement une association dans un modèle de régression.....	51
7.	Interprétation à partir d'un exemple issu de la littérature	51
8.	Quelles expositions faut-il inclure dans un modèle de régression multivarié ?....	51
IX.	Comment conclure à partir d'un OR ou un RR estimé dans un échantillon	52
A.	Démarche générale pour conclure à partir d'un OR ou d'un RR	52
B.	Différence entre inférence « possible » et inférence « faisable »	53
C.	Conclusion à partir d'un OR ou d'un RR brut	53
1.	OR ou RR brut significativement différent de 1 ($p \leq 0,05$)	53
2.	OR ou RR brut non significativement différent de 1 ($p > 0,05$)	54
3.	OR ou RR ajusté (sur une ou plusieurs expositions) significativement différent de 1 ($p \leq 0,05$).....	54
4.	OR ou RR ajusté (sur une ou plusieurs expositions) non significativement différent de 1 ($p > 0,05$).....	54
5.	Illustration sur un exemple issu d'un article.....	55
X.	Introduction à l'evidence-based veterinary medicine	55
A.	Définition et pratique de l' « evidence-based veterinary medicine ».....	55
B.	Escalier de preuve pour conclure au lien causal	57
XI.	Points-clés du polycopié d'épidémiologie clinique.....	58
XII.	Références.....	59
XIII.	Annexes	61
A.	Le biais de sélection.....	61
1.	Introduction	61
2.	Un échantillon « non représentatif » ne conduit pas forcément à du biais de sélection	62
B.	Illustration de l'impact d'un biais de classement non différentiel sur la force de l'association.....	62
C.	Démonstration de « l'égalité des indicateurs comparés est synonyme d'absence d'association statistique » (rappels de biostatistique)	65

I. INTRODUCTION

A. Commentaires préliminaires

Les concepts et outils qui vont être présentés dans ce polycopié font référence à l'épidémiologie analytique. Les définitions et les méthodes utilisées en épidémiologie analytique dépendent du contexte d'utilisation (sociologie, génétique, médecine d'élevage, recherche clinique, etc). Ce polycopié ne traite que l'utilisation de l'épidémiologie analytique dans un contexte de recherche clinique, dont l'objectif est le suivant (Broadbent, 2011) : « permettre d'acquérir des connaissances sur les liens causaux qui existent entre des expositions et des maladies. Ces liens sont souvent identifiés avant qu'ils ne soient compris : en général, l'épidémiologie [analytique] ne possède pas les outils pour expliquer comment une exposition peut causer une maladie, mais elle peut souvent fournir des preuves convaincantes que cette exposition cause effectivement cette maladie. »

B. Pré-requis

Pour être capable de suivre les enseignements du module « Epidémiologie clinique » de l'UC-0324, l'étudiant doit maîtriser les objectifs d'apprentissage (OA) qui ont été vus dans le module de Biostatistique en médecine vétérinaire de l'UC-0213 :

Définir l'inférence statistique ;

Interpréter l'intervalle de confiance à 95% d'une estimation (moyenne, pourcentage) en faisant de l'inférence statistique ;

Communiquer de façon pertinente à l'issue d'un résultat du test statistique non significatif ;

Appliquer la définition du risque d'erreur de première espèce (α) d'un test statistique significatif présenté dans une étude clinique ;

Citez dans une phrase les pourcentages qui vont être comparés puis testés, à partir d'effectifs fournis dans un tableau, dans l'objectif de tester l'association entre deux caractères binaires ;

Identifier la situation dans laquelle un message issu de la recherche clinique nécessite un groupe contrôle ;

Identifier la situation dans laquelle l'utilisation d'un test statistique (et ses conclusions qui vont avec) n'est pas pertinente ;

Conclure à partir des résultats d'un test statistique (incluant la valeur du degré de signification) comparant plusieurs (deux ou plus) moyennes (que les données soient appariées ou non), plusieurs médianes, et plusieurs pourcentages, en faisant entre autres de l'inférence statistique si elle est possible à faire.

C. Définitions fondamentales

1. Définition de « maladie »

Une **maladie** au sens épidémiologique du terme est toute « chose » relative à l'état de santé d'un animal dont on cherche à identifier les facteurs de risque ou protecteurs. Ce terme pourrait être remplacé par l'expression « critère de jugement » (« outcome » en anglais) pour couvrir davantage de situations en médecine vétérinaire (gestation, réforme, problèmes de comportements, effets indésirables, décès, ...), mais j'utiliserai le terme « maladie » pour des raisons de commodité pédagogique, qui sera noté « M » dans tout le polycopié.

Par ailleurs, je vais tout le temps considérer dans ce polycopié que M est un événement binaire. En pratique, M pourrait être un événement en 3 classes ou plus (par exemple, le stade IRIS d'une maladie rénale, qui va de 1 à 4), ou quantitatif (le rapport entre la taille de l'atrium gauche et celle de l'aorte, par exemple). Toute la théorie décrite dans ce polycopié pourra être appliquée à ces cas de figure, seule la pratique pourra différer (notamment le choix du modèle de régression).

2. Définition d'« exposition »

Une **exposition** au sens épidémiologique du terme est une caractéristique intrinsèque d'un animal (âge, sexe, race, concentration en urée, ...), ou extrinsèque (environnement, traitements reçus, ...), qui ne soit pas M, la « maladie » étudiée. Elle sera notée « E » ou « X » dans tout ce polycopié. Par ailleurs, je ne traiterai que des expositions binaires (animal exposé / animal non exposé). Mais tous les raisonnements énoncés sont valables pour des expositions quantitatives (par exemple, l'âge de l'animal), ou en 3 classes ou plus (par exemple, la race de l'animal).

3. Définition d'« exposition d'intérêt »

Une exposition d'*intérêt* est une exposition dont on cherche à savoir si elle est causalement associée à M. Une exposition peut être recueillie par les investigateurs d'une étude sans qu'elle ne soit néanmoins d'intérêt. Nous verrons dans la suite de ce polycopié qu'il peut en effet être nécessaire de prendre en compte dans les analyses statistiques une exposition A, qui ne soit pas d'intérêt, pour étudier la relation causale entre une exposition d'intérêt B et la survenue de M.

4. Illustration des définitions d'exposition et de maladie sur un exemple

Attention, il ne faut pas confondre « exposition » au sens médical du terme et « exposition » au sens épidémiologique du terme ; de même il ne faut pas confondre « maladie » au sens médical du terme avec « maladie » au sens épidémiologique du terme. En effet, une maladie au sens médical du terme peut être considérée, épidémiologiquement parlant dans une étude, comme une « exposition » (d'intérêt). L'exemple qui suit en est une illustration.

L'étude de Ferasin et coll. (Ferasin et al., 2013) avait pour objectif d'identifier les facteurs de risque de toux chez des chiens qui présentent une maladie valvulaire mitrale dégénérative (MMVD) (cf. **Figure 1**). Les auteurs de cette étude avaient observé que l'insuffisance cardiaque n'était pas significativement associée à la toux, contrairement à la présence d'une augmentation de la taille de l'atrium gauche. Ainsi, la « toux » était la « maladie » au sens épidémiologique du terme, et l'insuffisance cardiaque et l'augmentation de la taille du

ventricule gauche étaient deux « expositions » au sens épidémiologique du terme (bien qu'elles soient des « maladies » au sens médical du terme). Enfin, dans cette même étude, la MMVD n'était, au sens épidémiologique, ni une « maladie », ni une « exposition » (cf. définitions de « maladie » et d' « exposition » ci-dessus).

J Vet Intern Med 2013;27:286–292

Risk Factors for Coughing in Dogs with Naturally Acquired Myxomatous Mitral Valve Disease

L. Ferasin, L. Crews, D.S. Biller, K.E. Lamb, and M. Borgarelli

Background: Cough often is reported as the primary clinical sign of congestive heart failure (CHF) in dogs with chronic degenerative myxomatous mitral valve disease (MMVD). Concurrent airway disease and compression of the left mainstem bronchus by a large left atrium also have been proposed as potential causes of coughing in these patients.

Objectives: To investigate the association between the presence of coughing and different potential causes of cough, including CHF, abnormal radiographic airway pattern, and cardiomegaly in dogs affected by naturally acquired MMVD.

Animals: Two hundred six client-owned dogs.

Methods: Retrospective analysis performed on medical records of dogs affected by MMVD that underwent full cardiac evaluation, including echocardiographic examination and thoracic radiography.

Results: Univariate analyses showed that CHF is not a predictor of coughing (OR = 1.369; 0.723, 2.594), whereas abnormal radiographic airway pattern (OR = 3.650; 2.051, 6.496) and increased left atrial size observed radiographically (OR = 3.637; 1.904, 6.950) or echocardiographically (OR = 2.553; 1.436, 4.539) were significantly associated with coughing in dogs with MMVD. The same risk factors were significant in multivariate analyses.

Conclusions and Clinical Importance: This study indicates that CHF is not significantly associated with coughing in dogs with MMVD. Instead, abnormal radiographic airway pattern and left atrial enlargement are associated with coughing in these patients. This important finding should be taken into account when considering diagnosis and clinical management of CHF in these dogs.

Key words: Airway disease; CHF; MMVD; Pulmonary edema.

Figure 1. Abstract issu de l'article de Ferasin et coll, 2013.

II. CAUSALITE, FACTEUR DE RISQUE, ET INFERENCE

A. La relation causale

Il existe une relation causale (c'est-à-dire, une relation, ou un lien, de cause à effet) entre une exposition E et une maladie M si la présence de E a un impact sur la survenue de M (Parascandola and Weed, 2001). Autrement dit, E est causalement liée (ou « causalement associée ») à M si, toute chose égale par ailleurs, la seule présence de cette exposition E chez l'animal va conduire à augmenter ou à diminuer ses *risques* de survenue de M.

B. Facteur de risque et facteur protecteur

1. Définition théorique

Un **facteur de risque** ou un **facteur protecteur** de M est une exposition liée causalement à l'apparition de M. Un facteur de risque de M augmente les risques que M survienne, et un facteur protecteur diminue les risques que M survienne.

Un facteur de risque de M n'est ni une condition nécessaire ni une condition suffisante pour que M survienne. Un facteur protecteur de M n'est, de la même façon, ni une condition nécessaire ni une condition suffisante pour que ne M survienne pas.

En effet, chez l'Homme, la consommation de cigarettes est un facteur de risque de cancer du poumon, et pourtant, certaines personnes fumant un grand nombre de cigarettes par jour ne déclareront jamais de cancer du poumon, tandis que d'autres n'ayant jamais fumé au cours de leur vie déclareront un cancer du poumon.

2. Comment suggérer qu'une exposition semble être un facteur de risque (ou protecteur) d'une maladie ?

Pour suggérer qu'une exposition d'intérêt E *semble être* un facteur de risque (ou protecteur) de M dans la population à partir d'une étude clinique réalisée sur un échantillon, il faut que cette exposition remplisse les 3 conditions ci-dessous, appelée « règle des 3A »¹ :

- (1) E doit être **associée** significativement ($p \leq 0,05$) à M dans l'échantillon,
- (2) E, telle qu'elle a été collectée dans l'étude, doit être **antérieure** à la survenue M dans l'étude,
- (3) Il doit y avoir **absence** de biais sélection, de biais de classement différentiel, et de biais de confusion dans l'estimation de l'association entre E et M dans l'échantillon.

3. Point d'attention sur le critère d'antériorité

Quand j'ai écrit ci-dessus « E, telle qu'elle a été collectée dans l'étude, doit être **antérieure** à la survenue M dans l'étude », cela signifie que la valeur de l'exposition E qui a été collectée dans l'étude pour les analyses statistiques doit être celle *avant* que la maladie ne *surviene* chez l'animal. Vous savez au moins autant que moi qu'une maladie peut *survenir* sans être diagnostiquée, avec un diagnostic survenant des années plus tard. Si l'exposition est collectée après que la maladie est² survenue chez l'animal mais avant son diagnostic, le critère d'antériorité n'est plus vérifié (cf. **Figure 2**), et E ne devra pas être considérée comme un facteur de risque (ou protecteur) de M, même si les autres conditions de la règle des 3A sont vérifiées. En effet, dans une telle situation (association significative entre E et M et absence des trois biais cités ci-dessus), il est tout à fait possible que la valeur de E soit davantage une conséquence de M qu'un facteur de risque (ou protecteur) de M !

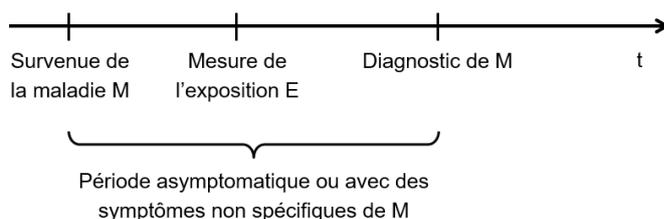


Figure 2. Situation où l'exposition E est collectée après la survenue de M mais avant son diagnostic. M est restée asymptomatique (ou bien l'animal présentait des symptômes non spécifiques de M) jusqu'à une certaine date, et à partir de la date d'apparition de symptômes spécifiques de M, le vétérinaire a posé le diagnostic de M. Dans cette situation, le critère d'antériorité n'est pas vérifié.

Evidemment, pour des expositions constantes au cours du temps (sexe, race, ...), E sera toujours antérieure à la survenue de M. Mais pour des expositions évoluant au cours du temps

¹ D'autres conditions existent, mais ce sera cette règle qui sera utilisée dans l'UC-0324.

² Oui, l'indicatif après un « après que » est dur à lire...

(alimentation, traitements, âge, biomarqueurs, ...), il faudra s'assurer de façon indispensable que le critère d'antériorité de la règle des 3A est bien vérifié.

C. Association statistique versus association causale

Il existe une « association statistique » entre E et M si M est plus (ou moins) fréquente parmi les animaux exposés (E+) que parmi les animaux non exposés (E-), quelle qu'en soit la raison, ou bien si M survient plus (ou moins) rapidement parmi les animaux E+ que parmi les animaux E-.

Il existe une « association causale » entre E et M s'il existe une relation de cause à effet entre E et M (E est alors un facteur de risque ou un facteur protecteur de M).

Dans tout le reste de ce polycopié, lorsque le mot « association » sera utilisé seul, sans le mot « causal » derrière « association », je parlerai d' « association statistique ».

Dans ce polycopié et lors de TD, nous allons très souvent utiliser des flèches pour relier des expositions entre elles, ou des expositions à une maladie, que ces flèches soient unidirectionnelles ou bidirectionnelles. Il est fondamental de correctement interpréter le sens de ces flèches.

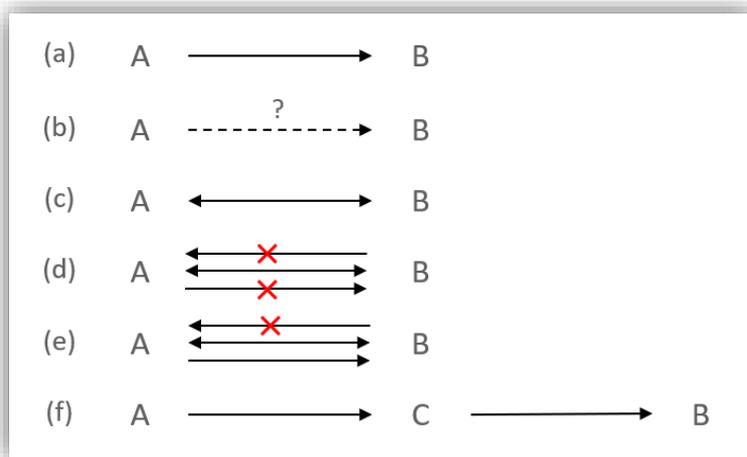


Figure 3. Interprétation du sens des flèches qui vont être utilisées durant tout le module d'épidémiologie clinique. A, B, et C sont des caractéristiques individuelles quelconques, liées ou non à un état de santé. Les interprétations des figures (a) à (e) sont énoncées dans le texte.

La flèche unidirectionnelle entre A et B dans la **Figure 3.a** signifie que A est causalement associée à B dans le sens où A est un facteur de risque, ou un facteur protecteur, de B (on peut aussi dire que « A est une cause de B »). La flèche unidirectionnelle entre A et B en pointillé avec un « ? » au-dessus dans la **Figure 3.b** signifie que l'on se pose la question de savoir si A est causalement associée à B dans le sens de la flèche de A vers B. La flèche bidirectionnelle entre A et B dans la **Figure 3.c** signifie que A et B sont associées statistiquement. La **Figure 3.d** indique que A et B sont associées statistiquement sans que A ne soit ni une cause, ni une conséquence, de B. Nous allons passer beaucoup de temps sur cette **Figure 3.d** en TD, car il n'est pas du tout intuitif que deux choses soient statistiquement associées sans que l'une ne soit ni une cause ni une conséquence de l'autre (et pourtant, il y a tellement de situations comme celle-là en médecine...). La **Figure 3.e** indique que A et B sont statistiquement associées, A pouvant être une cause de B (mais pas nécessairement), mais A n'étant pas une

conséquence de B. La **Figure 3.f** indique que A est une cause de C, et que C est une cause de B. Dans cette situation-là, on dit que C est un « facteur intermédiaire » dans la chaîne causale allant de A à B. La **Figure 3.f** explicite, en faisant intervenir C, la **Figure 3.a**.

D. Inférence statistique versus inférence causale

1. L'inférence statistique dans le cas d'une association

Rappelons la définition de l'inférence : l'inférence est le fait d'inférer, c'est-à-dire le fait d'étendre les résultats obtenus dans un échantillon à la population cible.

Si l'association entre E et M est significative ($p \leq 0,05$), alors l'inférence **statistique** consistera à dire (sous certaines conditions que nous verrons plus loin dans le polycopié) : « il y a des chances pour que, dans la population cible, M soit réellement plus (ou moins) fréquente parmi les animaux exposés à E que parmi les animaux non exposés à E » ou bien « il y a des chances pour que, dans la population cible, M surviennent réellement plus (ou moins) rapidement parmi les animaux exposés à E que parmi les animaux non exposés à E ».

2. L'inférence causale

Si l'association entre E et M est significative ($p \leq 0,05$), alors l'inférence **causale** consistera à dire (sous certaines conditions que nous verrons plus loin dans le polycopié) : « il y a des chances pour que, dans la population cible, E soit réellement un facteur de risque (ou protecteur) de M. »

3. Signification du mot « réel » ou « réellement »

Comme cela avait déjà été le cas en 2^{ème} année, le mot « réel » (ou « réellement ») fait référence à la population cible. Ainsi, écrire « il existe une association réelle » est équivalent à écrire « il existe une association *dans la population cible* ». Parfois, je renforcerai le message en écrivant « il existe une association réelle dans la population cible ». Ce que je viens d'écrire a une conséquence majeure : le mot « réel » ne fait pas référence au mot « causal ». Ainsi, écrire « il existe une association réelle entre la stérilisation de la chienne avant ses premières chaleurs et la survenue de tumeurs mammaires » *ne veut pas* dire que la stérilisation de la chienne avant ses premières chaleurs est causalement associée à la survenue de tumeurs mammaires ; cela veut dire que les tumeurs mammaires surviennent plus fréquemment parmi les chiennes stérilisées (avant leurs premières chaleurs) que parmi les autres chiennes.

En épidémiologie analytique, toute la question sera de savoir si l'association estimée dans l'échantillon est proche de l'association causale réelle (sous-entendu, dans la population cible). Pour répondre à la question, il faudra discuter et/ou prendre en compte les biais que nous verrons dans la suite de ce polycopié.

III. ETUDES EPIDEMIOLOGIQUES POUR TENTER D'ETABLIR UNE ASSOCIATION CAUSALE

Il existe deux grandes familles d'études épidémiologiques pour tenter d'établir une relation causale entre une exposition d'intérêt et une maladie : les **études interventionnelles** et les **études observationnelles**.

A. Comment distinguer une étude interventionnelle d'une étude observationnelle

Dans une étude observationnelle, les animaux recrutés dans l'étude sont exposés à des expositions ou ne le sont pas, mais sans qu'il n'y ait eu d'intervention de la part des investigateurs de cette étude : on *observe* les données (expositions et maladie étudiée) des animaux recrutés, et on les recueille dans une base de données avant d'entreprendre les analyses statistiques.

Dans une étude interventionnelle (quelle qu'elle soit, parmi celles qui vont être présentées ci-dessous), c'est un des investigateurs de l'étude (ou l'une des personnes impliquées dans l'étude) qui *décide* quels sont les animaux qui vont *devenir* exposés à l'exposition d'intérêt, et quels sont ceux qui ne vont pas *devenir* exposés à cette même exposition d'intérêt. Ainsi, dans une étude interventionnelle, il y a eu *intervention* de la part d'une personne impliquée dans l'étude dans le fait que les animaux recrutés dans l'étude sont ou ne sont pas exposés à l'exposition d'intérêt. De plus, les données recueillies dans une étude interventionnelle ont forcément pour objectif d'être analysées dans le cadre d'une étude.

B. Les études interventionnelles

1. Caractéristiques communes de toutes les études interventionnelles (essais cliniques)

Toutes les études interventionnelles, qu'elles soient randomisées ou non, ou qu'elles soient en aveugle ou non, partagent, en règle très générale, les caractéristiques communes suivantes :

- Les animaux recrutés dans l'étude sont indemnes de M au moment du recrutement (à J0) ;
- Il n'y a qu'une exposition d'intérêt dont on cherche à étudier le lien causal avec M ;
- L'attribution de l'exposition d'intérêt (devenir exposé ou devenir non exposé³) se réalise à J0 par l'investigateur de l'étude ou par une personne impliquée dans l'étude ;
- Les animaux sont suivis dans le temps à partir de l'attribution de l'exposition d'intérêt (J0), et l'on note la date de survenue de M.

Très souvent (mais ce n'est pas systématique), l'analyse de survie est utilisée pour comparer les animaux exposés à l'exposition d'intérêt par rapport à ceux qui ne le sont pas sur la survenue de M après J0.

Un grand nombre d'expositions ne peuvent pas être étudiées dans des études interventionnelles : toutes les expositions sur lesquelles il n'est pas possible d'agir (âge, sexe,

³ Par exemple : recevoir le traitement étudié (⇔ « être exposé à l'exposition d'intérêt ») ou recevoir le placebo (⇔ « ne pas être exposé à l'exposition d'intérêt »).

race, ...) ou pour lesquelles il n'est pas faisable ou pas éthique d'imposer à certains animaux l'exposition d'intérêt (environnement pollué, ...). Dans ces situations-là, seules les études observationnelles peuvent être utilisées.

Dans la mesure où (quasiment) toutes les études interventionnelles sont des essais cliniques, je parlerai dans la suite de ce polycopié d' « essai clinique » (« clinical trial » en anglais) quand il s'agira d'une étude interventionnelle.

2. L'essai clinique non randomisé

Un essai clinique non randomisé est une étude dans laquelle l'exposition d'intérêt a été attribuée de façon *non aléatoire*. Citons par exemple l'essai clinique en médecine humaine intitulée « Surgical induction of chorioretinal venous anastomosis in ischaemic central retinal vein occlusion: a non-randomised controlled clinical trial » (Mirshahi et al., 2005).

3. L'essai clinique randomisé

Un essai clinique randomisé (« randomised controlled trial » en anglais) est une étude dans laquelle l'exposition d'intérêt a été attribuée aléatoirement par l'investigateur de l'étude. Attention, j'insiste sur le fait que ce qui est randomisé, c'est l'exposition d'intérêt : une fois que l'animal a été inclus dans l'essai clinique (selon des critères bien stricts, qui eux, n'ont rien à voir avec le hasard !⁴), une table de randomisation va tirer au sort le fait que l'animal soit ou ne soit pas exposé à l'exposition d'intérêt. Pour cela, l'investigateur peut utiliser une « table de randomisation » (voir le site <http://www.randomization.com/> si vous souhaitez concevoir une étude randomisée pour votre thèse ou travail de recherche). L'exemple ci-dessous provient du site Internet cité, qui a attribué, dans un ordre aléatoire, le traitement ou le placebo dans un essai clinique incluant au total 10 animaux. Les deux premiers animaux recrutés dans l'étude recevront le traitement, le 3^{ème} recevra le placebo, le 4^{ème} recevra le traitement, et ainsi de suite.

A Randomization Plan
from
<http://www.randomization.com>

1. Traitement _____
2. Traitement _____
3. Placebo _____
4. Traitement _____
5. Placebo _____
6. Traitement _____
7. Placebo _____
8. Traitement _____
9. Placebo _____
10. Placebo _____

10 subjects randomized into 1 block
To reproduce this plan, use the seed 15162
along with the number of subjects per block/number of blocks
and (case-sensitive) treatment labels as entered originally.
Randomization plan created on 23/12/2019 à 11:15:15

⁴ Ainsi, les animaux recrutés dans un essai clinique ne sont pas du tout tirés au sort à partir d'une population beaucoup plus large !

4. L'essai clinique en simple aveugle et double aveugle

L'essai clinique en simple aveugle est un essai clinique au sein duquel le sujet de l'étude (ou, en médecine vétérinaire, le propriétaire de l'animal) ne sait pas s'il est ou s'il n'est pas exposé à l'exposition d'intérêt, mais celui qui évalue la survenue de M après J0 le sait quant à lui.

L'essai clinique en double aveugle est un essai clinique au sein duquel non seulement le sujet de l'étude ne sait pas s'il est ou s'il n'est pas exposé à l'exposition d'intérêt, mais celui qui évalue la survenue de M après J0 ne le sait pas non plus. Par exemple, si l'exposition d'intérêt est un traitement A (*versus* placebo) sous forme de comprimé, la mise en double aveugle pourra être réalisée en proposant le même emballage et la même couleur du comprimé que ce soit un comprimé du traitement A ou bien du placebo, mais avec un code de chiffres et de lettres, sans que ni le propriétaire ni celui qui évalue la survenue de M après J0 ne connaisse la clé du code utilisé.

Il existe des essais cliniques au sein desquels celui qui évalue la survenue de M après J0 ne sait pas si le sujet est exposé à l'exposition d'intérêt ou non, alors que le sujet, lui, le sait. Nous verrons par la suite que cette situation conduit à la même discussion sur les biais (de classement notamment) que dans la situation de double aveugle.

Attention, il ne faut surtout pas confondre « aveugle » (« blinded » en anglais) et « randomisé ». Il existe bien entendu des essais cliniques non randomisés mais en aveugle, comme par exemple celle-ci : « The Effect of Pivotal Response Treatment in Children with Autism Spectrum Disorders: A Non-randomized Study with a Blinded Outcome Measure » (Duifhuis et al., 2017).

C. Les études observationnelles

1. Caractéristiques communes de toutes les études observationnelles

Les études observationnelles sont des études qui sont réalisées entre autres lorsque l'attribution (aléatoire ou non, avec ou non mise à l'aveugle) de l'exposition d'intérêt par l'investigateur de l'étude n'est pas éthique ni/ou faisable, ou lorsque l'étude comprend plus d'une exposition d'intérêt. La démarche consiste à *observer* les données (les expositions et la maladie étudiée M), recueillies sur le « terrain », constituant ainsi une base de données qui servira par la suite aux analyses statistiques. Le gros inconvénient de ces études est qu'elles ne permettent pas d'établir avec certitude la relation causale entre une exposition d'intérêt et M (à cause de la présence inhérente de biais de confusion – cf. plus loin dans le polycopié). Il existe trois types d'études observationnelles : l'étude transversale, l'étude cas-témoins, et l'étude de cohorte.

Il faut noter que l'exposition d'intérêt dans une étude observationnelle est moins « homogène » que dans une étude interventionnelle. « Homogène » signifie ici « tous les animaux exposés à l'exposition d'intérêt y sont exposés de la même façon ». En effet, dans les études interventionnelles, l'exposition d'intérêt (sa présence, sa dose, la façon de l'avoir attribuée à l'animal, ...) est fixée par protocole, donc elle est très homogène. Dans une étude observationnelle au contraire, comme on observe les données sur le « terrain », il n'y a pas de standardisation de la façon dont les animaux sont exposés à l'exposition d'intérêt. Par exemple, la mise sous traitement (exposition d'intérêt) dans une étude observationnelle ne sera pas aussi homogène entre les animaux ayant reçu le traitement par rapport à une étude interventionnelle (par exemple, sur le « terrain », la posologie est fixée par le vétérinaire, et

peut varier d'un vétérinaire à l'autre, alors que dans une étude interventionnelle, tous les animaux qui reçoivent le traitement le reçoivent exactement de la même façon, car fixée par le protocole).

2. L'étude transversale

L'étude transversale n'est *a priori* pas une étude à visée analytique (recherche de facteurs de risque), mais plutôt à visée descriptive. Le recrutement des animaux s'effectue en général de la façon suivante : les animaux sont recrutés dans un endroit (une clinique, dans différentes fermes, ...) au moment où ils se présentent (ou au moment où l'investigateur se présente à eux). Certains d'entre eux sont *déjà* malades de la maladie étudiée M tandis que d'autres ne le sont pas, certains d'entre eux sont déjà exposés à l'exposition d'intérêt E tandis que d'autres ne le sont pas. C'est une étude rapide à mettre en place ; elle ne devrait cependant pas être mise en place si M est rare ou si l'exposition d'intérêt E est rare. En effet, si M est rare ou si l'exposition d'intérêt E est rare en population générale, elle risquera de l'être tout autant dans l'étude, et alors l'étude manquera de puissance statistique pour obtenir une association significative dans l'échantillon entre E et M.

3. L'étude cas-témoins

Le recrutement dans une étude cas-témoins s'effectue de la façon suivante : on recrute les cas (les animaux malades de la maladie étudiée) dans les endroits où il est plus facile de les recruter (par exemple, à l'hôpital, à partir de dossiers médicaux, ...), et l'on recrute des témoins (les animaux non malades de la maladie étudiée) de telle sorte que s'ils avaient été malades, ils auraient appartenu à la population source des cas. L'étude cas-témoins est une étude particulièrement adaptée au cas où la maladie étudiée M est rare en population générale (car on va aller chercher les cas là où ils se trouvent). Comme dans une étude transversale, les animaux du groupe « malade » sont des animaux qui sont *déjà* malades au moment du recrutement dans une étude cas-témoins. La différence avec l'étude transversale est que dans une enquête cas-témoins, l'investigateur va aller chercher les animaux malades là où ils sont, et il ne va pas attendre (comme c'est le cas dans l'enquête transversale) qu'ils se présentent à eux. Par ailleurs, en général, on recrute autant de témoins que de cas, de telle sorte à avoir dans l'échantillon environ 50% d'animaux malades. Ainsi, la puissance statistique d'une étude cas-témoins est bien meilleure que celle d'une étude transversale.

4. L'étude de cohorte

Le recrutement (J0) dans une étude de cohorte s'effectue de la façon suivante (cf. le cours d'analyse de survie dans l'UC-0313) : on recrute des animaux à J0 qui doivent *tous* être indemnes de la maladie étudiée M, on recueille toutes les informations nécessaires à ce moment-là, on suit dans le temps les animaux, et on note la date de survenue de M. L'analyse de survie est la méthode statistique pour analyser les données d'une étude de cohorte. Dans la mesure où les animaux du groupe « malade » sont des animaux qui sont *devenus* malades *après* J0, l'antériorité des expositions vis-à-vis de M est assurée (condition n°2 de la règle des 3A), contrairement aux études transversale et cas-témoins. Cependant, le prix à payer est le délai pour obtenir les premiers résultats qui peut être important !

Attention, il y a une grande différence entre l'étude de cohorte et l'essai clinique ! Dans une étude de cohorte, l'exposition d'intérêt n'est pas attribuée (aléatoirement ou non) par une personne impliquée dans l'étude. Elle l'a été « naturellement » (à la naissance, ou via le mode

de vie de l'animal, ou décidée « naturellement » par le clinicien de la clinique qui n'a pas de rapport avec l'étude).

IV. MESURES D'ASSOCIATION ENTRE E ET M

A. Introduction

Dans la suite de ce chapitre, nous allons utiliser les données du tableau suivant (les lettres dans les cases sont des nombres d'animaux) :

		Maladie M		Total
		M+	M-	
Exposition E	E+	a	b	e ₁
	E-	c	d	e ₀
Total		m ₁	m ₀	N

Notations :

$R_0 = c/e_0$ (pourcentage de M+ parmi les animaux E-)

$R_1 = a/e_1$ (pourcentage de M+ parmi les animaux E+)

$PE_{M-} = b/m_0$ (pourcentage de E+ parmi les animaux M-)

$PE_{M+} = a/m_1$ (pourcentage de E+ parmi les animaux M+)

B. Risque relatif (RR)

1. Calcul d'un RR

Le risque relatif (RR) est le rapport du pourcentage d'incidence de M⁵ parmi les animaux exposés (E+) sur le pourcentage d'incidence de M parmi les animaux non exposés (E-). (Il peut aussi être le rapport des pourcentages de prévalence de M⁶, seulement si ces pourcentages dans l'échantillon sont proches de ceux dans la population cible.) La formule (à connaître) est la suivante :

$$RR = \frac{R_1}{R_0}$$

Dans la mesure où le nombre de cas (animaux M+) est fixé à l'avance dans une étude cas-témoins, les pourcentages de prévalence de M dans les études cas-témoins n'ont absolument aucun sens et sont de fait très éloignés de ceux dans la population cible. Par conséquent, il est interdit de calculer un RR dans une étude cas-témoins. Les RR sont principalement calculés dans les études de cohortes et les essais cliniques.

⁵ Pourcentage d'animaux qui sont devenus M+ durant une période de temps donnée.

⁶ Pourcentage d'animaux qui sont déjà M+ au moment du recrutement.

2. Interprétation de la valeur d'un RR dans l'échantillon

La valeur du RR a une interprétation simple dans l'échantillon. Un $RR_{E \rightarrow M} = k$ ($k \in]0 ; +\infty[$) quantifiant l'association entre E et M signifie que dans l'échantillon, M était k fois plus fréquente parmi les animaux E+ que parmi les animaux E-. Je vais illustrer cela avec deux exemples.

Le premier exemple est présenté ci-dessous.

		Maladie M		Total
		M+	M-	
Exposition E	E+	3	7	10
	E-	1	9	10
Total		4	16	20

Dans l'exemple ci-dessus, le pourcentage de M+ parmi les E+ est de $3/10 = 30\%$, et le pourcentage de M+ parmi les E- est de $1/10 = 10\%$. La maladie M était donc 3 fois plus fréquente parmi les E+ que parmi les E-. C'est ce qu'indique d'ailleurs la valeur du RR lorsqu'on le calcule :

$$RR = \frac{R_1}{R_0} = \frac{3/10}{1/10} = 3$$

Le deuxième exemple est présenté ci-dessous.

		Maladie M		Total
		M+	M-	
Exposition E	E+	3	7	10
	E-	6	4	10
Total		9	11	20

Dans l'exemple ci-dessus, le pourcentage de M+ parmi les E+ est de $3/10 = 30\%$, et le pourcentage de M+ parmi les E- est de $6/10 = 60\%$. La maladie M était donc 2 fois *moins* fréquente parmi les E+ que parmi les E-. On peut le dire aussi autrement⁷ : la maladie est 0,5 fois *plus* fréquente parmi les E+ que parmi les E-. Et là encore, c'est ce qu'indique la valeur du RR lorsqu'on le calcule :

$$RR = \frac{R_1}{R_0} = \frac{3/10}{6/10} = 0,5$$

Le point ci-dessus se généralise ainsi : si la maladie est k ($k \in]0 ; +\infty[$) fois plus fréquente parmi les animaux E+ que parmi les animaux E-, elle est aussi $1/k$ fois *moins* fréquente parmi les animaux E+ que parmi les animaux E- (par exemple, « 0,5 fois plus » est synonyme de « 2 fois moins »).

Pour résumer, un $RR_{E \rightarrow M} = k$ ($k \in]0 ; +\infty[$) quantifiant l'association entre E et M signifie que dans l'échantillon, M était k fois plus fréquente parmi les animaux E+ que parmi les animaux

⁷ Même si j'ai bien conscience que la tournure n'est pas des plus naturelles !

E-. Si k est > 1 , alors M était *plus* fréquente parmi les animaux E+ que parmi les animaux E-, et si $k < 1$, alors M était *moins* fréquente parmi les animaux E+ que parmi les animaux E-.

Autre chose très importante : si $RR_{E \rightarrow M} = 1$, cela signifie que, dans l'échantillon, M était 1 fois plus fréquente parmi les animaux E+ que parmi les animaux E-. Autrement dit, M était *aussi* fréquente parmi les animaux E+ que parmi les animaux E-. Dit encore autrement, dans l'échantillon, il n'existait absolument aucune association entre E et M.

3. Calcul de l'intervalle de confiance et interprétation

Pour calculer l'intervalle de confiance à 95% d'un RR ($[RR_{min} ; RR_{max}]_{95\%}$), il faut tout d'abord passer à l'échelle logarithmique (logarithme népérien). Ensuite il faut calculer la « standard error » (SE) de $\ln(RR)$. (Les deux formules ci-dessous ne sont pas à connaître.)

$$\text{Standard error } [\ln(RR)] = SE[\ln(RR)] = \sqrt{\frac{b}{a \cdot (a + b)} + \frac{d}{c \cdot (c + d)}}$$

$$[RR_{min} ; RR_{max}]_{95\%} = [e^{\ln(RR) - 1,96 \cdot SE[\ln(RR)]} ; e^{\ln(RR) + 1,96 \cdot SE[\ln(RR)]}]$$

L'interprétation des valeurs des bornes de l'intervalle de confiance à 95% d'un RR est équivalente à celle que vous avez vue en biostatistique dans l'UC-0213 (ce qui est écrit entre crochets dans ce qui suit doit être remplacé par vous) : « si l'estimation du RR quantifiant l'association entre [E et M] n'est pas biaisée, [dans la population cible], il y a 95% de chances pour que le RR réel correspondant soit compris entre [la borne inférieure] et [la borne supérieure] ». Attention, un RR *calculé* dans l'échantillon a quant à lui 100% de chances d'être inclus dans son intervalle de confiance à 95% (cf. formules ci-dessus) ! Cette remarque vaudra bien entendu aussi pour l'Odds Ratio (cf. ci-dessous).

C. Odds Ratio (OR)

1. Calcul d'un OR

« Odds » en anglais se traduit par « cote »⁸, en français. Ainsi, l'OR est le rapport des cotes. La cote chez les animaux exposés vaut $R_1 / (1 - R_1)$ et celle chez les animaux non exposés vaut $R_0 / (1 - R_0)$.

$$OR = \frac{\frac{R_1}{(1 - R_1)}}{\frac{R_0}{(1 - R_0)}}$$

On peut aussi montrer que :

$$OR = \frac{\frac{PE_{M+}}{(1 - PE_{M+})}}{\frac{PE_{M-}}{(1 - PE_{M-})}}$$

⁸ On retrouve la notion de « cote » dans les courses ou les paris hippiques : lorsqu'un cheval est coté à 3 contre 1, cela signifie que 3 parieurs l'ont joué perdant contre 1 gagnant. (« Cote » est sans accent circonflexe sur le « o ».)

L'OR peut s'estimer dans tout type d'étude (transversale, cas-témoins, cohorte, et études interventionnelles). On peut enfin montrer que (c'est la formule à connaître, celle que l'on utilisera dans les TD) :

$$OR = \frac{a \cdot d}{b \cdot c}$$

2. Interprétation de la valeur d'un OR dans l'échantillon

Dans la mesure où l'OR n'est *pas* un rapport de pourcentages d'incidence (ou de prévalence) de M, on ne peut *pas* dire qu'un $OR_{E \rightarrow M} = k$ ($k \in]0; +\infty[$) signifie que M était k fois plus fréquente parmi les animaux E+ que parmi les animaux E- (comme on pouvait le faire avec un RR).

Ce que l'on peut en revanche dire (tout comme on le disait déjà avec le RR), c'est qu'avec un $OR_{E \rightarrow M} = k$ avec $k > 1$, M était plus fréquente parmi les animaux E+ que parmi les animaux E-. Si $k < 1$, on peut dire que M était moins fréquente parmi les animaux E+ que parmi les animaux E-. De plus, et de la même façon qu'avec le RR, si $OR_{E \rightarrow M} = 1$, alors cela signifie que, dans l'échantillon, il n'existait absolument aucune association entre E et M.

Remarque : dans la situation où la maladie est **rare** en population générale (pourcentage de prévalence de M bien inférieur à 1%), la valeur de l'OR peut être interprétée comme celle d'un RR. Cela dit, très souvent, en recherche clinique vétérinaire, la maladie étudiée ne peut pas être considérée comme rare. Ainsi, dans ce polycopié (et dans tous les exercices d'enseignement en épidémiologie clinique de l'UC-0324), je ne partirai *jamais* du principe que la maladie est rare. Dans des études épidémiologiques conduites chez des animaux de production, il arrive que la maladie étudiée soit rare. Ainsi, contrairement à l'UC-0324, les OR calculés dans l'UC-0322 (où il est question d'épidémiologie conduite auprès d'animaux de production) pourront éventuellement être interprétés comme des RR.

3. Calcul de l'intervalle de confiance et interprétation

Pour calculer l'intervalle de confiance à 95% de l'OR ($[OR_{min}; OR_{max}]_{95\%}$), comme avec le RR, il faut tout d'abord passer à l'échelle logarithmique (logarithme népérien). Ensuite il faut calculer la « standard error » de $\ln(OR)$. (Les deux formules ci-dessous ne sont pas à connaître.)

$$\text{Standard error } [\ln(OR)] = SE[\ln(OR)] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

$$[OR_{min}; OR_{max}]_{95\%} = [e^{\ln(OR) - 1,96 \cdot SE[\ln(OR)]}; e^{\ln(OR) + 1,96 \cdot SE[\ln(OR)]}]$$

L'interprétation de l'intervalle de confiance à 95% d'un OR est équivalente à celle de l'intervalle de confiance du RR : « si l'estimation de l'OR quantifiant l'association entre [E et M] n'est pas biaisée, [dans la population cible], il y a 95% de chances pour que l'OR réel correspondant soit compris entre [la borne inférieure] et [la borne supérieure] ».

D. Largeur d'un intervalle de confiance à 95% et précision

Plus les nombres d'animaux dans chacune des quatre cases du tableau d'effectif diminuent (les lettres « a », « b », « c », et « d » ci-dessus), et moins le RR ou l'OR sera précis. En effet, plus ces effectifs diminuent, et plus la SE[Ln(RR)] et la SE[Ln(OR)] augmente (cf. les deux formules ci-dessus). Et la SE d'une estimation (ici, Ln(RR) ou Ln(OR)) quantifie la précision d'une estimation : plus la SE augmente, plus la précision de l'estimation diminue.

Les formules des intervalles de confiance à 95% des OR et RR ci-dessus nous rappellent que plus la SE[Ln(RR)] ou la SE[Ln(OR)] augmente, plus la largeur de l'intervalle de confiance à 95% (IC_{95%}) augmente.

Ainsi, une autre interprétation, subjective cette fois-ci, mais que l'on utilise beaucoup plus souvent en recherche clinique que les interprétations des valeurs des bornes d'un IC_{95%} d'un OR ou d'un RR que j'ai mentionnées ci-dessus est la suivante : « la largeur d'un IC_{95%} donne une indication de la précision avec laquelle le RR ou l'OR a été calculé. Plus la largeur de l'IC_{95%} est importante, moins le RR ou l'OR est précis ».

Même s'il n'y a pas de consensus universel sur les valeurs des bornes d'un IC_{95%} pour dire qu'un RR ou qu'un OR n'est pas précis, j'utilise les valeurs de 500 (pour la borne supérieure) ou son symétrique 1/500 = 0,002 (pour la borne inférieure). Ainsi, par exemple, un RR de 7,4 dont l'IC_{95%} est [0,1 ; 992,3] sera considéré comme très imprécis.

E. Tester un OR ou un RR

1. Au sein de l'UC-0324

Je vous rappelle que s'il y a absence totale d'association entre E et M, le RR_{E→M} et l'OR_{E→M} sont égaux à 1. Par conséquent, tester s'il existe une association significative entre E et M revient à tester si le RR ou l'OR est significativement différent de 1. Si le RR ou l'OR est significativement différent de 1, il existe alors une association significative entre E et M.

Une première façon de savoir si un RR ou un OR est significativement différent de 1 est de lire son IC_{95%} : si « 1 » n'appartient pas à l'IC_{95%}, le RR ou l'OR sera significativement différent de 1 (**Figure 4.a** et **Figure 4.b**) ; si « 1 » appartient à l'IC_{95%}, le RR ou l'OR ne sera pas significativement différent de 1 (**Figure 4.c** et **Figure 4.d**).

Légende

● —● OR ou RR (IC95%)

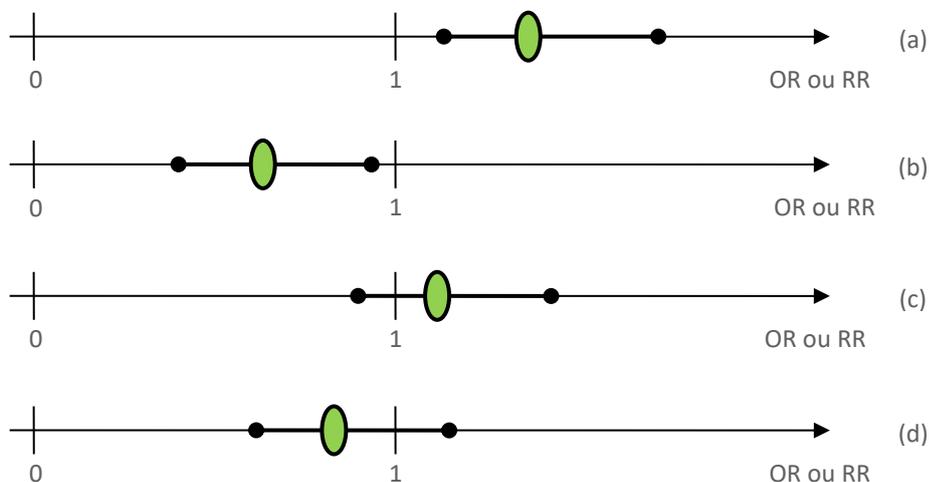


Figure 4. Différentes situations où l'OR* est significativement différent de « 1 » ((a) et (b)) et où l'OR* n'est pas significativement différent de « 1 » ((c) et (d)).

Une seconde façon de savoir si un OR ou un RR est significativement différent de 1 consiste à lire le degré de signification fourni par le logiciel statistique dans le cadre de l'utilisation de modèle de régression univarié ou multivarié (cf. partie VIII.D.6 dans ce polycopié).

Ainsi, dans le cadre de l'UC-0324, pour savoir si un OR ou un RR est significativement différent de 1 (donc pour savoir s'il existe une association significative entre E et M), on utilisera soit l'IC_{95%} de l'OR ou du RR, soit le degré de signification fourni par le modèle de régression.

2. Définition du terme « significativité »

Dans la suite de ce polycopié, la « significativité » d'un OR ou d'un RR fait référence au fait que l'OR ou le RR est *ou n'est pas* significativement différent de 1. Par exemple, si une situation est capable de modifier la significativité d'un OR, cela veut dire, par exemple, que cette situation peut conduire au fait que cet OR peut tout à fait devenir significativement différent de 1 à cause de cette situation, alors qu'il n'aurait pas été significativement différent de 1 sans cette situation.

3. Différence avec l'UC-0322

Dans l'UC-0322, le test de l'association entre E et M peut se réaliser à l'aide des méthodes présentées ci-dessus ou à partir d'un test du Chi-2. La démarche utilisant le test du Chi-2 ne sera pas utilisée au sein de l'UC-0324.

F. Association « positive, « négative », et « forte » : signification et illustration avec un OR ou un RR

Toute cette partie a pour objectif de vous rendre capable de répondre à la question suivante, à partir de la lecture du tableau 1 de l'article de Sogaro-Robinson et coll. (Sogaro-Robinson et al., 2009) testant chez des chevaux l'association entre différentes expositions et la présence de résultats anormaux au scanner (la maladie étudiée, abrégée par « RAS ») : « dans l'étude, l'association entre la présence de RAS et celle d'un traumatisme crânien (« head trauma ») était-elle plus forte ou moins forte que celle entre la présence de RAS et la présence de comportements de type convulsion (« seizure-like activity ») (cf. cadre ci-dessous dans le tableau) ? »

Table 1—Results of univariate analysis of predictive factors for abnormal results of CT in 57 horses with neurologic deficits.

Variable	Category	No. of horses with abnormal CT results	No. of horses with normal CT results	OR	95% CI	P value*
Sex	Sexually intact male	2	10	1.00	Referent	NA
	Gelding	6	21	1.43	0.24–8.37	0.693
	Female	5	13	1.92	0.31–12.05	0.485
Breed	Quarter Horse†	7	12	1.00	Referent	NA
	Thoroughbred	3	13	0.40	0.08–1.89	0.245
	Arabian and Standardbred	1	10	0.17	0.02–1.64	0.126
	Other	2	9	0.38	0.06–2.29	0.292
Age (y)	< 1	1	4	1.00	Referent	NA
	1–4	4	6	2.67	0.21–33.50	0.447
	5–15	6	19	1.26	0.12–13.59	0.847
	> 15	2	15	0.53	0.04–7.49	0.641
Head trauma	No	10	43	1.00	Referent	NA
	Yes	3	1	12.9	1.21–137.37	0.034
Seizure-like activity	No	8	5	1.00	Referent	NA
	Yes	5	39	0.08	0.02–0.34	0.001

Tout d'abord, re-précisons qu'un OR ou un RR de valeur égale à 1 signifie « absence d'association entre E et M ».

On dit qu'une association entre E et M est « positive » si M est *plus* fréquente parmi les animaux E+ que parmi les animaux E-. Cette situation arrive lorsque l'OR_{E→M} et le RR_{E→M} ont une valeur supérieure à 1. On dit qu'une association entre E et M est « négative » si M est *moins* fréquente parmi les animaux E+ que parmi les animaux E-. Cette situation arrive lorsque l'OR_{E→M} et le RR_{E→M} ont une valeur inférieure à 1.

Si une exposition E est *fortement et positivement* associée à une maladie M, cela signifie que la maladie est *beaucoup plus* fréquente parmi les animaux E+ que parmi les animaux E-. Dans ce cas-là, l'OR_{E→M} ou le RR_{E→M} sera très éloigné de « 1 », par valeur *supérieure*.

Si une exposition E est *fortement et négativement* associée à M, cela signifie que M est *beaucoup moins* fréquente parmi les animaux E+ que parmi les animaux E-. Dans ce cas-là, l'OR_{E→M} ou le RR_{E→M} sera « très » éloigné de « 1 » par valeur *inférieure*, donc il sera très proche de 0 (tout en restant supérieur à 0, bien entendu).

Parce que cette notion est un peu compliquée, je vais utiliser le RR (car c'est plus simple que l'OR pour comprendre les choses), et prendre deux exemples dont les données sont présentées dans les deux tableaux ci-dessous : le premier étudiant l'association entre une exposition E₁ et M, et le second étudiant l'association entre une exposition E₂ et M.

		Maladie M		Total
		M+	M-	
Exposition E ₁	E ₁ +	1	99	100
	E ₁ -	10	90	100
Total		11	189	200

Le pourcentage de M+ parmi les E₁+ est de 1% (1/100), et le pourcentage de M+ parmi les E₁- est de 10% (10/100). Par conséquent, M était dix fois *moins* fréquente parmi les E₁+ que parmi les E₁- (l'association entre E₁ et M est donc dite « négative »).

		Maladie M		Total
		M+	M-	
Exposition E ₂	E ₂ +	5	95	100
	E ₂ -	10	90	100
Total		15	185	200

Le pourcentage de M+ parmi les E₂+ est de 5% (5/100), et le pourcentage de M+ parmi les E₂- est de 10% (10/100). Par conséquent, M était deux fois *moins* fréquente parmi les E₂+ que parmi les E₂- (l'association entre E₂ et M est donc là encore dite « négative »).

Par conséquent, l'association entre E₁ et M était plus forte que l'association entre E₂ et M (« dix fois moins fréquente » est plus fort que « deux fois moins fréquente »).

En utilisant les formules du calcul du RR de la partie IV.B.1, on obtient :

$$RR_{E_1 \rightarrow M} = \frac{R_1}{R_0} = \frac{1/100}{10/100} = 0,1$$

$$RR_{E_2 \rightarrow M} = \frac{R_1}{R_0} = \frac{5/100}{10/100} = 0,5$$

On observe bien que le RR quantifiant l'association la plus forte des deux (celle entre E₁ et M) est bien le RR le plus éloigné de « 1 » par valeur inférieure parmi les deux RR (« 0,1 » est plus éloigné de « 1 » que ne l'est « 0,5 »).

En prenant un autre exemple, un RR de 0,2 traduit une association plus forte qu'un RR de 4, car un RR de 0,2 signifie que M est cinq fois moins fréquente parmi les E+ que parmi les E- (5 = 1/0,2), et « cinq fois moins fréquente » est plus fort que « quatre fois plus fréquente ». Vous pouvez donc vous rendre compte qu'une association « négative » (RR ou OR < 1) peut être plus *forte* qu'une association « positive » (RR ou OR > 1).

En résumé, plus la valeur d'un OR_{E→M} ou d'un RR_{E→M} estimée dans un échantillon s'éloigne de « 1 » (par valeur supérieure ou inférieure), plus cet OR ou RR traduit une association forte entre E et M dans cet échantillon, mais pas de façon proportionnelle entre à droite et à gauche de « 1 » ! La relation entre valeur du RR (ou de l'OR) et force de l'association est décrite dans **Figure 5** ci-dessous.

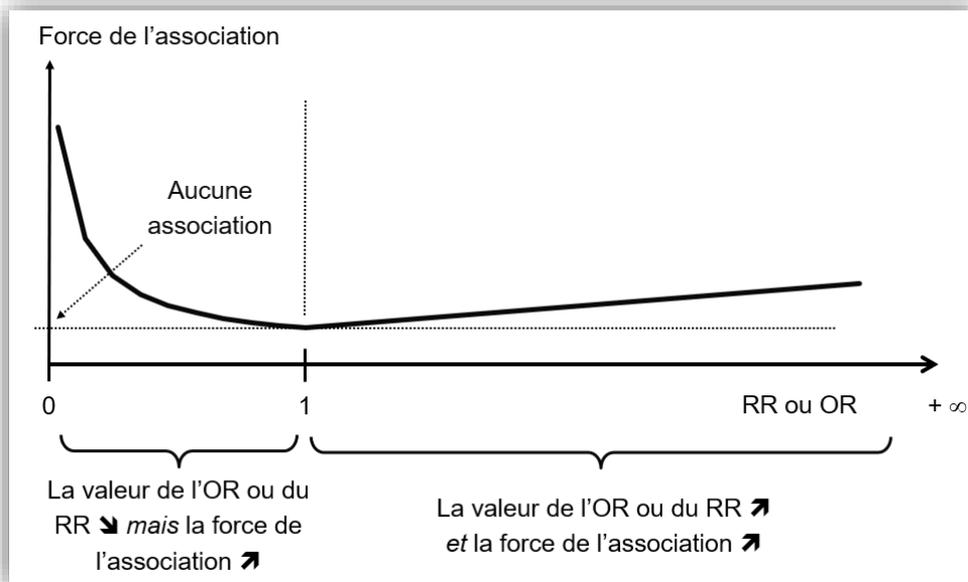


Figure 5. Relation entre la valeur d'un OR ou d'un RR et la force de l'association qu'il quantifie.

On peut généraliser le raisonnement en appliquant la règle suivante pour comparer les forces des associations à partir de la valeur des RR estimés dans un échantillon : il faut toujours comparer des valeurs de RR supérieures à 1, et si un RR est inférieur à 1, il faut le faire passer de l'autre côté de « 1 » en passant par l'inverse du RR ($\Leftrightarrow 1/RR$).

On peut même généraliser cette règle avec les OR en appliquant *exactement* la même règle avec deux OR estimés dans un échantillon. Pour reprendre l'exemple de l'article de Sogaro-Robinson et coll. ci-dessus, l'OR comportements de type convulsion \rightarrow RAS (OR = 0,08 ; $p = 0,001$) traduit-il une association plus forte ou moins forte que l'OR traumatisme crânien \rightarrow RAS (OR = 12,9 ; $p = 0,034$) ? Pour répondre à la question, il faut comparer les valeurs des OR en plaçant les deux OR du même côté de « 1 » par valeur supérieure. La valeur de 12,9 est du bon côté de « 1 ». Il faut désormais placer la valeur de 0,08 de l'autre côté de « 1 » en faisant : $1/0,08 = 12,5$. Par conséquent, puisque $12,9 > 12,5$, alors on peut dire que, *dans l'échantillon*, l'association entre la présence de traumatisme crânien et celle de RAS était plus forte que celle avec les comportements de type convulsion.

Vous voyez que si vous aviez comparé les valeurs des degrés de signification (0,001 *versus* 0,034), vous auriez dit l'inverse ! De façon générale, pour comparer les forces des associations *dans un échantillon*, il ne faut surtout pas comparer les valeurs des degrés de signification, mais comparer les valeurs des OR ou des RR en les plaçant du même côté de « 1 » par valeur supérieure. En particulier, dire qu'un OR non significatif traduit une association moins forte qu'un OR significatif parce que « le premier n'est pas significatif alors que le second l'est » est une très belle erreur de raisonnement (on a déjà vu dans l'UC-0213 ce point fondamental). En effet, le degré de signification ne donne pas d'information quant à l'importance clinique, ou la force, d'une association dans un échantillon ; il quantifie uniquement la rareté d'un phénomène observé dans un échantillon sous H_0 (ce qui permet de ne faire qu'une seule chose, à savoir rejeter H_0 ⁹, et le rejet de H_0 ne porte *que* sur la population cible).

⁹ Si et seulement si le degré de signification est $\leq 0,05$.

V. GENERALITE SUR LES BIAIS D'ASSOCIATION

A. Remarques préliminaires

Pour correctement discuter la présence et/ou l'impact d'un biais d'association entre une exposition E et une maladie M, il est **in-dis-pen-sable** que vous ayez écrit sur un bout de papier ce que sont E et M. Si, à la lecture d'un protocole, vous identifiez mal E et/ou M, toutes les discussions sur les biais seront ... fausses.

Par ailleurs, pour éviter des lourdeurs d'écriture, je vais désormais écrire « OR* » pour éviter d'écrire « OR ou RR » dans tout ce qui suit dans le polycopié (sauf dans les titres).

B. Définition d'un biais d'association

Un biais d'association est un écart¹⁰ *systématique* entre la valeur de l'association estimée (quantifiée par un OR*) et la valeur réelle de l'association que l'on vise dans la population cible. (Cf. la définition générale d'un « biais » qui figure dans le polycopié de Biostatistique en Médecine Vétérinaire de l'UC-0213.) Je vous rappelle qu'un OR* estimé dans un échantillon sera forcément différent de celui, correspondant, dans la population cible, tout d'abord à cause du hasard (autrement dit, de la fluctuation d'échantillonnage), et ensuite, parce qu'il peut y avoir du biais d'association. Le biais est la partie de l'écart entre l'OR* estimé et l'OR* réel qui n'est pas due à la fluctuation d'échantillonnage. Cette partie d'écart est « systématique » car elle a pour origine une erreur dans la mise en place et/ou dans l'exécution du protocole de recrutement, de collecte des données, et d'analyses statistiques. En effet, si l'on refaisait une deuxième étude recrutant d'autres animaux mais ayant le même objectif que la première étude, avec le même protocole (de recrutement, de collecte des données, et d'analyses statistiques), la partie de l'écart entre le nouvel OR* estimé dans cette nouvelle étude et le même OR* réel qui n'est pas due à la fluctuation d'échantillonnage sera *a priori* identique à celle dans la première étude (d'où le mot « systématique » dans la définition du biais).

C. Présentation des biais d'association

En toute théorie, pour pouvoir faire de l'inférence causale avec un $OR^*_{E \rightarrow M}$ estimé dans un échantillon, l'association entre E et M ne doit être biaisée par aucun biais d'association. On peut dénombrer trois biais d'association, dont deux sont au programme de l'UC-0324 : le biais de **classement**, le biais de **confusion**. Le biais de **sélection** est le troisième biais d'association, mais il est hors programme (discussion difficile et très peu d'études cliniques comportent des biais de sélection ; mais ce biais est présenté en annexe A).

¹⁰ Et non pas une « erreur » ! Un biais n'est *jamais* une erreur. En revanche, un biais a pour *origine* une erreur.

D. OR (ou RR) estimé versus OR (ou RR) causal

1. Aspect théorique

L'OR* **estimé** est l'OR* calculé dans l'échantillon. Nous verrons dans la suite de ce polycopié que nous aurons trois types d'OR* estimés : l'OR* brut, l'OR* stratifié sur une exposition, et l'OR* ajusté sur une/plusieurs exposition(s).

L'OR*_{E→M} **causal** quantifie l'association causale entre une exposition E et M dans la population cible. Cet OR* causal nous¹¹ est donc inconnu. Il ne pourra jamais être calculé. L'objectif de l'épidémiologie analytique est justement d'estimer un OR* dans un échantillon qui soit le plus proche possible de cet OR* causal.

Si l'OR*_{E→M} causal dans la population est égal à 1 (sans que nous ne le sachions), cela signifie que E n'est ni un facteur de risque, ni un facteur protecteur de M. Si l'OR*_{E→M} causal dans la population est inférieur à 1, cela signifie que E est un facteur protecteur de M. Si l'OR*_{E→M} causal dans la population est supérieur à 1, cela signifie que E est un facteur de risque de M.

2. Illustration sur un exemple

Prenez l'exemple de la stérilisation chez la chienne avant les premières chaleurs et la survenue par la suite de tumeurs mammaires. La valeur de l'OR*_{stérilisation → tumeurs mammaires} causal est inconnue. Est-ce qu'elle est plus grande que « 1 » (auquel cas, la stérilisation augmenterait les risques de tumeurs mammaires par la suite), plus petite que « 1 » (auquel cas, la stérilisation serait un facteur protecteur de survenue de tumeurs mammaires), ou égale à « 1 » (auquel cas, la stérilisation n'aurait aucun impact sur la survenue de tumeurs mammaires) ? Pour répondre à cette question, nous devons mettre en place des études cliniques rigoureuses au sein desquelles l'OR*_{stérilisation → tumeurs mammaires} estimé dans l'étude sera le plus proche possible de l'OR*_{stérilisation → tumeurs mammaires} causal. Comment ? Facile, en limitant au maximum les biais de classement, confusion, et sélection ☺ (sauf qu'en 2012, de telles études semblaient ne pas encore exister (Beauvais et al., 2012)).

VI. LE BIAIS DE CLASSEMENT

A. Origine du biais de classement et commentaires

L'OR*_{E→M} est biaisé par du biais de classement dès qu'il existe une erreur de classement (aléatoire ou systématique) commise sur E et/ou dès qu'il existe une erreur de classement (aléatoire ou systématique) commise sur M. Ces erreurs de classement proviennent d'une mauvaise façon de recueillir les données sur E et/ou sur M dans l'étude.

Une erreur de classement commise sur E signifie que, lors du recueil des données, on va considérer un animal exposé alors qu'il ne l'est pas en réalité, ou bien que l'on va considérer un animal non exposé alors qu'il l'est en réalité.

¹¹ Nous, tous petits êtres humains que nous sommes.

Une erreur de classement commise sur M signifie que, lors du recueil des données, on va considérer un animal « malade » alors qu'il ne l'est pas en réalité, ou bien que l'on va considérer un animal non « malade » alors qu'il l'est en réalité.

J'insiste (car c'est une chose contre-intuitive) : même si une erreur de classement sur E est commise *aléatoirement*, ou même si une erreur de classement sur M est commise *aléatoirement*, l' $OR^*_{E \rightarrow M}$ sera quand même biaisé par du biais de classement.

Par ailleurs, j'ai écrit ci-dessus : « L' $OR^*_{E \rightarrow M}$ [...] est biaisé par du biais de classement dès qu'il existe une erreur de classement [...] commise sur E **et/ou** [...] sur M. » Ce « et/ou » indique quelque chose de très important dans le biais de classement : **E et M jouent des rôles totalement interchangeables**.

Enfin, il existe deux types de biais de classement : le biais de classement **non différentiel** et le biais de classement **différentiel**. Le type de biais de classement dépend de la *façon* dont l'erreur de classement sur E (et/ou sur M) a été commise.

B. Evaluation objective ou subjective de l'exposition ou de la maladie étudiée

Une évaluation « objective » de E ou de M est une évaluation qui ne comprend aucune subjectivité de la part de celui ou celle qui recueille l'information sur E ou M. Par exemple, la mesure de E ou M à l'aide d'un instrument de mesure est forcément objective, ce qui ne veut pas dire que cette évaluation sera de fait *sans* erreur de classement.

Une évaluation « subjective » de E ou de M est une évaluation qui comprend une part de subjectivité de la part de celui ou celle qui recueille l'information sur E ou M. Une évaluation subjective de E ou de M conduira forcément à quelques erreurs de classement.

C. Le biais de classement non différentiel

1. Origine du biais de classement non différentiel

L' $OR^*_{E \rightarrow M}$ est biaisé par du biais de classement non différentiel si l'**erreur** de classement commise sur E est **indépendante** du statut malade / non malade de l'animal.

Qu'est-ce qu'une erreur de classement commise sur E indépendamment de M ? Cela signifie que la personne qui commet une erreur de classement sur E le fait *a priori* de la même façon que les animaux soient « malades » (M+) ou non « malades » (M-). Autrement dit, la personne qui recueille l'information sur E n'est pas influencée par le statut M+/M- de l'animal au moment de recueillir l'information sur E. Ainsi, l'erreur de classement est *non* différentielle car l'erreur de classement sur E ne se commet pas de façon différentielle selon que les animaux sont M+ ou M-. Les erreurs de classement *aléatoires* sur E sont, de fait, des erreurs de classement sur E indépendantes du statut M+/M-. Par conséquent, ces erreurs de classement aléatoires sur E conduiront à du biais de classement non différentiel.

Remarque : n'oubliez pas que dans le biais de classement, E et M sont interchangeables. Par conséquent, relisez ce que je viens d'écrire en remplaçant dans votre tête « E » par « M » et « M » par « E ».

2. Illustration sur un exemple

Supposons une étude dont l'objectif est de quantifier, à l'aide d'un OR, l'association entre une alimentation de type « supermarché » (*versus* « alimentation de marque ») et la présence d'un lymphome chez le chat : $OR_{\text{alimentation supermarché} \rightarrow \text{lymphome}}$. Dans cette étude, le type d'alimentation (E) est recueilli par un questionnaire que le propriétaire remplit avec le vétérinaire pendant la consultation, à partir de la question suivante : « Nourrissez-vous votre chat principalement avec une alimentation de supermarché, ou bien principalement avec une alimentation de marque (Royal Canin, Hills, ...) ? ». Le diagnostic de lymphome (M) est posé à partir de prélèvements dont les résultats ne sont disponibles qu'une semaine après la consultation chez le vétérinaire. On va supposer qu'il n'y a pas d'erreur de classement sur M (\Leftrightarrow pas d'erreur de diagnostic). En revanche, on peut tout à fait imaginer des erreurs de classement commises sur E : le recueil de données sur l'alimentation est subjectif, donc sujet à erreurs (ici, on peut penser que certains propriétaires vont classer à tort leur chat dans la catégorie « alimentation de marque » parce qu'ils n'osent pas dire à leur vétérinaire qu'ils donnent principalement de l'alimentation de supermarché). Sauf que ces erreurs de classement sur E (l'alimentation du chat) sont forcément commises de façon *indépendante* du statut M+/M- du chat, car ce statut M+/M- n'était pas connu *au moment* où le propriétaire répondait au questionnaire. Autrement dit, le propriétaire et le vétérinaire n'ont pas pu être influencés par le fait que le chat ait ou non un lymphome au moment de classer le chat de façon exposée (alimentation de type « supermarché ») ou non exposée (« alimentation de marque »). Ainsi, dans cette étude, $OR_{\text{alimentation supermarché} \rightarrow \text{lymphome}}$ sera biaisé par du biais de classement non différentiel.

3. Impact du biais de classement non différentiel sur la valeur de l'OR ou du RR

L'impact du biais de classement non différentiel sur l'OR* est tel qu'il diminue toujours la force de l'association (revoyez la **Figure 5** si besoin) : la valeur de l'OR* estimé et biaisé est plus proche de « 1 » que ne l'est la valeur de l'OR* réel correspondant dans la population (donc non biaisé).

En quelque sorte, les erreurs de classement non différentielles (sur E et/ou sur M) créent un biais de classement non différentiel « qui brouille les pistes ». Ces erreurs de classement ajoutent du « bruit » de sorte que l'association biaisée en raison de ces erreurs de classement non différentielles va être moins forte que si ces erreurs n'avaient pas été commises.

Une illustration de cela est présentée en annexe B, en fin de polycopié.

4. Impact du biais de classement non différentiel sur la significativité d'un OR ou d'un RR

Si un OR*, significativement différent de 1 ($p \leq 0,05$), est biaisé par du biais de classement non différentiel, la présence de ce biais ne remet pas en cause la significativité de l'OR*.

Explication : si l'OR* n'avait pas été biaisé par du biais de classement non différentiel, sa **force** aurait été plus importante, il aurait donc été encore plus éloigné de « 1 » qu'il ne l'est en étant biaisé, et son $IC_{95\%}$ aurait continué de ne pas inclure « 1 » (**Figure 6.a** et **Figure 6.b**). Par conséquent, l'OR*, s'il n'avait pas été biaisé par du biais de classement non différentiel, serait resté significativement différent de 1.

Si un OR*, non significativement différent de 1 ($p > 0,05$), est biaisé par du biais de classement non différentiel, la présence de ce biais remet *potentiellement* en cause la significativité de l'OR*.

Explication : si l'OR* n'avait pas été biaisé par du biais de classement non différentiel, sa **force** aurait été plus importante, il aurait donc été plus éloigné de « 1 » qu'il ne l'est en étant biaisé, et son IC_{95%} aurait cette fois-ci pu ne pas inclure « 1 » (**Figure 6.c** et **Figure 6.d**). Par conséquent, l'OR*, s'il n'avait pas été biaisé par du biais de classement non différentiel, aurait pu *devenir* significativement différent de 1, et la conclusion aurait été alors différente. Cependant (d'où la raison du mot « potentiellement » ci-dessus), on aurait aussi pu être dans les situations des **Figure 6.e** et **Figure 6.f**, situations dans lesquelles l'OR* reste cette fois-ci non significativement différent de 1, bien qu'étant plus éloigné de « 1 » sans ce biais. La différence entre la **Figure 6.c** (ou **Figure 6.d**) et la **Figure 6.e** (ou **Figure 6.f**) correspond à l'importance du biais de classement. Le biais de classement non différentiel est plus important dans la **Figure 6.c** (ou **Figure 6.d**) que dans la **Figure 6.e** (ou **Figure 6.f**).

Légende

- (red hatched oval) ● OR* (IC95%) biaisé par du biais de classement non différentiel
 - (green oval) ● OR* (IC95%) avant d'être biaisé par du biais de classement non différentiel
- ↘ Impact du biais de classement non différentiel

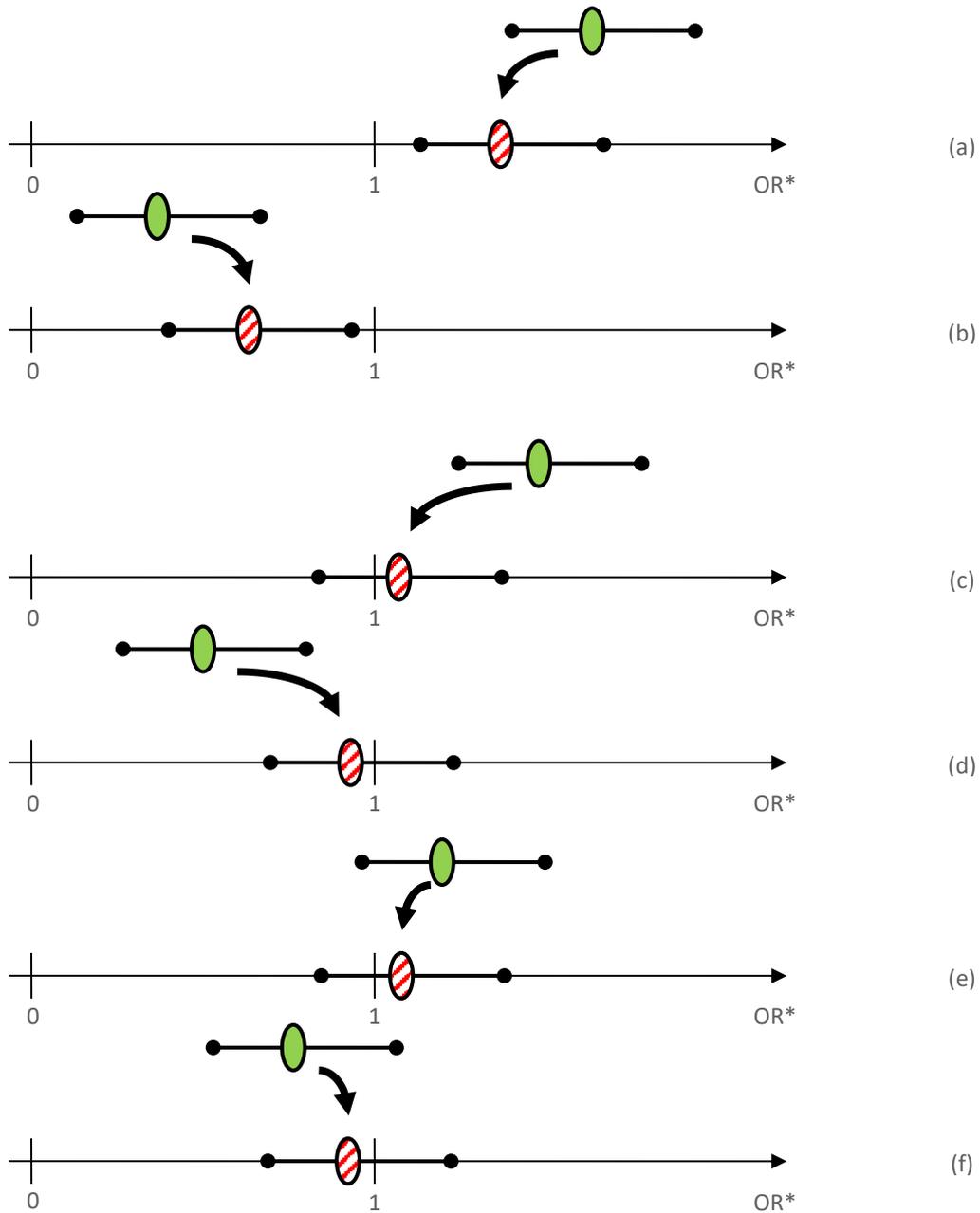


Figure 6. Impact du biais de classement non différentiel sur la valeur et la significativité d'un OR*.

D. Le biais de classement différentiel

1. Origine du biais de classement différentiel

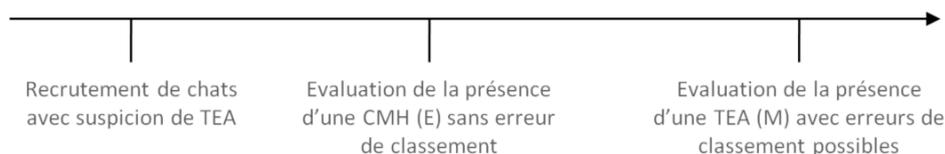
L'OR*_{E→M} est biaisé par du biais de classement différentiel si l'**erreur** de classement commise sur E est **dépendante** du statut M+/M- de l'animal.

Qu'est-ce qu'une erreur de classement commise sur E dépendante de M ? Cela signifie que la personne qui commet une erreur de classement sur E le fait d'une façon *a priori* différente selon que les animaux sont ou non « malades » : la connaissance du statut M+/M- de l'animal va *influencer* cette personne quand elle va évaluer l'exposition E de cet animal si cette personne suspecte un lien entre E et M. Ainsi, cette personne va avoir tendance à sur-estimer la présence de l'exposition E auprès des animaux dont ils savent qu'ils sont M+ et/ou à sous-estimer la présence de E auprès des animaux dont ils savent qu'ils sont M-.

Remarque : n'oubliez pas que dans le biais de classement, E et M sont interchangeable. Par conséquent, relisez ce que je viens d'écrire en remplaçant dans votre tête « E » par « M » et « M » par « E ».

2. Illustration sur un exemple

Supposons une étude clinique réalisée par des cliniciens suspectant fortement un lien entre la présence d'une cardiomyopathie hypertrophique (CMH) et celle d'une thrombo-embolie aortique (TEA) chez les chats. Ici, supposons que la CMH joue le rôle d'exposition (E) et la TEA joue le rôle de la maladie (M). Pour réaliser cette étude, les cliniciens recrutent des chats avec une suspicion de TEA. Parmi les chats recrutés dans l'étude, ces cliniciens commencent par évaluer la présence d'une CMH (évaluation de E). Supposons qu'il n'y ait aucune erreur dans cette évaluation. Supposons qu'*ensuite*, les cliniciens évaluent de façon subjective la présence d'une TEA à l'aide d'une échographie qu'ils réalisent eux-mêmes, avec cette fois-ci des erreurs possibles commises par ces cliniciens (cf. illustration ci-dessous).



Que peut-il se passer ? Dans la mesure où, (1) il peut exister des erreurs de classement sur la présence d'une TEA, (2) la présence d'une TEA est évaluée subjectivement, (3) au moment d'évaluer la présence d'une TEA, les cliniciens savent déjà que les chats ont ou non une CMH, et (4) ils suspectent déjà un lien entre TEA et CMH, alors il y a des risques pour qu'il se produise la chose suivante : une sur-estimation de la présence de TEA dans le cas où le chat a une CMH (E+), et/ou une sous-estimation de la présence de TEA dans le cas où le chat n'a pas de CMH (E-). En effet, les cliniciens ont tout à fait pu se laisser influencer par leur intuition quant à l'existence d'un lien entre CMH et TEA. Par conséquent, *dans les données recueillies*, la maladie M (la TEA) sera *encore plus* fréquente parmi les E+ que parmi les E- par rapport à ce qu'il peut exister dans la réalité, à cause des erreurs de classement commises sur M (la TEA).

3. Essai clinique en double aveugle et biais de classement différentiel

Dans un essai clinique, l'attribution de l'exposition d'intérêt E en double aveugle supprime le biais de classement différentiel dans l'estimation de l'OR*_{E→M}. En effet, les erreurs de classement sur M (forcément évaluée après E dans un essai clinique) ne peuvent pas dépendre de E puisque dans un essai en double aveugle, celui qui évalue M ne sait pas si l'animal est ou n'est pas exposé à l'exposition d'intérêt E. (Et par ailleurs, les erreurs de classement sur E¹² ne peuvent pas dépendre de M puisque, dans un essai clinique, M est évaluée au cours de suivi, donc forcément *après* avoir attribué E.) De même, dans un essai clinique au sein duquel le propriétaire de l'animal sait si son animal est exposé à E ou non, mais au sein duquel celui qui évalue M ne le sait pas, il n'y a pas de biais de classement différentiel. Ce qui compte dans la mise en aveugle au sein d'un essai clinique, pour supprimer les biais de classement différentiels, c'est que celui qui évalue M ne connaisse pas le statut E+/E- de l'animal.

4. Impact du biais de classement différentiel sur la valeur de l'OR ou du RR

L'impact du biais de classement différentiel sur l'OR* est tel qu'il augmente toujours la force de l'association (en fait, *quasiment* toujours¹³) : la valeur de l'OR* estimé et biaisé est plus éloignée de « 1 » que ne l'est la valeur de l'OR* réel correspondant dans la population.

5. Impact du biais de classement différentiel sur la significativité d'un OR ou d'un RR

Si un OR*, significativement différent de 1, est biaisé par du biais de classement différentiel, la présence de ce biais remet *potentiellement* en cause la significativité de l'OR*.

Explication : si l'OR* n'avait pas été biaisé par ce biais de classement différentiel, sa force aurait été plus faible, il se serait donc rapproché de « 1 », et son IC_{95%} aurait pu inclure « 1 » (**Figure 7.a** et **Figure 7.b**). Par conséquent, l'OR*, s'il n'avait pas été biaisé par du biais de classement différentiel, serait potentiellement *devenu* non significativement différent de 1, et la conclusion aurait été différente. Cependant (d'où le mot « potentiellement » écrit ci-dessus), sans ce biais de classement différentiel, on aurait aussi pu être dans la situation de la **Figure 7.c** ou de la **Figure 7.d**, sans remise en cause de la significativité de l'OR* (l'OR* serait resté significativement différent de 1 sans ce biais).

Si un OR*, non significativement différent de 1, est biaisé par du biais de classement différentiel, la présence de ce biais ne remet pas en cause la significativité de l'OR* (sans ce biais, l'OR* serait resté non significativement différent de 1).

Explication : si l'OR* n'avait pas été biaisé par ce biais de classement différentiel, sa force aurait été plus faible, il se serait donc rapproché de « 1 », et son IC_{95%} aurait toujours inclus « 1 » (**Figure 7.e** et **Figure 7.f**). Par conséquent, l'OR*, s'il n'avait pas été biaisé par du biais de classement différentiel, serait *resté* non significativement différent de 1.

¹² Considérer à tort un animal exposé alors qu'il ne l'est pas, ou inversement (dans un essai clinique, cette erreur serait, par exemple, de considérer à tort qu'un animal a reçu le traitement étudié, si cet animal n'a en pratique, jamais pris le traitement qu'il devait prendre au sein de l'essai).

¹³ Dans de très rares cas (hors programme), il diminue la force de l'association.

Légende

- (red hatched oval) ● OR* (IC95%) biaisé par du biais de classement différentiel
 - (green oval) ● OR* (IC95%) avant d'être biaisé par du biais de classement différentiel
- ↘ Impact du biais de classement différentiel

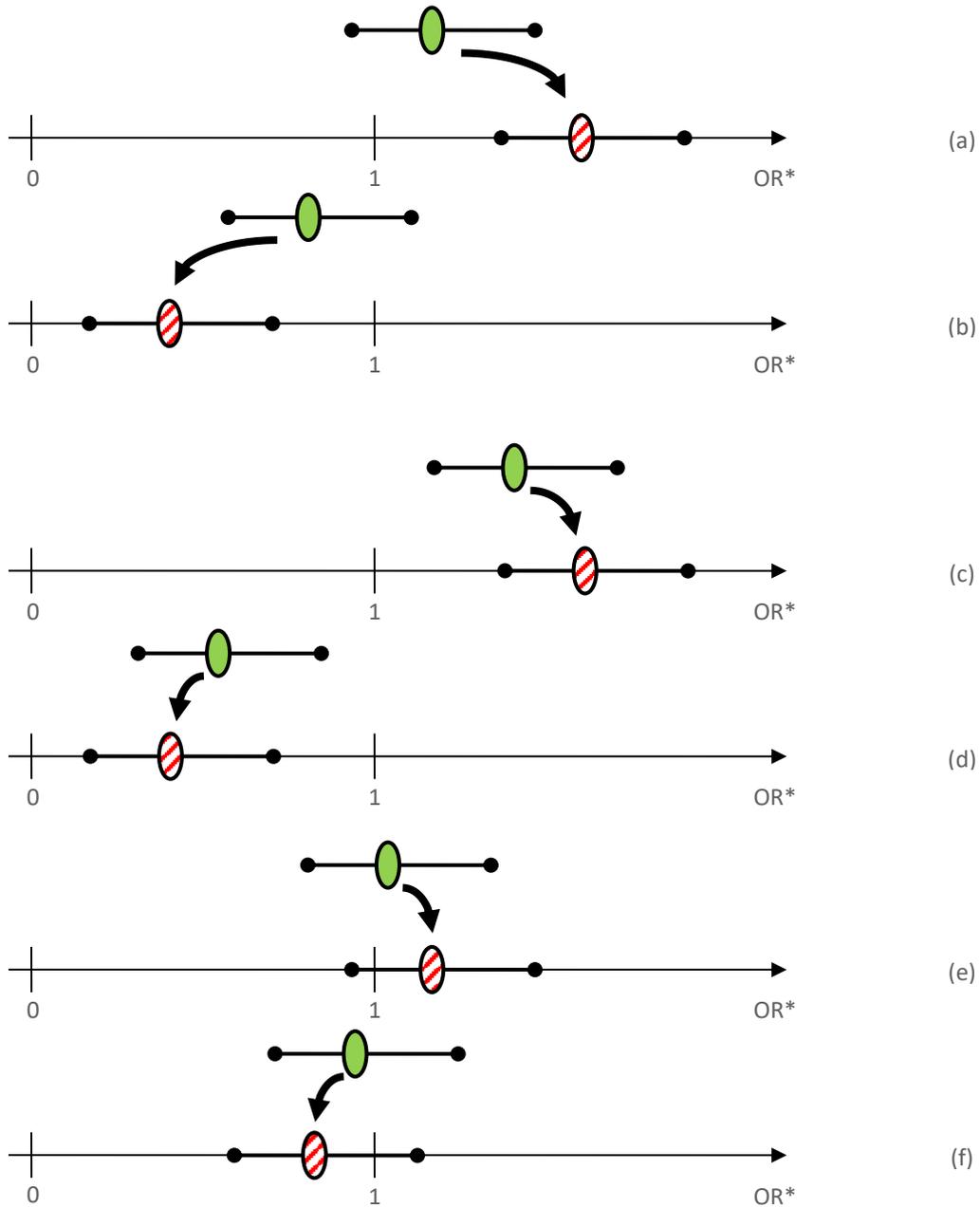


Figure 7. Impact du biais de classement différentiel sur la valeur et la significativité d'un OR*.

E. Discussion de l'importance des biais de classement

Le biais de classement (qu'il soit différentiel ou non différentiel) est d'autant plus important que les erreurs de classement sont nombreuses. Pour pouvoir affirmer que du biais de classement est certes présent mais n'est pas important, il faut argumenter pour convaincre que les erreurs de classement sont très peu fréquentes. L'« importance » du biais de classement est représentée par la longueur des flèches noires (représentant l'impact du biais) dans la **Figure 6** et dans la **Figure 7**.

F. Biais de classement différentiel ou non différentiel ?

1. Dangerosité du biais de classement différentiel

Comme nous avons pu le voir ci-dessus, si un OR^* , significativement différent de 1, est biaisé par du biais de classement non différentiel, la présence de ce biais ne remet pas en cause la significativité de l' OR^* . En revanche, si un OR^* , significativement différent de 1, est biaisé par du biais de classement différentiel, la présence de ce biais remet potentiellement en cause la significativité de l' OR^* . C'est la raison pour laquelle le biais de classement différentiel est beaucoup plus dangereux que le biais de classement non différentiel. En effet, si une association entre E et M est significative, et si cette association est biaisée par du biais de classement différentiel, on ne peut pas du tout être convaincu que cette association existe réellement dans la population, bien qu'elle soit significative dans l'échantillon. Ainsi, il est fondamental d'identifier la situation où du biais de classement différentiel existe.

Pour discuter du type de biais de classement entre le biais de classement différentiel et le biais de classement non différentiel, il est indispensable de savoir dans quel ordre ont été recueillies E et M (« E puis M », ou bien « M puis E » ?).

Je vous rappelle que la question du type de biais de classement (différentiel *versus* non différentiel) ne se pose que s'il existe des erreurs de classement sur E et/ou sur M. En effet, s'il n'y a aucune erreur de classement ni sur E ni sur M, alors l' OR^* ne sera pas biaisé par du biais de classement.

2. Logigramme pour évaluer la présence et le type de biais de classement

Le logigramme ci-dessous (**Figure 8**) permet de savoir si l' $OR^*_{E \rightarrow M}$ est biaisé par du biais de classement, et si oui, s'il est biaisé par du biais de classement non différentiel ou par du biais de classement différentiel. Comme indiqué dans la légende de la figure, « A » et « B » sont soit E soit M, de telle sorte que ce soit A qui soit évaluée avant B dans l'étude clinique. La suspicion de lien entre A et B est celle de la part de la personne qui évalue B.

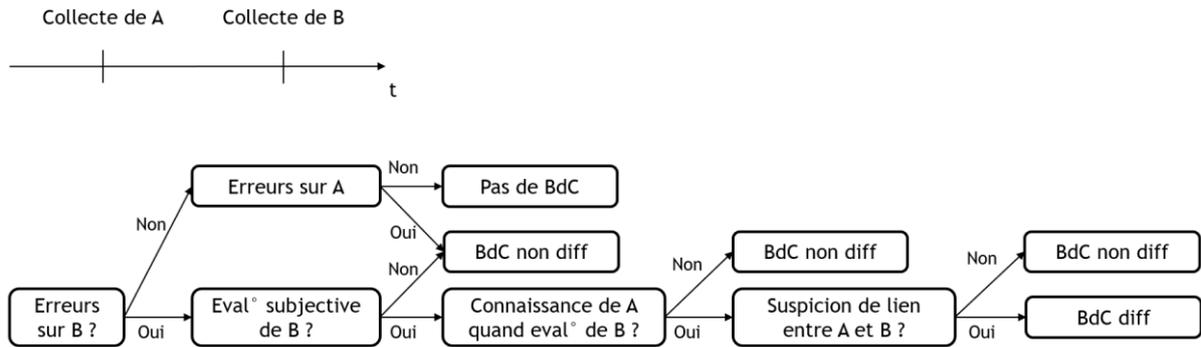


Figure 8. Logigramme permettant d'identifier la présence et le type de biais de classement dans l'étude de l'association entre une exposition d'intérêt E et une maladie étudiée M. « A » et « B » sont soit E soit M de telle sorte que ce soit A qui soit évaluée avant B dans l'étude clinique. « BdC » signifie « biais de classement ».

3. Priorité au biais de classement différentiel

Supposons que M soit évaluée après E (donc « B » est M, et « A » est E dans la **Figure 8**), et que les quatre conditions soient remplies pour que l'OR*_{E→M} soit biaisé par du biais de classement différentiel : erreurs de classement sur M, évaluation subjective de M, connaissance du statut E+/E- de l'animal au moment d'évaluer M, et suspicion d'un lien entre E et M par la personne qui évalue M. Supposons de plus qu'il existe des erreurs de classement sur E. Ces erreurs de classement sur E ne peuvent par définition pas dépendre de M, puisque E a été évaluée *avant* M. Ainsi, en plus d'être biaisée par du biais de classement différentiel, l'OR*_{E→M} sera biaisé *aussi* par du biais de classement non différentiel. Dans cette situation-là, le biais de classement différentiel est prioritaire dans la discussion des biais de classement, et on fera comme si l'OR*_{E→M} n'était biaisé *que* par du biais de classement différentiel.

VII. LE BIAIS DE CONFUSION ET LE FACTEUR DE CONFUSION

A. Théorie

1. Définition et origine du biais de confusion

Le biais de confusion est l'écart systématique entre l'OR*_{E→M} **estimé** dans l'échantillon et l'OR*_{E→M} **causal** dans la population, écart systématique provoqué par la non prise en compte d'au moins un facteur de confusion dans les analyses statistiques (cette non prise en compte est une erreur dans le protocole d'analyses statistiques).

2. Définition et identification d'un facteur de confusion

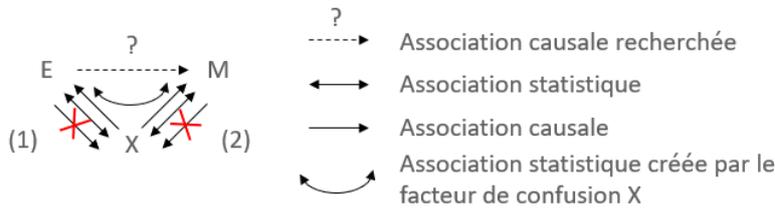
Un facteur de confusion est une exposition qui génère du biais de confusion si elle n'est pas prise en compte dans les analyses statistiques. Une exposition X est un facteur de confusion dans l'étude de l'association causale entre E et M si elle vérifie les deux critères ci-dessous :

Critère n°1 : X est associée statistiquement, ou causalement, à E sans en être une conséquence directe ;

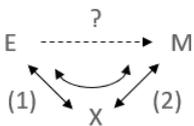
Critère n°2 : X est associée statistiquement, ou causalement, à M sans en être une conséquence directe.

Par conséquent, si une exposition X est une conséquence directe de E ou une conséquence directe de M, alors X ne sera pas un facteur de confusion dans l'étude de l'association causale entre E et M.

Le schéma suivant illustre ces deux critères :



Le schéma simplifié est celui-ci, et c'est celui que j'utiliserai dans la suite de ce polycopié :



Le biais de confusion généré par X est d'autant plus important que l'association entre X et E et/ou l'association entre X et M est/sont importante(s).

Une fois qu'un facteur de confusion est pris en compte, il ne génèrera plus de biais de confusion. Il est donc indispensable de **prendre en compte** dans les analyses statistiques tous les facteurs de confusion afin d'estimer un OR* dans l'échantillon qui soit le plus proche possible de l'OR* causal.

B. Biais de confusion et comparabilité clinique

1. Rappel sur « l'égalité des indicateurs est synonyme d'absence d'association »

Il est indispensable de se souvenir (cf. cours et TD de biostatistique) qu'une égalité des indicateurs comparés (par exemple, deux pourcentages ou deux moyennes) est synonyme d'absence d'association statistique entre deux caractères. J'ai repris la démonstration (car nous avons déjà beaucoup abordé cela dans le module de biostatistique de l'UC-213) dans l'annexe C de ce polycopié.

Le corollaire de « l'égalité des indicateurs est synonyme d'absence d'association » est le suivant : plus les deux indicateurs comparés (moyennes, médianes, pourcentages) sont proches, moins l'association statistique entre les deux caractères est importante.

2. De « égalité des indicateurs » à « comparabilité clinique »

En pratique, dans un échantillon, les indicateurs comparés ne seront jamais numériquement égaux. En revanche, ils pourront être considérés comme cliniquement voisins (par exemple, une moyenne d'âge de 8,9 ans chez des chiens castrés/stérilisés et une moyenne d'âge de 9,4 ans chez des chiens entiers), ou bien comme cliniquement différents (auquel cas, la différence entre les deux indicateurs sera jugée importante d'un point de vue clinique). (Je vous rassure tout de suite, je ne vous demanderai jamais de me dire si deux indicateurs comparés sont cliniquement voisins ou cliniquement différents. C'est moi qui vous le dirai dans les énoncés des TD et de l'examen.)

Lorsque, dans un échantillon, les deux indicateurs comparés (un par groupe) seront considérés comme cliniquement voisins, on dira que les deux groupes comparés de l'échantillon sont cliniquement comparables (dans le sens de « similaires »¹⁴). Et lorsque les deux indicateurs comparés (un par groupe) seront considérés comme cliniquement différents, on dira que les deux groupes comparés ne sont pas cliniquement comparables.

3. De « comparabilité clinique » à « absence d'association »

Ainsi, si l'on considère que deux indicateurs comparés (un par groupe) sont, dans l'échantillon, cliniquement comparables, alors on considèrera qu'il n'y a (quasiment) pas d'association entre les deux caractères dans l'échantillon. Pour reprendre l'exemple de l'âge et du statut castré/stérilisé, si dans l'échantillon la moyenne d'âge des chiens castrés/stérilisés est de 8,9 ans et si elle est de 9,4 ans chez les chiens entiers, et si l'on considère que ces deux moyennes sont cliniquement voisines, alors cela signifie que, dans l'échantillon, il n'y a (quasiment) pas d'association entre l'âge et le statut castré/stérilisé. Donc, le fait que, dans l'échantillon, les chiens castrés/stérilisés soient cliniquement comparables aux chiens entiers sur l'âge dans l'échantillon conduit au fait qu'il n'existe (quasiment) pas d'association entre l'âge et le statut castré/stérilisé dans l'échantillon.

4. De « comparabilité clinique » à « absence de biais de confusion »

a) Petite introduction

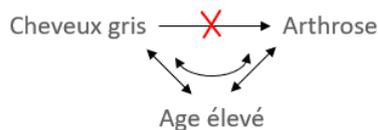
Une fois que vous avez compris que « comparabilité clinique » était synonyme de « absence d'association », alors il devient (plus) facile de comprendre que si, dans l'échantillon, les animaux exposés à E (E+) sont cliniquement comparables aux animaux non exposés à E (E-) sur X, cela signifie que, dans l'échantillon, E et X ne sont (quasiment) pas associées. Et ainsi, le critère n°1 pour que X soit un facteur de confusion n'est pas vérifié, et donc X ne sera pas un facteur de confusion dans l'étude de l'association causale entre E et M.

b) Exemple de l'association entre cheveux gris et arthrose

Comme vous le savez, dans la population humaine, il n'existe aucun lien *causal* entre la présence de cheveux gris (E) et la présence d'arthrose (M) (\Leftrightarrow les cheveux gris ne sont ni une cause ni une conséquence d'arthrose). Et pourtant, dans cette même population, il existe une association *statistique* réelle entre ces deux caractéristiques : la proportion de personnes qui ont de l'arthrose parmi les personnes qui ont les cheveux gris est bien supérieure à la proportion d'arthrose parmi les personnes qui n'ont pas les cheveux gris. (Et là, vous pouvez revenir sur la **Figure 3.d**, page 8, qui aurait pu vous poser initialement quelques soucis !)

Pour quelle raison ? Parce que l'âge est un facteur de confusion dans l'étude de l'association causale entre les cheveux gris et l'arthrose. L'âge répond en effet aux deux critères ci-dessus : (1) un âge élevé (X) est associé de façon cliniquement importante aux cheveux gris (E) sans en être une conséquence (les personnes qui ont les cheveux gris sont *en moyenne* beaucoup plus âgées que celles qui n'ont pas les cheveux gris), et (2) un âge élevé (X) est associé de façon cliniquement importante à l'arthrose (M) sans en être une conséquence (les personnes qui ont de l'arthrose sont *en moyenne* beaucoup plus âgées que celles qui n'ont pas d'arthrose). Nous retrouvons le schéma ci-dessous.

¹⁴ Attention, dans tout le module d'EpiClin, l'adjectif « comparable » signifiera *toujours* « similaire » ou « voisin ».



Pourquoi l'arthrose est-elle réellement plus fréquemment rencontrée parmi les personnes avec les cheveux gris que parmi les personnes sans cheveux gris, alors qu'il n'existe pas de lien causal ? On peut répondre à cette question en utilisant la notion de « comparabilité clinique » : « parce que les personnes qui ont les cheveux gris ne sont pas cliniquement comparables aux personnes qui n'ont pas les cheveux gris sur une exposition (l'âge) qui est par ailleurs associée à l'arthrose ». En effet, les personnes qui ont les cheveux gris ont un âge moyen cliniquement bien plus élevé que celles qui n'ont pas les cheveux gris (\Leftrightarrow « non comparabilité clinique » sur l'âge entre les personnes qui ont les cheveux gris et les personnes qui n'ont pas les cheveux gris).

c) Généralisation

Une exposition X est un facteur de confusion dans l'étude de l'association causale entre une exposition d'intérêt E et une maladie M si (1) les animaux exposés à E ne sont pas cliniquement comparables aux animaux non exposés à E sur cette exposition X (qui ne doit pas être une conséquence directe de E) et si (2) X est par ailleurs associée à M sans en être une conséquence directe. Le point (1) dans la phrase précédente correspond au critère n°1 pour identifier un facteur de confusion, et le point (2) correspond au critère n°2.

C. Vérification des deux critères pour identifier un facteur de confusion

Je dois tout d'abord vous rappeler une chose fondamentale¹⁵. Un test statistique ne sert qu'à une seule chose : à faire de l'inférence à partir d'une association significative ($p \leq 0,05$). Un test statistique ne dit pas si une association est forte (ou faible) *dans un échantillon*. En effet, une association peut être forte dans un échantillon, et pourtant non significative (parce que la taille de l'échantillon est faible). Une association peut être faible dans un échantillon, et pourtant significative (parce que la taille de l'échantillon est importante). Ce petit rappel de l'UC-0213 effectué, je peux donc poursuivre 😊.

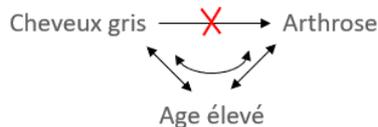
La vérification du critère n°1 doit être exclusivement « clinique ». C'est-à-dire que pour savoir si le critère n°1 est vérifié ou non, il faut utiliser son cerveau et ses yeux de clinicien. Si le degré de signification est utilisé pour vérifier le critère n°1, c'est une belle erreur, car on peut très souvent conclure *à tort* que les deux groupes E+ et E- sont comparables sur X.

La vérification du critère n°2 doit aussi être « clinique ». Cela dit (et pour des raisons qui dépassent largement le programme de l'UC-0324), en pratique, il n'est pas rare que l'on utilise la valeur du degré de signification testant l'association entre X et M dans l'échantillon. Lorsque cela est fait, les auteurs peuvent considérer que le critère n°2 est vérifié si p est inférieur ou égal à un seuil fixé à l'avance (qui n'a rien à voir avec α – le risque d'erreur α reste toujours fixé à 0,05 pour dire qu'une association est « significative »). En fonction des études, ce seuil peut prendre la valeur de 0,25, 0,20, ou 0,05.

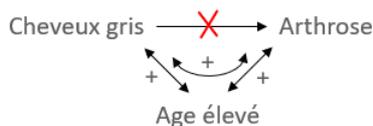
¹⁵ Et oui, encore une... 😞

D. Sens du biais de confusion (« positif » ou « négatif »)

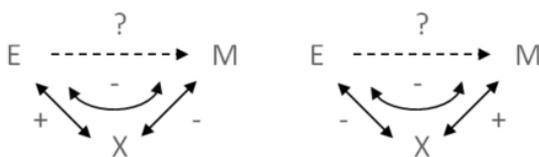
Avant de lire cette partie VII.D, je vous recommande de relire la partie IV.F « Association « positive, « négative », et « forte » : signification et illustration avec un OR ou un RR » (page 22). Dans le schéma ci-dessous, nous voyons que les cheveux gris et l'arthrose sont statistiquement associés à cause du biais de confusion généré par l'âge (présence d'une flèche doublement orientée, et incurvée, entre « cheveux gris » et « arthrose »).



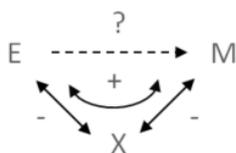
L'étape d'après est de mettre un sens (« + » ou « - ») à cette flèche incurvée et doublement orientée. Pour pouvoir le faire, il faut mettre un sens aux deux flèches qui génèrent le biais de confusion : celle entre E et X, et celle entre X et M. Puisque les personnes qui ont les cheveux gris sont en moyenne *plus* âgées que celles qui n'ont pas les cheveux gris, on met un « + » entre « Cheveux gris » et « Age élevé ». Puisque les personnes qui ont de l'arthrose sont en moyenne *plus* âgées que celles qui n'ont pas d'arthrose, on met un « + » entre « Arthrose » et « Age élevé ». C'est l'existence de ces deux phénomènes « positifs » qui a conduit au fait que l'on observe que l'arthrose est *plus* fréquente parmi les personnes qui ont les cheveux gris que parmi les personnes qui n'ont pas les cheveux gris (donc un « + » entre « Cheveux gris » et « Arthrose »). Le biais de confusion est alors dit « positif » :



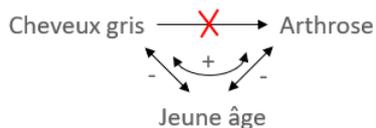
Nous verrons en TD que lorsque E et X sont associées négativement et lorsque X et M sont associées positivement, ou bien lorsque E et X sont associées positivement et lorsque X et M sont associées négativement, le biais de confusion qui en résulte est alors « négatif » :



Enfin (mais c'est hors programme), lorsque E et X sont associées négativement et lorsque X et M sont aussi associées négativement, le biais de confusion qui en résulte est alors « positif » :



Pour vous en convaincre, reprenez le schéma ci-dessus pour cheveux gris et arthrose, et remplacez « Age élevé » par « Jeune âge ». Voici ce que vous obtenez :



En fait, la règle du signe du biais de confusion est exactement la même que celle de la multiplication : multiplier deux nombres positifs ou deux nombres négatifs conduit à un résultat positif, et multiplier un nombre positif par un nombre négatif conduit à un résultat négatif.

E. Commentaires

Le biais de confusion *provient* d'une erreur de la part de la personne qui analyse et interprète les résultats. Cette erreur, magistrale ☺, est celle de la non prise en compte d'au moins un facteur de confusion avant de faire de l'inférence causale. Nous allons voir dans la partie VIII comment empêcher ou réduire, grâce au protocole, la présence de facteurs de confusion, ou bien comment les prendre en compte dans les analyses statistiques si le protocole d'étude n'a pas pu empêcher l'existence de facteurs de confusion.

Ces facteurs de confusion existent car, dans la « nature » (au sens le plus général qui soit), beaucoup de choses sont associées entre elles, et ces associations entraînent d'autres associations par « réactions en chaîne » (d'où la raison pour laquelle le biais de confusion est inhérent dans les études observationnelles qui représentent « la vraie vie » ou la « nature »).

Par ailleurs, pour apprécier l'existence d'un phénomène de confusion, il faut penser « population », et non pas « individu » : bien sûr que ce n'est pas *parce qu'*une personne est âgée qu'elle *va* avoir les cheveux gris, ni qu'elle *va* avoir de l'arthrose. Mais, *en général*, il est plus fréquent de présenter une arthrose et des cheveux gris lorsqu'une personne est âgée que lorsqu'une personne est jeune.

F. Définition d'association brute

Une association **brute** entre E et M est une association qui ne prend en compte *aucune* exposition dans son calcul ou son estimation.

Dans la population, nous observons¹⁶ une association brute entre « cheveux gris » et « arthrose » (l'arthrose est en effet réellement plus fréquente parmi les personnes qui ont les cheveux gris que parmi les personnes qui n'ont pas les cheveux gris). Cette association est « brute » car elle ne prend en compte aucune exposition (en particulier, elle ne prend pas en compte l'âge). Si l'on devait quantifier cette association brute par un OR*, cet OR*_{cheveux gris → arthrose} **brut** serait > 1 (si ce n'est pas évidemment, alors relisez la partie IV.C.2, page 19). Alors que l'OR*_{cheveux gris → arthrose} **causal** est quant à lui égal à 1 car il n'y a aucune association causale entre « cheveux gris » et « arthrose ».

L'OR* brut estimé dans un échantillon est d'autant plus éloigné de l'OR* causal que les facteurs de confusion sont nombreux et/ou importants sans avoir été pris en compte dans les

¹⁶ Si nous étions dotés d'un drone capable de repérer la couleur des cheveux et la présence d'arthrose, survolant le monde entier !

analyses statistiques (il ne faut pas oublier que l'OR* brut est éloigné de l'OR* causal lorsqu'il est aussi biaisé par du biais de classement ou de sélection).

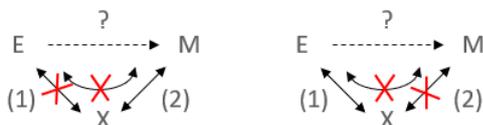
Dans la recherche des facteurs de risque ou protecteurs de la maladie étudiée en utilisant une étude non randomisée, partez *toujours* du principe qu'il existe forcément des facteurs de confusion. Par conséquent, partez toujours du principe que, dans une étude non randomisée, l'OR* brut estimé dans l'échantillon est forcément biaisé par du biais de confusion, et qu'il est donc forcément éloigné de l'OR* causal.

VIII. COMMENT EMPECHER OU REDUIRE LES BIAIS DE CONFUSION

A. Introduction

Il y a deux grandes façons d'empêcher ou réduire les biais de confusion : grâce au **protocole** de l'étude, ou grâce aux **analyses statistiques** (Braga et al., 2012). Dans les analyses statistiques, on dit que l'on « prend en compte » un facteur de confusion. « Prendre en compte un facteur de confusion » signifie « empêcher qu'un facteur de confusion ne joue son rôle de confusion ».

Théoriquement, pour empêcher la présence de biais de confusion dû à une exposition X dans l'estimation de l'association entre E et M, il faut empêcher l'association entre E et X (empêcher que le critère n°1 ne soit vérifié) ou empêcher l'association entre X et M (empêcher que le critère n°2 ne soit vérifié ; cf. schémas ci-dessous).



Si, à l'issue des analyses statistiques, il existe toujours du biais de confusion dans l'étude de l'association causale entre E et M, on dit qu'il existe du biais de confusion **résiduel**. A ce sujet, je vous cite un passage de l'article Kyriacou et Lewis (Kyriacou and Lewis, 2016) : « du biais de confusion résiduel peut persister dans les études cliniques en dépit du protocole de l'étude et des analyses statistiques dont l'objectif était justement d'éliminer le biais de confusion. Ce biais de confusion résiduel peut se produire lorsque la randomisation échoue à rendre comparables les deux groupes de traitements sur les expositions associées par ailleurs à la maladie étudiée (notamment dans la situation des essais cliniques avec des groupes de petites tailles). Il peut se produire surtout dans les études observationnelles quand les analyses statistiques ne prennent pas en compte les facteurs de confusion. »

Pour finir cette courte introduction, je vous cite un passage de Ewald et Coll. (Ewald et al., 2019) : « L'essai clinique randomisé (ECR) est généralement considéré comme le type d'étude de référence pour identifier les relations causales afin de choisir la bonne option de traitement (Guyatt et al., 2008). Cependant, les ECR sont pas disponibles pour de nombreuses questions importantes sur les soins de santé. [...] Dans le "monde réel", le choix du traitement n'est pas aléatoire, mais dépend souvent de facteurs spécifiques, tels que des comorbidités ou des préférences personnelles des patients et du personnel soignant, qui sont par ailleurs associés à l'évolution de l'état de santé (biais de confusion par indication (Kyriacou and Lewis, 2016)). [...] Plusieurs méthodes sont utilisées dans des études observationnelles pour tenter de

résoudre les problèmes de biais de confusion, par exemple, les méthodes de régression multivariées ou l'utilisation des scores de propension (Lonjon et al., 2014). »

B. Empêcher ou réduire les biais de confusion grâce au protocole d'étude

1. Attribution aléatoire (randomisation) de l'exposition d'intérêt

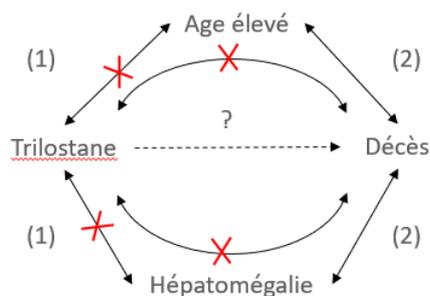
a) En théorie

Le protocole d'étude qui empêche ou réduit les biais de confusion est celui de l'essai clinique randomisé (cf. partie III.B.3, page 13). Attribuer aléatoirement (\Leftrightarrow randomiser) l'exposition d'intérêt E est la méthode la plus efficace pour éliminer le *maximum* de biais de confusion (en théorie, elle les élimine tous).

Dans tout ce qui suit dans cette partie, et pour être plus proche de ce qu'il se passe en pratique, je vais me mettre dans la situation où les animaux exposés à l'exposition d'intérêt (E) seront les animaux traités au traitement étudié (E+), et les animaux non exposés à l'exposition d'intérêt seront les animaux sous placebo (E-). Bien entendu, tout ce qui suit est valable pour tout « type de traitement » (par exemple, un groupe recevant un traitement médical tandis qu'un autre recevant une intervention chirurgicale, avec pour objectif de montrer que le traitement médical est plus efficace que le traitement chirurgical).

Pourquoi la randomisation du type de traitement supprime-t-elle, en théorie, *tous* les biais de confusion ? Pour répondre à cette question, supposons une étude clinique mise en place pour montrer que, chez les chiens atteints d'un syndrome de Cushing, la mise sous traitement Trilostane au moment du diagnostic du syndrome de Cushing prolonge la vie de ces chiens par rapport à une mise sous placebo (la « maladie » M étudiée est donc le décès). Supposons par ailleurs qu'un âge avancé ainsi que la présence d'une hépatomégalie soient tous les deux des facteurs de risque de décès.

Supposons que l'on recrute 200 chiens atteints d'un syndrome de Cushing pour tester l'efficacité du Trilostane. Supposons que parmi ces 200 chiens recrutés le jour du diagnostic, la moyenne d'âge soit de 10 ans, et 20% de ces 200 chiens aient présenté une hépatomégalie. A partir de ces 200 chiens, constituons au hasard deux groupes de 100 chiens, dont le premier reçoit le Trilostane (G_T) et le second reçoit un placebo (G_P). (Le type de traitement a ainsi été randomisé.) En toute théorie (c'est-à-dire, si le hasard fait *parfaitement* bien les choses), la moyenne d'âge dans le groupe G_T est *égale* à celle dans le groupe G_P , et elle est *égale* à 10 ans ; de même, le pourcentage d'hépatomégalie dans le groupe G_T est *égal* à celui dans le groupe G_P , et il est *égal* à 20%. En toute théorie, le hasard fait en effet très bien les choses : il équilibre parfaitement les deux groupes. Autrement dit, il rend les deux groupes G_T et G_P *parfaitement comparables* sur l'âge et sur la présence d'hépatomégalie. Dans une telle situation (idéale), il n'existe donc pas d'association du tout, parmi les 200 chiens de l'étude, entre le type de traitement (Trilostane *versus* placebo) et l'âge des chiens, ni entre le type de traitement (Trilostane *versus* placebo) et la présence d'hépatomégalie (égalité des deux moyennes d'âge et égalité des deux pourcentages d'hépatomégalie entre le groupe G_T et le groupe G_P). Voici donc le schéma :



Bien que le critère n°2 pour qu'une exposition soit un facteur de confusion soit toujours vérifié pour l'âge et l'hépatomégalie (en effet, ce n'est pas parce que le Trilostane a été randomisé que l'effet de l'âge ou de l'hépatomégalie sur le décès a disparu !), le critère n°1 a quant à lui disparu : la randomisation du Trilostane a empêché (en théorie) la présence d'association avec l'âge ou l'hépatomégalie. Ainsi, en théorie, l'âge et l'hépatomégalie ne peuvent plus être des facteurs de confusion (et ne joueront donc pas de rôle de confusion dans l'estimation de l'association entre le type de traitement et le décès des chiens atteints d'un syndrome de Cushing).

Et la force de la randomisation (et donc celle des essais cliniques randomisés), c'est que le raisonnement que l'on vient de tenir pour l'âge et l'hépatomégalie est valable pour *toutes* les expositions liées, causalement ou non, au décès (y compris celles dont on n'a pas encore idée aujourd'hui), donc pour *toutes* les expositions jouant potentiellement un rôle de confusion. En théorie, avec une randomisation de l'exposition d'intérêt, *aucune* exposition ne sera un facteur de confusion.

Par conséquent, en théorie, l'estimation dans un essai clinique randomisé de l'OR* brut quantifiant l'association entre le type de traitement et la survenue de la maladie M étudiée ne sera pas biaisée par du biais de confusion, et aura par conséquent des chances d'être proche de l'OR* causal (sous réserve d'absence de biais de classement et de biais de sélection). Voilà la raison pour laquelle il est beaucoup plus facile de faire de l'inférence causale à partir d'un essai clinique *randomisé* que de la faire à partir d'une étude non randomisée (essai clinique non randomisé, ou n'importe quelle étude observationnelle).

Attention, la randomisation garantit en théorie l'absence de biais de confusion, mais elle ne garantit absolument pas du tout l'absence de biais de classement. En effet, l'origine d'un biais de classement étant une erreur de classement (par exemple, on dit à tort qu'un animal est M+ alors qu'il ne l'est pas). Or, le fait de randomiser l'exposition d'intérêt E n'empêche pas l'erreur de classement faite par le vétérinaire ou le propriétaire, ni lors de l'évaluation de la présence de M (par examen clinique ou par auto-questionnaire), ni lors de l'évaluation de E (dire que l'animal a effectivement reçu le traitement d'intérêt ou n'a effectivement pas reçu le traitement d'intérêt) !

b) En pratique

En théorie, tout marche très bien. Mais en pratique, il se peut tout à fait que le hasard (dans l'attribution aléatoire de l'exposition d'intérêt) fasse mal les choses (Nguyen and Xie, 2021) ! En effet, en théorie, s'il y a 20% d'hépatomégalie parmi les 200 chiens avant randomisation (soit n = 40 chiens avec hépatomégalie), il devrait y avoir 20% (pile) d'hépatomégalie dans chacun des deux groupes (soit 20 chiens avec hépatomégalie par groupe). En pratique, si le hasard faisait mal les choses, sur ces 40 chiens avec hépatomégalie, 5 (au lieu de 20)

pourraient recevoir du Trilostane (soit 5% des chiens sous Trilostane) et 35 (au lieu de 20) pourraient recevoir du placebo (soit 35% des chiens sous placebo). Dans une telle situation, le pourcentage de chiens avec hépatomégalie est cliniquement beaucoup plus important parmi les chiens sous placebo (35%) que parmi ceux sous Trilostane (5%), le critère n°1 devient par conséquent vérifié, et l'hépatomégalie devient un facteur de confusion (parce que le critère n°2 était déjà vérifié). Dans une telle situation, l'autre façon d'identifier que l'hépatomégalie est un facteur de confusion, c'est de se rendre compte que les chiens sous Trilostane ne sont pas cliniquement comparables aux chiens placebo sur la présence d'hépatomégalie (5% *versus* 35% d'hépatomégalie, respectivement), elle-même associée au décès des chiens.

Si le hasard fait mal les choses sur certaines expositions associées par ailleurs à M, ces expositions seront des facteurs de confusion qu'il faudra prendre en compte dans les analyses statistiques.

c) Impact de la taille de l'échantillon sur les « bienfaits » de la randomisation

Comme on vient de le voir, en pratique, le hasard peut mal faire les choses, c'est-à-dire qu'il peut conduire au fait que les deux groupes (le groupe traité et le groupe sous placebo par exemple) soient cliniquement non comparables sur une exposition X par ailleurs associée à M (cette exposition X sera donc un facteur de confusion). Et il faut savoir que plus la taille de l'échantillon (ou la taille de chacun des deux groupes) est faible, plus il y a de risques pour que le hasard fasse mal les choses au point de rendre non comparables, cliniquement parlant, les deux groupes. Autrement dit, la randomisation garantit d'autant plus la comparabilité clinique entre les deux groupes (donc l'absence de biais de confusion) que la taille des deux groupes est importante. Autrement dit encore, en pratique, une étude randomisée garantit d'autant *moins* l'absence de biais de confusion que la taille des groupes est *faible*.

2. Autres protocoles d'études empêchant ou réduisant les biais de confusion

Les autres protocoles d'études pour empêcher ou réduire les biais de confusion ne sont pas au programme de l'UC-0324, car ce sont des protocoles que l'on ne rencontre quasiment jamais dans les études cliniques vétérinaires. Vous verrez quelques-uns de ces protocoles dans le cadre de l'UC-0322. En voici deux parmi d'autres :

- L'appariement (par exemple, les enquêtes cas-témoins appariée) ;
- La restriction.

C. Empêcher ou réduire les biais de confusion grâce aux analyses statistiques : stratification et ajustement

1. La stratification

a) Description de la méthode

Supposons que X soit un facteur de confusion binaire. Stratifier sur X, c'est quantifier l'association entre E et M dans chacune des deux strates de X (**Figure 9**), à l'aide d'un OR*.

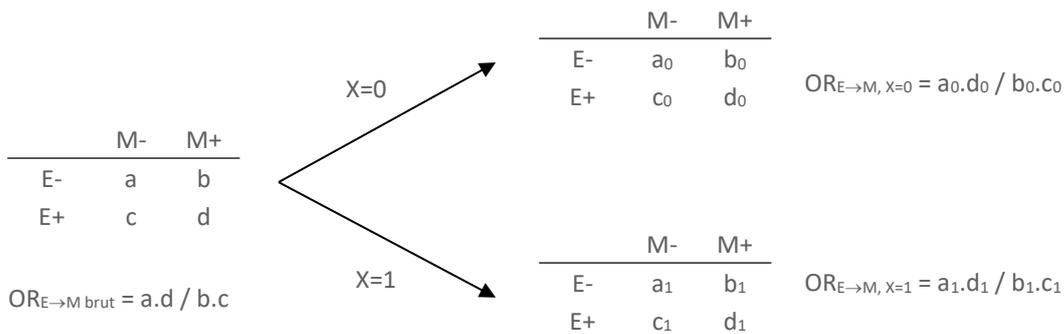
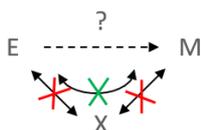


Figure 9. Schéma illustrant la stratification sur un facteur de confusion X binaire, où E est l'exposition d'intérêt et M la maladie. Par notation, $OR_{E \rightarrow M, X=i} = OR_{E \rightarrow M}$ stratifié sur X et calculé dans la strate où $X = i$. La figure présente la stratification avec des OR, mais la démarche reste bien évidemment la même avec des RR.

Dans la **Figure 9**, on peut voir que l'association brute (quantifiée par l' $OR_{E \rightarrow M}$ brut) ne prend effectivement pas en compte X dans le calcul, alors que les deux associations stratifiées (quantifiées par les $OR_{E \rightarrow M, X=0}$ et $OR_{E \rightarrow M, X=1}$) prennent en compte X, puisqu'elles sont calculées dans chacune des strates de X.

La stratification sur X conduit au fait que X, initialement un facteur de confusion, ne peut plus jouer de rôle de confusion dans l'estimation de l'association entre E et M. Pourquoi ? Parce que dans chacune des strates de X, il n'y a plus d'association ni entre E et X, ni entre X et M : X ne vérifie plus, *après* stratification, les deux critères d'identification de facteur de confusion (cf. page 35) qu'elle vérifiait *avant* stratification. En effet, dans la strate où $X=0$, tous les animaux ont la valeur « 0 » pour l'exposition X ; de même, dans la strate où $X=1$, tous les animaux ont la valeur « 1 » pour X. Par conséquent, dans chacune des strates de X, il ne peut plus y avoir d'association ni entre E et X ni entre X et M¹⁷. La stratification a mis les deux croix rouges sur les flèches ci-dessous (la croix verte étant la conséquence des deux croix rouges) :



¹⁷ Prenons un exemple : pensez-vous que dans un échantillon de femmes uniquement, il existe une association entre le fait de fumer et le sexe des individus ? C'est peut-être vrai dans la population générale, mais ça ne l'est absolument pas *dans cet échantillon*, puisqu'il n'est constitué que de femmes.

Si X est un facteur de confusion, les OR* stratifiés sur X et estimés dans l'échantillon sont donc plus proches de l'OR* causal que ne l'était l'OR* brut estimé dans l'échantillon, car chacun de ces OR* stratifiés n'est plus biaisé par du biais de confusion dû à X.

D'un point de vue « comparabilité clinique », dans chacune des strates de X, les animaux exposés à E sont parfaitement comparables sur X aux animaux non exposés à E puisque X ne prend qu'une seule valeur dans chacune des strates. Cette parfaite comparabilité clinique sur X conduit au fait que X ne joue plus de rôle de facteur de confusion après stratification sur X.

b) Illustration graphique

Supposons une maladie M et une exposition E toutes deux quantitatives, et une exposition X binaire. Supposons 50 animaux sur lesquels on mesure M, E, et X.

La **Figure 10.a** représente l'association brute entre E et M. On peut voir que, sans prendre en compte X, globalement, lorsque E augmente, M augmente très légèrement (pente = +0,1). En revanche, si l'on stratifie l'association entre E et M sur X, c'est-à-dire que l'on trace la droite de régression séparément parmi les animaux pour lesquels X=0 (droite rouge) et parmi les animaux pour lesquels X=1 (droite verte ; **Figure 10.b**), on peut voir que dans chacune de ces deux strates, l'association entre E et M est beaucoup plus forte qu'elle ne l'était en brut (pentes = -1,2). Par conséquent, vous pouvez vous rendre compte de quelque chose qui est contre-intuitif : l'association brute (globale) n'est pas la moyenne des associations dans chacun des sous-échantillons (ici, des « strates ») qui constituent l'échantillon global !

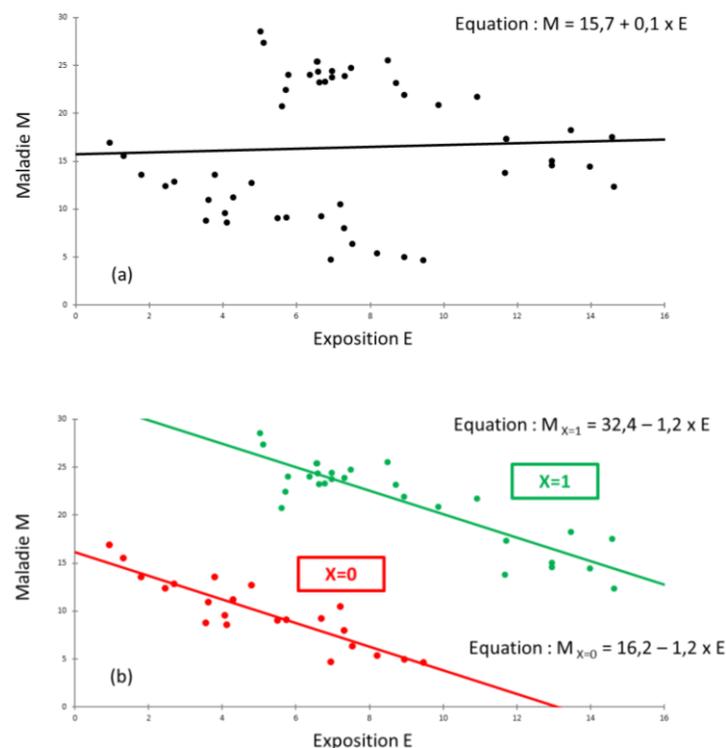


Figure 10. Les deux figures (a) et (b) représentent les mêmes données, et la pente de la droite de régression quantifie l'association entre l'exposition E et la maladie M. Dans le panel (a), toutes les données sont représentées en noir sans distinction des animaux exposés à X ou non. Dans le panel (b), les animaux exposés et non exposés à X sont représentés avec une couleur différente (rouge et vert). L'association brute, avant stratification, entre E et M est très faible (pente = +0,1 ; (a)). Les deux associations stratifiées sur X sont beaucoup plus fortes (pentes = -1,2 ; (b)). (Les échelles sont identiques.)

La prise en compte de X dans l'association entre E et M (par stratification ici) a modifié l'association entre E et M par rapport à la situation sans prise en compte de X (modification des pentes entre avant et après stratification).

c) Inconvénient de la stratification

L'inconvénient majeur de la stratification est qu'après stratification sur un facteur de confusion X binaire, on obtient deux $OR_{E \rightarrow M}^*$ (un par strate de X). Lequel choisir ? Conserver les deux et présenter les deux ? (Sachant bien entendu qu'il n'y a aucune chance pour qu'ils soient tous les deux exactement égaux, ce serait trop simple !) Le choix serait cornélien...

2. L'ajustement

a) Démarche de calcul

L'ajustement pallie l'inconvénient de la stratification précédemment cité. Par ailleurs, l'ajustement est la méthode la plus fréquemment employée dans le domaine de l'épidémiologie clinique pour éliminer du biais de confusion.

Il existe plusieurs méthodes pour calculer l' $OR_{E \rightarrow M}^*$ ajusté sur un facteur de confusion X. Une première méthode de calcul est celle qui suit la stratification. Dans ce cas-là, l' $OR_{E \rightarrow M}^*$ ajusté sur X est une moyenne pondérée des $OR_{E \rightarrow M}^*$ calculés dans chacune des strates de X.

En pratique, soit $OR_{E \rightarrow M, X=0}^*$ et $OR_{E \rightarrow M, X=1}^*$ les deux $OR_{E \rightarrow M}^*$ calculés dans chacune des deux strates de X. L' OR^* ajusté sera calculé plus ou moins de la façon suivante¹⁸ (calculs hors programme de l'UC-0324). :

$$OR_{E \rightarrow M}^* \text{ ajusté sur } X = \frac{w_0 \times OR_{E \rightarrow M, X=0}^* + w_1 \times OR_{E \rightarrow M, X=1}^*}{w_0 + w_1}$$

Les coefficients de pondération w_0 et w_1 dépendent du nombre d'animaux dans chacune des strates de X (w_0 et w_1 sont toujours > 0) : plus une strate contiendra un nombre important d'animaux, plus le coefficient de pondération sera élevé¹⁹.

La condition pour pouvoir calculer un $OR_{E \rightarrow M}^*$ ajusté sur X est la suivante : les $OR_{E \rightarrow M}^*$ calculés dans chacune des strates de X doivent être considérés comme « voisins ». Cette vérification n'est pas au programme de l'UC-0324, et nous partirons toujours du principe que les $OR_{E \rightarrow M}^*$ stratifiés sont voisins.

¹⁸ Les calculs sont bien entendu plus compliqués, mais l'objectif n'est pas que vous sachiez calculer un OR^* ajusté, mais que vous compreniez la démarche pour le calculer.

¹⁹ Explication (hors programme) : chaque poids w_i est proportionnel à la précision de l' $OR_{E \rightarrow M, X=i}^*$. Or (cf. cours et TD de Biostatistique de l'UC-0213), la précision d'une estimation est quantifiée par la « standard error » de l'estimation. Et vous voyez dans les parties IV.B.3 et IV.C.3 que la « standard error » d'un OR^* dépend du nombre d'animaux dans chacune des 4 cases du tableau. Par conséquent, les poids w_i dépendent des effectifs dans chacune des strates, dans le sens suivant : plus les effectifs dans une strate sont faibles, plus la « standard error » de l' OR^* stratifié est grande, moins cet OR^* stratifié sera précis, et donc moins cet OR^* stratifié aura de poids dans le calcul de l' OR^* ajusté sur X.

b) Mise en évidence d'un rôle de confusion important

La confrontation entre l'OR*_{E→M} brut et l'OR*_{E→M} ajusté sur X apporte une information très importante : plus ces deux OR* seront éloignés l'un de l'autre, et plus X aura joué un rôle de confusion important (et par conséquent, plus il sera indispensable d'ajuster sur l'exposition X pour empêcher le biais de confusion qu'elle provoquerait si elle n'était pas prise en compte).

Pour quantifier le rôle de confusion qu'a pu jouer une exposition X dans l'estimation de l'association entre E et M, on calcule la différence relative entre l'OR*_{E→M} brut et l'OR*_{E→M} ajusté sur X (Δ), en utilisant la formule (à savoir) suivante :

$$\Delta \text{ (exprimée en \%)} = \frac{|OR_{E \rightarrow M}^* \text{ ajusté sur } X - OR_{E \rightarrow M}^* \text{ brut}|}{OR_{E \rightarrow M}^* \text{ brut}} \times 100$$

Si $\Delta \geq 20\%$, alors on dira que X a joué un rôle important de confusion. Si $\Delta \in [10\% ; 20\%[$, alors on dira que X a joué un rôle de confusion modéré. Si $\Delta < 10\%$, alors on dira que X a joué un rôle de confusion minime, et il ne sera alors pas indispensable de prendre en compte X dans les analyses statistiques.

c) Pourquoi ajuster rapproche l'association estimée de l'association causale ?

L'OR*_{E→M} ajusté sur X est une moyenne (pondérée) des OR*_{E→M} stratifiés sur X. Or, nous avons vu (partie VIII.C.1 ci-dessus) que la stratification élimine le biais de confusion dû au facteur de confusion X, lors de l'estimation de l'association entre E et M (quantifiée par l'OR*_{E→M}). Par conséquent, puisque l'OR* ajusté est une moyenne (pondérée) des OR* stratifiés sur X, et puisque les OR* stratifiés sur X ne sont plus biaisés par le biais de confusion dû à X, l'OR* ajusté sur X ne l'est donc plus non plus. Par conséquent, l'association ajustée estimée devient alors plus proche de l'association causale que ne l'était l'association brute estimée avant ajustement.

d) L'ajustement dans les articles de recherche clinique

Dans la quasi totalité des articles de recherche clinique, l'OR* ajusté n'est pas calculé à partir des OR* stratifiés, mais à partir d'une autre méthode : l'utilisation d'un modèle de régression multivarié (évidemment, les deux techniques de calcul conduisent à la même valeur de l'OR* ajusté !). J'ai néanmoins présenté la méthode de calcul de l'OR* ajusté en tant que moyenne d'OR* stratifiés car elle vous permet de comprendre pour quelle raison un OR* ajusté sur X n'est pas biaisé par du biais de confusion dû à X.

D. Introduction aux modèles de régression multivariés

Cette partie pourrait faire l'objet d'un document à lui seul. Je vais donc ici me limiter à une « courte » introduction. (Si le sujet vous intéresse, par exemple pour votre thèse vétérinaire, je vous recommande fortement la lecture de l'article de McNamee (McNamee, 2005).) Cette introduction est cependant indispensable pour être capable de réaliser une analyse critique d'articles dans la mesure où (heureusement) de plus en plus d'articles de recherche clinique vétérinaire utilisent des modèles de régression multivariés, en en présentant les résultats dans un tableau (cf. **Figure 11**, issu de (Murray et al., 2009)) ou dans le texte. Ainsi, pour comprendre les résultats d'un article et faire l'analyse critique des messages cliniques qui sont

délivrés à partir de ces résultats, il faut comprendre les résultats issus d'un modèle de régression !

Table 4. Multivariable conditional logistic regression model of OR, 95% CI and *P*-values of risk factors associated with FIV positive cats tested by CP during 2004

Variable	Number of cases (%) (<i>n</i> = 180)	Number of controls (%) (<i>n</i> = 543)	Adjusted* OR (95% CI)	<i>P</i> -value
<i>Gender</i>				
Female†	43 (23)	292 (54)	1.00	<0.001
Male	137 (76)	251 (46)	3.31 (2.11–5.20)	
<i>Previous history</i>				
Owned†	57 (32)	282 (52)	1.00	<0.001
Stray/feral	123 (68)	261 (48)	3.09 (1.88–5.08)	
<i>Health</i>				
Good/fair†	98 (54)	447 (82)	1.00	<0.001
Poor	82 (46)	96 (18)	4.22 (2.54–7.01)	

*Adjustment is for all variables shown.
†Reference category.

Figure 11. Exemple de résultats issus de l'utilisation d'un modèle multivarié (ici, un modèle de régression logistique).

1. Quand utiliser un modèle de régression multivarié ?

On utilise un modèle multivarié lorsque l'on souhaite quantifier une association entre E et M en ajustant sur au moins un facteur de confusion. Ainsi, en pratique, en épidémiologie clinique, l'utilisation des modèles multivariés est systématique (dès lors que les auteurs souhaitent ajuster leurs associations).

2. Écriture d'un modèle de régression

Soit M la maladie étudiée et E_i une exposition. Un modèle de régression *univarié* contient une seule exposition (notée E_1), et s'écrit de la façon suivante :

$$f(M) = \alpha + \beta_1.E_1$$

Un modèle de régression *multivarié* contient plusieurs expositions, et s'écrit de la façon suivante :

$$f(M) = \alpha + \beta_1.E_1 + \beta_2.E_2 + \dots + \beta_k.E_k$$

Où $f(x)$ est une fonction mathématique qui dépend du type de M (binaire, qualitative, ou quantitative). Dans la suite de cette partie sur les modèles de régression, je ne considérerai que les expositions binaires, de type exposé/non exposé (E_{i+}/E_{i-}).

3. Trois modèles de régression

Le modèle de régression dépend du type de M. Si M est quantitative, le modèle sera un modèle de régression **linéaire** avec $f(M) = M$. Si M est binaire de type « malade / non malade » sans qu'il n'y ait de délai de survenue de la maladie, le modèle sera un modèle de régression **logistique** avec $f(M) = \ln(P/(1-P))$ et $P = \text{proba}(M+)$. Si M est binaire de type « malade / non malade » assorti d'un délai de survenue de M (cf. cours sur l'analyse de survie dans l'UC-0313), le modèle sera un modèle de **Cox à risques proportionnels** avec une expression de $f(M)$ un peu compliquée !

D'autres modèles de régression existent, mais je viens de citer les trois principaux que l'on rencontre dans les articles. Ces trois modèles de régression et leurs conditions d'utilisation sont à savoir pour le module d'épidémiologie clinique de l'UC-0324.

4. Interprétation des résultats issus d'un modèle de régression univarié

Le modèle de régression univarié s'écrit : $f(M) = \alpha + \beta_1.E_1$

Un modèle de régression univarié ne quantifie qu'une seule association : celle entre E_1 et M . Par conséquent, aucune exposition n'est prise en compte statistiquement dans la quantification de cette association entre E_1 et M . L'association dans un modèle de régression univarié est donc toujours une association *brute*.

Dans le cas de la régression logistique, $\beta_1 = \text{Ln}(\text{OR}_{E_1 \rightarrow M} \text{ brut})$. Donc, $\text{OR}_{E_1 \rightarrow M} \text{ brut} = e^{\beta_1}$. Ainsi, la régression logistique univariée fournit un OR brut quantifiant l'association brute entre E_1 et M (après passage par l'exponentielle).

Dans le cas du modèle de Cox, $\beta_1 = \text{Ln}(\text{RR}_{E_1 \rightarrow M} \text{ brut})$. Donc, $\text{RR}_{E_1 \rightarrow M} \text{ brut} = e^{\beta_1}$. Ainsi, le modèle de Cox univarié fournit un RR brut quantifiant l'association brute entre E_1 et M (après, là encore, passage par l'exponentielle). Notez que dans les articles, vous ne verrez pas écrit « RR » mais « HR » pour le modèle de Cox : « HR » pour « Hazard Ratio ». Mais le HR a exactement la même interprétation qu'un RR.

Dans le cas de la régression linéaire, $\beta_1 =$ différence brute de moyennes de M entre les animaux E_{1+} et les animaux E_{1-} .

5. Interprétation des résultats issus d'un modèle de régression multivarié

Supposons le modèle de régression multivarié incluant trois expositions E_1 , E_2 , et E_3 . Il s'écrit : $f(M) = \alpha + \beta_1.E_1 + \beta_2.E_2 + \beta_3.E_3$

Ce modèle de régression multivarié ci-dessus quantifie trois associations (une par paramètre β_i) : celle entre E_1 et M (quantifiée par β_1), celle entre E_2 et M (quantifiée par β_2), et celle entre E_3 et M (quantifiée par β_3). La quantification de l'association entre E_1 et M *prend en compte* les expositions E_2 et E_3 car elles ont été incluses dans le modèle. Par conséquent, l'association entre E_1 et M est ajustée sur les expositions prises en compte, c'est-à-dire ajustée sur E_2 et E_3 ²⁰. De la même façon, l'association entre E_2 et M est ajustée sur E_1 et E_3 , et l'association entre E_3 et M est ajustée sur E_1 et E_2 . Les associations estimées dans un modèle de régression multivarié sont donc toujours des associations *ajustées*. Pour information, l'ordre des expositions dans le modèle n'a aucun impact sur l'interprétation des résultats du modèle.

Si le modèle ci-dessus contenant E_1 , E_2 , et E_3 est une régression logistique, $\beta_1 = \text{Ln}(\text{OR}_{E_1 \rightarrow M} \text{ ajusté sur } E_2 \text{ et } E_3)$. Donc, l' $\text{OR}_{E_1 \rightarrow M}$ ajusté sur E_2 et E_3 vaut e^{β_1} . De la même façon, l' $\text{OR}_{E_2 \rightarrow M}$ ajusté sur E_1 et E_3 vaut e^{β_2} , et l' $\text{OR}_{E_3 \rightarrow M}$ ajusté sur E_1 et E_2 vaut e^{β_3} . Ainsi, une régression logistique multivariée fournit des OR ajustés (après, là encore, passage par l'exponentielle).

Si le modèle ci-dessus contenant E_1 , E_2 , et E_3 est un modèle de Cox, $\beta_1 = \text{Ln}(\text{RR}_{E_1 \rightarrow M} \text{ ajusté sur } E_2 \text{ et } E_3)$. Donc, le $\text{RR}_{E_1 \rightarrow M}$ ajusté sur E_2 et E_3 vaut e^{β_1} . De la même façon, le $\text{RR}_{E_2 \rightarrow M}$ ajusté sur

²⁰ La démonstration est hors programme, mais faites-moi confiance. Si vous voulez cependant la démonstration, demandez-la moi !

E_1 et E_3 vaut e^{β_2} , et le $RR_{E_3 \rightarrow M}$ ajusté sur E_1 et E_2 vaut e^{β_3} . Ainsi, un modèle de Cox multivarié fournit des RR ajustés (après, là encore, passage par l'exponentielle).

Si le modèle ci-dessus contenant E_1 , E_2 , et E_3 est une régression linéaire, β_1 = différence de moyennes de M entre les animaux E_{1+} et les animaux E_{1-} , ajustée sur E_2 et E_3 . (Idem pour β_2 et β_3 qui sont des différences de moyennes ajustées.)

6. Tester statistiquement une association dans un modèle de régression

Pour savoir si l'association entre E_i et M, quantifiée par β_i , est significative, le logiciel de statistique avec lequel le modèle de régression est réalisé fournit le test statistique qui teste si β_i est significativement différent de 0²¹. Ce test statistique est le test de Wald.

7. Interprétation à partir d'un exemple issu de la littérature

Reprenons la **Figure 11**. Le tableau présenté est issu d'une étude sur les facteurs de risque de FIV chez des chats âgés d'au moins six mois²² (Murray et al., 2009). L'étude était une étude transversale. Les chats étaient soit infectés par le FIV soit non infectés au moment du recrutement dans l'étude. La colonne « Adjusted OR » présente les OR ajustés. La note de bas de tableau précise que les OR sont ajustés sur les expositions présentes dans le tableau : sexe (mâle *versus* femelle), histoire du chat (chat de rue *versus* chat de propriétaire), et état de santé apparent (mauvais *versus* bon/acceptable). Le modèle de régression logistique²³ multivarié s'écrit de la façon suivante : $f(\text{FIV}) = \alpha + \beta_1.\text{sexe} + \beta_2.\text{histoire} + \beta_3.\text{état_santé}$.

Les auteurs ont ensuite calculé e^{β_1} , e^{β_2} , e^{β_3} , ce qui a conduit aux valeurs des OR suivantes : 3,31, 3,09, et 4,22 (cf. **Figure 11**). L'OR_{sexe→FIV} de 3,31 est l'OR quantifiant l'association entre le sexe du chat et la présence de FIV, ajusté sur l'histoire du chat et son état de santé. L'OR_{histoire→FIV} de 3,09 est l'OR quantifiant l'association entre l'histoire du chat et la présence de FIV, ajusté sur le sexe du chat et son état de santé. Et l'OR_{état_santé→FIV} de 4,22 est l'OR quantifiant l'association entre l'état de santé apparent du chat et la présence de FIV, ajusté sur le sexe du chat et son histoire. La colonne « P-value » est la colonne des degrés de signification provenant du test statistique de Wald testant chacun de ces trois OR ajustés.

8. Quelles expositions faut-il inclure dans un modèle de régression multivarié ?

La réponse est toute sauf simple (Greenland, 1989). De plus, elle ne sera pas la même si je dois m'adresser à des étudiants de M2 d'épidémiologie analytique ou bien à des futurs vétérinaires que vous êtes. Je vais donc faire très court 😊.

²¹ En effet, si $\beta_i = 0$, alors l'OR (dans le cas de la régression logistique) ou le RR (dans le cas du modèle de Cox) sont égaux à 1 ($e^0 = 1$ 😊), ce qui traduit une absence d'association. Donc si β_i est significativement différent de 0, cela conduit de fait à montrer que l'OR ou le RR est significativement différent de 1, ce qui traduit la présence d'une association significative entre E_i et M.

²² La population cible de l'étude était donc : les chats âgés d'au moins 6 mois.

²³ Puisque la maladie étudiée M (statut infecté/non infecté par le FIV) est binaire et non assortie d'un délai de survenue de la maladie (c'est seulement dans les études de cohorte ou dans les essais cliniques, qui suivent les animaux à partir du recrutement dans l'étude, qu'il existe un délai de survenue de la maladie).

Un modèle multivarié dont l'objectif est d'estimer l'association entre une exposition E et un état de santé M la plus proche possible de l'association causale entre E et M doit en théorie inclure *tous* les facteurs de confusion.

Or, dans la mesure où l'augmentation du nombre d'expositions incluses dans un modèle augmente l'imprécision des estimations des paramètres β_i du modèle (et donc celle des OR ou RR des régressions logistiques ou modèles de Cox), il faut sélectionner les expositions à inclure les plus ... *pertinentes*. La règle communément admise est que le maximum d'expositions que l'on peut inclure dans un modèle est environ égal au nombre d'animaux M+ dans l'échantillon divisé par 10 (cf. exemple dans la note de bas de page²⁴) (Vittinghoff and McCulloch, 2007). Si ce nombre est bien plus élevé que la règle ne le recommande, le modèle fournira des estimations très imprécises qui ne voudront plus rien dire du tout, et l'ensemble des résultats du modèle ne sera pas crédible.

Un modèle de régression qui ne contient pas un facteur de confusion X important dans l'étude de l'association causale entre E et M est un modèle qui va conduire à une estimation β quantifiant l'association entre E et M fortement biaisée par du biais de confusion dû à X. Ainsi, l'OR_{E→M} fourni par cette régression logistique ou le RR_{E→M} fourni par ce modèle de Cox sera donc fortement biaisé par du biais de confusion \Rightarrow cet OR_{E→M} ou RR_{E→M} sera de fait fortement éloigné de l'OR_{E→M} ou du RR_{E→M} causal, même s'il est ajusté sur d'autres facteurs de confusion.

Les expositions qui ne doivent surtout pas être incluses dans un modèle sont des expositions qui peuvent être des conséquences de la maladie étudiée M. Notamment, dans la **Figure 11**, on peut voir que les auteurs ont inclus dans leur modèle l'état de santé apparent du chat. Il est tout à fait possible que cet état de santé soit évalué au moment du diagnostic. Si tel est le cas, cet état de santé serait donc bien davantage une *conséquence* du FIV qu'une *cause* ! Le modèle serait donc un modèle faux, qui fournirait des OR faux, et qui conduirait donc à des messages non scientifiquement prouvés de la part des auteurs. Dur dur.

IX. COMMENT CONCLURE A PARTIR D'UN OR OU UN RR ESTIME DANS UN ECHANTILLON

A. Démarche générale pour conclure à partir d'un OR ou d'un RR

Une ou bien deux phrases doivent être énoncées pour conclure à partir de la valeur d'un OR* **et de sa significativité**. La première phrase fait référence à la conclusion dans l'échantillon. La seconde phrase fait référence à l'inférence (qu'elle soit statistique et/ou causale), mais elle ne doit être énoncée que si l'OR* est significativement différent de 1. En effet, lorsque l'OR* est significativement différent de 1, l'hypothèse nulle H_0 est rejetée, et le risque d'erreur évoqué est $\alpha=0,05$. (L'erreur que l'on commet en rejetant H_0 n'est cependant pas égale à $\alpha=0,05$ (Desquilbet, 2020).) Lorsque l'OR* est non significativement différent de 1, H_0 ne peut pas être rejetée et l'inférence n'est pas possible à faire (car si elle était faite, l'erreur commise en disant que H_0 est vraie, qui dépend du risque d'erreur β de valeur inconnue, serait potentiellement très grande).

²⁴ Par exemple, si dans un échantillon de 350 animaux, 67 présentent la maladie étudiée M, le modèle ne pourra pas inclure plus de (environ) 7 variables, incluant l'exposition d'intérêt bien entendu.

B. Différence entre inférence « possible » et inférence « faisable »

Le mot « possible » fait référence à la statistique et notamment aux risques d'erreur de 1^{ère} espèce α (toujours fixé à 0,05) et de 2^{ème} espèce β (de valeur inconnue). Le mot « faisable » fait référence à l'épidémiologie et notamment à la présence de biais. Ainsi, l'inférence est « possible » lorsque l'on rejette H_0 (lorsque $p \leq 0,05$), et elle n'est « pas possible » lorsque $p > 0,05$. Elle est « faisable » lorsque les biais d'association pertinents sont absents ou bien *a priori* suffisamment faibles pour qu'ils ne remettent pas en cause l'inférence s'ils sont présents, elle n'est « pas faisable » si les biais d'association pertinents sont clairement présents et peuvent facilement remettre en cause l'inférence.

C. Conclusion à partir d'un OR ou d'un RR brut

Tout ce qui est écrit entre crochets ci-dessous doit être remplacé par vous, en fonction du contexte de la recherche clinique et des analyses qui ont été effectuées.

1. OR ou RR brut significativement différent de 1 ($p \leq 0,05$)

a) Conclusion au niveau de l'échantillon

« Dans l'échantillon, [M] était significativement [plus / moins] fréquente parmi [les animaux E+] que parmi [les animaux E-] (OR* brut = [valeur] ; IC_{95%} = [valeur inf ; valeur sup]). »

b) Inférence

Elle est possible à faire car H_0 a été rejetée (car $p \leq 0,05$).

Inférence statistique : « Sous réserve que les biais de classement différentiel et de sélection²⁵ soient absents, il y a des chances pour que, dans [la population cible], [M] soit réellement [plus / moins] fréquente parmi [les animaux E+] que parmi [les animaux E-]. »

Inférence causale : « Sous réserve que [l'exposition E] soit antérieure à la survenue de [M], sous réserve que les biais de classement différentiel et de sélection soient absents, et à condition qu'il y ait une *totale* absence de biais de confusion, il y a des chances pour que, dans [la population cible], [E] soit réellement un facteur [de risque / protecteur] de [M]. »

Cette condition d'une *totale* absence de biais de confusion n'est pas réaliste dans une étude non randomisée. Par conséquent, **l'inférence causale n'est pas faisable à partir d'une association brute dans une étude non randomisée** (et donc ce que j'ai écrit au paragraphe précédent ne peut pas être écrit dans le cas d'une étude non randomisée).

Dans une étude randomisée, en théorie, il n'y a aucun biais de confusion, donc en théorie, l'inférence causale serait faisable à partir d'une association brute significative. En pratique, dans une étude randomisée, il faut vérifier que les animaux E+ sont cliniquement comparables aux animaux E- sur *toutes* les expositions associées à M présentes dans un tableau comparant les animaux E+ aux animaux E-, et en croisant les doigts que le hasard de la randomisation ait conduit au fait que les expositions associées à M mais *non* présentes dans ce tableau soient distribuées de façon comparable entre les animaux E+ et les animaux E-. (Je vous rappelle, cf.

²⁵ Même si le biais de sélection et sa discussion ne sont pas au programme (mais tout cela étant dans l'annexe A), vous devez faire l'hypothèse qu'il est absent.

page 44, que plus la taille des deux groupes sera faible, moins il y aura de chances pour que ces deux groupes soient cliniquement comparables sur toutes les expositions pertinentes, après randomisation.)

2. OR ou RR brut non significativement différent de 1 ($p > 0,05$)

a) Conclusion au niveau de l'échantillon

« Dans l'échantillon, il n'existait pas d'association significative brute entre [l'exposition E] et [la maladie M] (OR* brut = [valeur] ; IC_{95%} = [valeur inf ; valeur sup]). »

b) Inférence

Elle n'est pas possible à faire car H_0 n'a pas pu être rejetée (car $p > 0,05$).

3. OR ou RR ajusté (sur une ou plusieurs expositions) significativement différent de 1 ($p \leq 0,05$)

Dans la conclusion d'un OR* ajusté, il faut absolument montrer que vous avez pris en compte le ou les facteurs de confusion (en ajustant dessus). Pour cela, il faut utiliser l'expression « indépendamment de » suivie du ou des facteurs de confusion qui ont été pris en compte (sur lesquels l'OR* a été ajusté). Ci-dessous, « X » fait référence à l'exposition ou aux expositions sur lesquelles l'OR* a été ajusté.

a) Conclusion au niveau de l'échantillon

« Dans l'échantillon, indépendamment de [X], [M] était significativement [plus / moins] fréquente parmi [les animaux E+] que parmi [les animaux E-] (OR* ajusté = [valeur] ; IC_{95%} = [valeur inf ; valeur sup]). »

b) Inférence

Elle est possible à faire car H_0 a été rejetée (car $p \leq 0,05$).

Inférence statistique : « Sous réserve que les biais de classement différentiel et de sélection soient absents, il y a des chances pour que, dans [la population cible], et indépendamment de [X], [M] soit réellement [plus / moins] fréquente parmi [les animaux E+] que parmi [les animaux E-]. »

Inférence causale (que l'étude soit randomisée ou non randomisée) : « Sous réserve que [l'exposition E] soit antérieure à la survenue de [M], sous réserve que les biais de classement différentiel et de sélection soient absents, et sous réserve que [X] soit (soient) le(s) seul(s) facteur(s) de confusion, il y a des chances pour que, dans [la population cible], [E] soit réellement un facteur [de risque / protecteur] de [M]. »

Après avoir ajusté sur une ou plusieurs expositions, l'inférence causale est donc faisable si les réserves émises sont vraies.

4. OR ou RR ajusté (sur une ou plusieurs expositions) non significativement différent de 1 ($p > 0,05$)

a) Conclusion au niveau de l'échantillon

« Dans l'échantillon, indépendamment de [X], il n'existait pas d'association significative entre [l'exposition E] et [la maladie M] (OR* ajusté = [valeur] ; IC_{95%} = [valeur inf ; valeur sup]). »

b) Inférence

Elle n'est pas possible à faire car H_0 n'a pas pu être rejetée (car $p > 0,05$).

5. Illustration sur un exemple issu d'un article

Reprenons la **Figure 11**, page 49. Les 3 OR présentés dans le tableau sont des OR ajustés, chacun ajusté sur les expositions présentes dans le tableau. Les lignes pour lesquelles « 1,00 » est indiqué dans la colonne « Adjusted* OR » correspondent aux animaux non exposés (par exemple, pour le sexe des chats, la catégorie « non exposée » correspond aux chats femelles). De plus, ces OR ajustés sont significativement différents de 1 car leur $IC_{95\%}$ ne contient pas « 1 » (et leur degré de signification p provenant du test statistique de Wald est $\leq 0,05$). Concluons à partir de l'OR_{sexe (mâle versus femelle) → FIV} dont la valeur est 3,31. (La population cible de l'étude était : « la population de chats âgés d'au moins 6 mois ».)

Conclusion au niveau de l'échantillon :

« Dans l'échantillon, indépendamment de l'histoire du chat et de son état de santé apparent, l'infection par le FIV était significativement plus fréquente parmi les chats mâles que parmi les chats femelles (OR_{sexe (mâle versus femelle) → FIV} ajusté = 3,31 ; $IC_{95\%} = [2,11 ; 5,20]$). »

Inférence statistique :

« Sous réserve que les biais de classement différentiel et de sélection soient absents, il y a des chances pour que, dans la population des chats âgés d'au moins 6 mois, et indépendamment de l'histoire du chat et de son état de santé apparent, l'infection par le FIV soit réellement plus fréquente parmi les chats mâles que parmi les chats femelles. »

Inférence causale :

« Sous réserve que le sexe des chats soit antérieur à la survenue de l'infection par le FIV²⁶, sous réserve que les biais de classement différentiel et de sélection soient absents, et sous réserve que l'histoire du chat *et* son état de santé apparent soient les seuls facteurs de confusion, il y a des chances pour que, dans la population des chats âgés d'au moins 6 mois, le fait d'être un mâle soit réellement un facteur de risque de FIV par rapport au fait d'être une femelle. ».

X. INTRODUCTION À L'EVIDENCE-BASED VETERINARY MEDICINE

A. Définition et pratique de l'« evidence-based veterinary medicine »

Pratiquer l'« evidence-based veterinary medicine » (« médecine fondée sur les preuves » en français), c'est **utiliser** de façon consciencieuse, explicite, et judicieuse la **meilleure preuve** disponible pour la **prise de décision** concernant le **soin** de l'animal (Keene, 2000). La « meilleure preuve » est la preuve établie à partir d'une démarche scientifique la **plus rigoureuse** possible. Par ailleurs, chaque étudiant vétérinaire issue d'une école nationale vétérinaire doit être capable de pratiquer l'evidence-based veterinary medicine (extrait ci-dessous du référentiel de diplôme vétérinaire) :

²⁶ Ce qui est bien entendu évident avec cette exposition d'intérêt qu'est le sexe des chats !

Sc2. Appliquer l'analyse critique de l'organisation des soins et la médecine fondée sur les preuves (Evidence-based veterinary medicine, EBVM)

Connaissances sous-jacentes

Biostatistiques
Médecine vétérinaire fondée sur les preuves

Sciences du vivant et vétérinaires
Techniques de documentation

Indicateurs (la compétence de l'étudiant sera évaluée sur...)

- Présentation de travaux expérimentaux et de cas cliniques d'intérêt selon les règles de présentation scientifique (thèse vétérinaire, article, présentation orale en congrès, posters)
- Présentation de synthèses méthodologiques de manière scientifique
- Dans l'exercice clinique, reconnaître des situations de pratique de la médecine fondée sur les preuves et d'autres types de situations

Capacités

Sc.2.1 Appliquer la médecine fondée sur les preuves sur des ressources déjà existantes

	4A	5A
Justifier scientifiquement une démarche clinique existante	a fait	sait faire
Analyser de manière critique les approches sémiologiques, cliniques et thérapeutiques		
Reconnaître les biais méthodologiques		
Savoir reconnaître le besoin d'une information complémentaire, dans un cadre clinique et thérapeutique		

Sc.2.2 Appliquer la médecine fondée sur les preuves dans la prise de décision clinique et thérapeutique

	4A	5A
Savoir assurer la traçabilité et la démarche scientifique pour prendre une décision	a fait	sait faire
Justifier scientifiquement une nouvelle démarche clinique, différente de la pratique habituelle		

Sc.2.3 Présenter de manière critique et pondérée selon leur pertinence les différents résultats publiés concernant une approche sémiologique, clinique ou thérapeutique, en utilisant pour cela une grille issue de la médecine fondée sur les preuves

4A : a vu
5A : sait faire

La pratique de l'evidence-based veterinary medicine se déroule en 5 étapes :

1. Convertir l'information en une question clinique (rédaction de cette question clinique, qui s'appelle la question PICOT²⁷)
2. Rechercher les preuves répondant à cette question (revue de la littérature)
3. Critiquer ces preuves pour identifier la « meilleure » (analyse critique d'articles)
4. Appliquer en clinique cette « meilleure preuve »
5. Evaluer la performance de cette application clinique

Tout ce que l'on a vu dans ce polycopié vous permet d'acquérir les compétences afin d'analyser de façon critique un article dans ses aspects méthodologiques : les auteurs ont-ils le droit de conclure ce qu'ils concluent à l'issue de leur étude au vu du protocole utilisé, des analyses statistiques réalisées, et de la présence de biais potentiels ? Nous répondrons à cette question à partir de plusieurs articles de recherche clinique vétérinaire lors de vos rotations en 5^{ème} année.

²⁷ Question qui sera enseignée en 5^{ème} année, lors des TD d'EBVM.

B. Escalier de preuve pour conclure au lien causal

Avant de « traiter » (au sens le plus large) un animal ou avant de délivrer un message de prévention auprès d'un propriétaire, chaque vétérinaire devrait avoir en tête le niveau de preuve concernant l'efficacité de ce « traitement » ou concernant le facteur de risque pour lequel le message de prévention est délivré. Par exemple, si un vétérinaire propose à un propriétaire de stériliser sa chienne avant ses premières chaleurs car il pense que faire cela réduit les risques de tumeurs mammaires, ce même vétérinaire doit savoir si le niveau de preuve dans ce lien causal est élevé ou faible. Car s'il est faible (en l'occurrence, il l'était en 2012 (Beauvais et al., 2012)), le vétérinaire doit en être conscient : il proposera alors une intervention qui aura un « coût » (financier, risques chirurgicaux, ...) sans être sûr qu'elle soit efficace. Et vous savez que ce n'est pas parce qu'un article est publié, y compris dans une revue dite de prestige, que le message clinique délivré est en accord avec les méthodes statistiques utilisées.

Un moyen rapide nécessaire mais non suffisant d'évaluer le niveau de preuve d'un article délivrant un message clinique causal consiste à savoir à quel niveau de preuve se situe l'étude clinique qui a été réalisée. Il s'agit de l'escalier de preuve pour conclure au lien causal, présenté dans la **Figure 12** ci-dessous.

Niveau de preuve pour conclure au lien causal

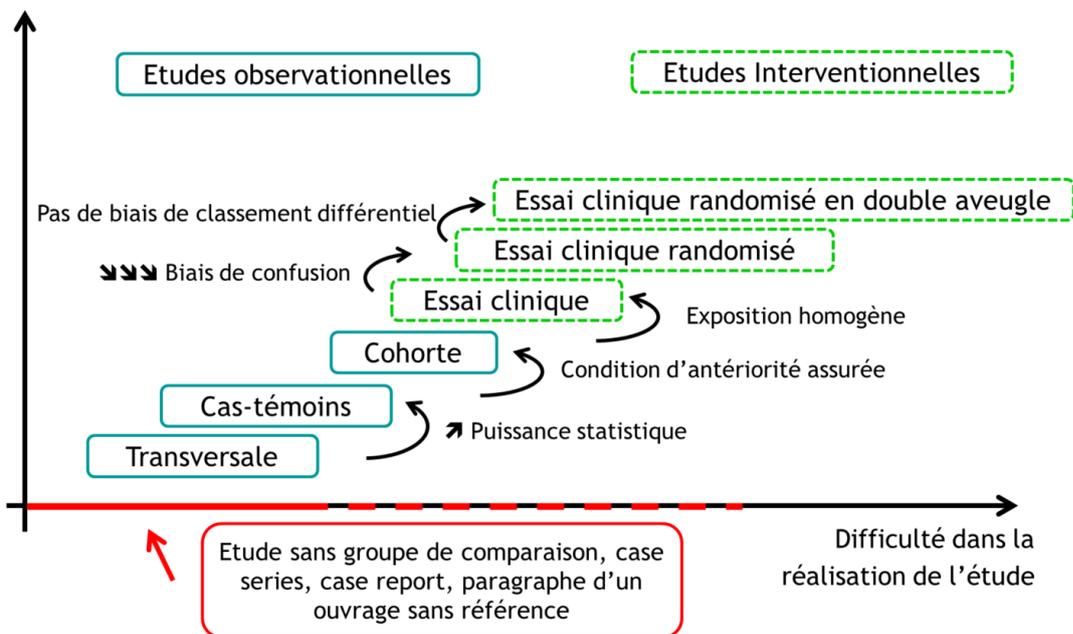


Figure 12. Escalier de preuve pour conclure au lien causal, avec la justification d'une telle hiérarchie.

Un « case report » est une étude au sein de laquelle est décrit un animal qui présente des choses suffisamment particulières pour être communiquées à la communauté scientifique (par exemple (Allen et al., 2020)). Un « case series » est une étude où seul un groupe d'animaux est suivi. Ces études ne comportant pas de groupe de comparaison (ou « groupe contrôle »), elles ont le niveau de preuve le plus bas (par exemple (McPeake and Mills, 2017)). Enfin, un paragraphe d'un ouvrage écrit par un auteur, sans référence bibliographique, a le niveau de preuve le plus bas. En effet, ce paragraphe n'engage que son auteur, et même si cet auteur est mondialement reconnu, un ouvrage ne nécessite pas de relecture par un comité d'experts indépendants.

XI. POINTS-CLES DU POLYCOPIE D'ÉPIDÉMIOLOGIE CLINIQUE

Voici 11 points qu'il faut absolument avoir *compris* à l'issue des enseignements en épidémiologie clinique, et qui vous serviront en 5^{ème} année pour réaliser une analyse critique d'articles. (Apprendre ces points-clés par cœur ne vous sera d'aucune utilité, en examen par exemple...)

1) Lors de la lecture de l'article, il faut identifier l'exposition d'intérêt (notée « E » dans tout ce photocopié) et la maladie étudiée (notée « M »). Si les auteurs de l'étude se sont intéressés à plusieurs expositions, alors l'analyse critique doit porter, individuellement, sur chacune de ces expositions. Faites attention, une « exposition » au sens épidémiologique du terme peut tout à fait être une « maladie » au sens médical du terme, et une « maladie » au sens épidémiologique du terme peut tout à fait ne pas être une « maladie » au sens médical du terme.

2) Plus l'OR* est éloigné de « 1 » par valeur inférieure ou supérieure, plus il traduit une association forte. Pour comparer les forces d'association (par exemple, l'association entre E₁ et M à comparer à l'association entre E₂ et M), il faut comparer les valeurs des OR* en les mettant du même côté de « 1 » (par valeur supérieure de préférence), et ne pas comparer l'aspect significatif / non significatif de ces deux associations.

3) Pour que l'on puisse suggérer qu'une exposition E semble être un facteur de risque ou protecteur de M²⁸, elle doit vérifier la règle des 3A : association statistique significative entre E et M, antériorité de E (telle qu'elle a été collectée) vis-à-vis de M, et absence de biais de classement différentiel, de sélection, et de confusion. Il suffit que l'une de ces trois conditions ne soit pas vérifiée de façon certaine pour qu'il soit impossible de suggérer que E semble être un facteur de risque / protecteur de M.

4) La mise en double aveugle supprime le biais de classement différentiel. Elle ne supprime ni le biais de classement non différentiel, ni le biais de sélection, ni le biais de confusion.

5) Du biais de classement différentiel est présent si les quatre conditions suivantes sont vérifiées dans la situation où M est évaluée après E : (1) erreurs de classement sur M possibles, (2) M est évaluée subjectivement, (3) la personne qui évalue M connaissait le statut E+/E- de l'animal au moment d'évaluer M, et (4) cette même personne suspectait un lien entre E et M (elle a donc pu être influencée par le statut E+/E- de l'animal au moment d'évaluer subjectivement M). Comme que E et M sont interchangeables, remplacez « M » par « E » et « E » par « M » dans la phrase précédente pour avoir toutes les situations où du biais de classement différentiel est présent.

6) Le biais de classement non différentiel diminue toujours la force de l'association, tandis que le biais de classement différentiel augmente (quasiment) toujours la force de l'association, et peut donc rendre une association significative alors qu'elle ne l'aurait pas été sans ce biais de classement différentiel. C'est pour cela que le biais de classement différentiel est dangereux. Le biais de classement non différentiel est quant à lui seulement « embêtant » car il peut rendre une association non significative, alors qu'elle l'aurait potentiellement été sans ce biais de classement non différentiel.

²⁸ Vous avez vu la couche des gants que l'on met, n'est-ce pas ?! 😊

7) Si les animaux E+ ne sont pas cliniquement comparables aux animaux E- sur une exposition X (\Leftrightarrow critère n°1 rempli) par ailleurs associée à la maladie étudiée M (\Leftrightarrow critère n°2 rempli), alors X sera un facteur de confusion qui devra être pris en compte dans les analyses.

8) La prise en compte d'un facteur de confusion X dans l'étude de l'association causale entre E et M empêche le biais de confusion dû à X. La stratification et l'ajustement (notamment via l'utilisation de modèles de régression multivariés) font partie des méthodes statistiques de prise en compte des facteurs de confusion.

9) L'OR* $_{E \rightarrow M}$ brut (\Leftrightarrow non ajusté) estimé dans une étude non randomisée est forcément biaisé par du biais de confusion (il y a forcément une exposition X en même temps associée à E et à M sans en être une conséquence directe). Ces facteurs de confusion devront être pris en compte dans les analyses statistiques.

10) La randomisation supprime *en théorie* tous les biais de confusion. Mais en pratique, le hasard peut mal faire les choses et rendre les animaux exposés à l'exposition d'intérêt non comparables aux animaux non exposés à l'exposition d'intérêt sur au moins une exposition par ailleurs associée à la maladie étudiée (\Rightarrow biais de confusion). Et ce hasard fait d'autant plus mal les choses que la taille des groupes est faible.

11) L'inférence causale à partir d'une association ajustée sur k expositions ($k \geq 1$) n'est possible qu'après avoir vérifié que ces k expositions étaient les seuls facteurs de confusion, et qu'il n'y a pas d'autres facteurs de confusion (non pris en compte, donc).

XII. REFERENCES

Allen, J., Peterson, N., Barrett, K. and Llamas, A., 2020. Graded balloon atrial septostomy for palliation of congenital pulmonary hypertension in a dog: A case report. *J Vet Intern Med.* 34, 283-288.

Beauvais, W., Cardwell, J.M. and Brodbelt, D.C., 2012. The effect of neutering on the risk of mammary tumours in dogs--a systematic review. *J Small Anim Pract.* 53, 314-22.

Braga, L.H., Farrokhyar, F. and Bhandari, M., 2012. Confounding: what is it and how do we deal with it? *Can J Surg.* 55, 132-8.

Broadbent, A., 2011. Conceptual and methodological issues in epidemiology: An overview. *Prev Med.* 53, 215-6.

Desquilbet, L., 2020. Enhancing Clinical Decision-Making: Challenges of making decisions on the basis of significant statistical associations. *J Am Vet Med Assoc.* 256, 187-193.

Duifhuis, E.A., den Boer, J.C., Doornbos, A., Buitelaar, J.K., Oosterling, I.J. and Klip, H., 2017. The Effect of Pivotal Response Treatment in Children with Autism Spectrum Disorders: A Non-randomized Study with a Blinded Outcome Measure. *J Autism Dev Disord.* 47, 231-242.

Ewald, H., Ioannidis, J.P.A., Ladanie, A., Mc Cord, K., Bucher, H.C. and Hemkens, L.G., 2019. Nonrandomized studies using causal-modeling may give different answers than RCTs: a meta-epidemiological study. *J Clin Epidemiol.* 118, 29-41.

- Ferasin, L., Crews, L., Biller, D.S., Lamb, K.E. and Borgarelli, M., 2013. Risk factors for coughing in dogs with naturally acquired myxomatous mitral valve disease. *J Vet Intern Med.* 27, 286-92.
- Greenland, S., 1989. Modeling and variable selection in epidemiologic analysis. *Am J Public Health.* 79, 340-9.
- Guyatt, G.H., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P. and Schunemann, H.J., 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 336, 924-6.
- Keene, B.W., 2000. Towards evidence-based veterinary medicine. *J Vet Intern Med.* 14, 118-9.
- Kyriacou, D.N. and Lewis, R.J., 2016. Confounding by Indication in Clinical Research. *JAMA.* 316, 1818-1819.
- Lonjon, G., Boutron, I., Trinquart, L., Ahmad, N., Aim, F., Nizard, R. and Ravaud, P., 2014. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg.* 259, 18-25.
- McNamee, R., 2005. Regression modelling and other methods to control confounding. *Occup Environ Med.* 62, 500-6, 472.
- McPeake, K.J. and Mills, D.S., 2017. The use of imepitoin (Pexion) on fear and anxiety related problems in dogs - a case series. *BMC Vet Res.* 13, 173.
- Mirshahi, A., Roohipoor, R., Lashay, A., Mohammadi, S.F. and Mansouri, M.R., 2005. Surgical induction of chorioretinal venous anastomosis in ischaemic central retinal vein occlusion: a non-randomised controlled clinical trial. *Br J Ophthalmol.* 89, 64-9.
- Murray, J.K., Roberts, M.A., Skillings, E., Morrow, L.D. and Gruffydd-Jones, T.J., 2009. Risk factors for feline immunodeficiency virus antibody test status in Cats Protection adoption centres (2004). *J Feline Med Surg.* 11, 467-73.
- Nguyen, T.L. and Xie, L., 2021. Incomparability of treatment groups is often blindly ignored in randomised controlled trials - a post hoc analysis of baseline characteristic tables. *J Clin Epidemiol.* 130, 161-168.
- Parascandola, M. and Weed, D.L., 2001. Causation in epidemiology. *J Epidemiol Community Health.* 55, 905-12.
- Sogaro-Robinson, C., Lacombe, V.A., Reed, S.M. and Balkrishnan, R., 2009. Factors predictive of abnormal results for computed tomography of the head in horses affected by neurologic disorders: 57 cases (2001-2007). *J Am Vet Med Assoc.* 235, 176-83.
- Vittinghoff, E. and McCulloch, C.E., 2007. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol.* 165, 710-8.

XIII. ANNEXES

A. Le biais de sélection

1. Introduction

Le biais de sélection est un biais d'association difficile à appréhender et à discuter. C'est la raison pour laquelle il est mis en annexe et n'est donc pas au programme de l'UC-0324. Je ne vais donc pas m'étendre trop longtemps sur ce biais. Je vais m'attacher à en dire l'essentiel.

L'OR*_{E→M} est biaisé par du biais de sélection si l'**association** entre E et M dans l'échantillon est, en raison du protocole recrutement des animaux dans l'échantillon, forcément différente de celle dans la population cible.

Le logigramme ci-dessous (**Erreur ! Source du renvoi introuvable.**) permet de discuter de la présence de biais de sélection.

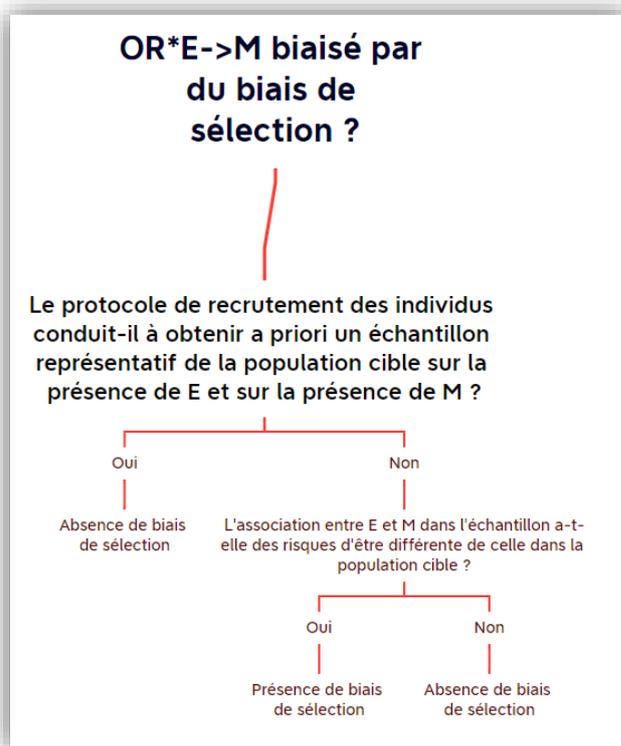


Figure 13. Logigramme permettant d'identifier la présence de biais de sélection dans l'estimation de l'OR*_{E→M}, avec E l'exposition d'intérêt et M la maladie.

Dans la **Figure 13**, la question de la représentativité ne porte que sur E et M. Ainsi, si le protocole de l'étude est tel que, *a priori*, l'échantillon est représentatif sur la présence de E et de M, mais qu'il ne l'est pas sur une autre exposition que E, ce n'est pas grave : l'OR*_{E→M} restera non biaisé par du biais de sélection. Dans le cas où le protocole ne garantit pas cette représentativité vis-à-vis de la population cible sur E et/ou sur M, toute la question est de savoir si l'association entre E et M dans l'échantillon a des risques d'être différente de celle dans la population cible. C'est une question difficile à répondre. Les connaissances médicales sont indispensables pour y répondre.

2. Un échantillon « non représentatif » ne conduit pas forcément à du biais de sélection

Prenez l'exemple de l'enquête cas-témoins (relisez au besoin la partie III.C.3, page 15, sur l'étude cas-témoin.) Par construction, l'échantillon d'une étude cas-témoins comprend souvent autant de cas que de témoins (animaux non malades). Par conséquent, un échantillon d'une étude cas-témoins n'est évidemment pas du tout représentatif de la population cible sur la présence de M (il y a environ 50% d'individus M+ dans l'échantillon, contre un pourcentage beaucoup plus faible dans la population cible). Mais alors, pensez-vous que les OR* estimés dans des études cas-témoins sont donc tous biaisés par du biais de sélection ?! Evidemment que non, car sinon, cela ferait belle lurette que l'on ne mettrait plus en place d'études cas-témoins pour quantifier une association entre une exposition et une maladie !

La règle générale est donc la suivante : ce n'est pas parce qu'un échantillon n'est pas représentatif de la population cible sur E (l'exposition d'intérêt) et/ou sur M (la maladie étudiée) que l'association entre E et M va de fait être biaisée par du biais de sélection.

B. Illustration de l'impact d'un biais de classement non différentiel sur la force de l'association

Illustrons cet impact sur l'exemple suivant. Supposons que l'on cherche à quantifier l'association entre le faible état corporel (EC) d'une vache (codé de façon binaire « faible » / « normal ») et la présence d'une maladie M. Supposons de plus qu'une vache soit dite « exposée » si elle a un EC « faible », et « non exposée » si elle a un EC « normal ». Soit 200 vaches pour lesquelles l'EC et la présence de la maladie ont été évaluées.

Supposons une 1^{ère} situation où aucune erreur de classement n'a été commise sur la maladie tout comme sur l'EC. Le tableau ci-dessous montre la répartition de ces 200 vaches.

		Malade		Total	
		Oui	Non		
EC	Faible	85	20	105	Situation n°1 : aucune erreur de classement sur l'EC
	Normal	15	80	95	
Total		100	100	200	

La répartition montre qu'il y a 85% de vaches exposées parmi les 100 vaches malades, contre 20% parmi les 100 vaches non malades. L'OR quantifiant l'association entre EC et M vaut $OR_{sit^1} = (85 \times 80) / (15 \times 20) = 22,7$.

Supposons la 2^{ème} situation suivante : une autre personne évalue l'EC de ces mêmes 200 vaches, et a tendance à sur-estimer l'EC d'une vache (donc à sous-estimer l'exposition), de sorte que 60% des vaches avec EC faibles vont être, à tort, classées « EC normal », indépendamment du statut malade / non malade de la vache (erreur de classement commise *globalement* sur 60% des vaches avec EC faible). Je rappelle que parmi les 100 vaches malades, sans erreur de classement, 85 ont un EC faible ; cette erreur de classement va conduire au fait que 60% de ces 85 vaches avec un EC faible vont être (mal) classées en « EC normal », ce qui conduit à un effectif de $85 - 85 \times (60/100)$, soit 34 vaches classées « EC faible ». Le même

raisonnement parmi les vaches non malades aboutit à $20 - 20 \times (60/100)$, soit 8 vaches classées « EC faible ». Cela donne le tableau suivant :

		Malade		Total
		Oui	Non	
EC	Faible	34	8	42
	Normal	66	92	158
Total		100	100	200

Situation n°2 :
situation n°1 avec une
sur-estimation
globale de l'EC

$$34 = 85 - 85 \times (60/100)$$

$$66 = 100 - 34$$

$$8 = 20 - 20 \times (60/100)$$

$$92 = 100 - 8$$

La répartition de la situation n°2 montre qu'il y a 34% de vaches exposées parmi les 100 vaches malades, contre 8% parmi les 100 vaches non malades. $L'OR_{sit^2} = (34 \times 92) / (8 \times 66) = 5,9$.

Supposons la 3^{ème} situation suivante : une troisième personne évalue l'EC de ces mêmes 200 vaches, et a tendance à sous-estimer l'EC d'une vache (donc à sur-estimer l'exposition), de sorte que 60% des vaches avec EC normal vont être, à tort, classées « EC faible », indépendamment du statut malade / non malade de la vache. Je rappelle que parmi les 100 vaches malades, sans erreur de classement, 15 ont un EC normal ; cette erreur de classement va conduire au fait que 60% de ces 15 vaches avec un EC normal vont être (mal) classées en « EC faible », ce qui conduit à un effectif de $15 - 15 \times (60/100)$, soit 6 vaches classées « EC normal ». Le même raisonnement parmi les vaches non malades aboutit à $80 - 80 \times (60/100)$, soit 32 vaches classées « EC faible ». Cela donne le tableau suivant :

		Malade		Total
		Oui	Non	
EC	Faible	94	68	162
	Normal	6	32	38
Total		100	100	200

Situation n°3 : situation
n°1 avec une sous-
estimation globale de
l'EC

$$6 = 15 - 15 \times (60/100)$$

$$94 = 100 - 6$$

$$32 = 80 - 80 \times (60/100)$$

$$68 = 100 - 32$$

La répartition de la situation n°3 montre qu'il y a 94% de vaches exposées parmi les 100 vaches malades, contre 68% parmi les 100 vaches non malades. $L'OR_{sit^3} = (94 \times 32) / (6 \times 68) = 7,4$.

Par conséquent, on voit que la situation n°2 (sous-estimation globale de l'exposition) ou la situation n°3 (sur-estimation globale de l'exposition) conduisent dans les deux cas à une

réduction de la **force** de l'association, puisque l'on passe d'un OR = 22,7 dans la situation n°1 (sans erreur de classement) aux OR de 5,9 et 7,4 respectivement dans les situations n°2 et n°3.

Je parle bien de « force de l'association », et non pas de « valeur de l'OR », pour avoir un discours qui fonctionne quelle que soit la valeur de l'OR. En effet, si dans la situation n°1, on avait eu un OR (non biaisé) de 0,4 par exemple, la sur-estimation ou la sous-estimation globale de E aurait conduit à un biais de classement non différentiel, qui aurait conduit à un nouvel OR « moins fort », c'est-à-dire dans cette situation-là supérieur à 0,4 (tout en restant inférieur à 1²⁹), par exemple, 0,7.

Pour illustrer ce phénomène (probablement contre-intuitif pour la situation n°3 : « une sur-estimation globale de l'exposition entraîne une diminution de la force de l'association »), utilisons la notion de contraste (**Figure 14**) : faisons correspondre le % de noir au % de vaches exposées dans deux cercles concentriques, dont le plus grand correspond aux vaches malades (M+), et le plus petit aux vaches non malades (M-). Dans la situation n°1, nous avons le grand cercle avec 85% de noir (gris foncé), et le cercle plus petit avec 20% de noir (gris assez clair) ; dans la situation n°2, nous avons le grand cercle avec 34% de noir (gris), et le cercle plus petit avec 8% de noir (gris clair) ; dans la situation n°3, nous avons le grand cercle avec 94% de noir (gris très foncé), et le cercle plus petit avec 68% de noir (gris assez foncé).

En observant les trois cercles de la **Figure 14**, on remarque que le contraste entre les deux cercles concentriques est plus élevé dans la situation 1 que dans les deux autres situations, ce qui correspond au fait que l'OR est plus élevé dans la situation 1 que dans les deux autres situations.

²⁹ Car le biais de classement non différentiel rapproche la valeur de l'OR de « 1 ».

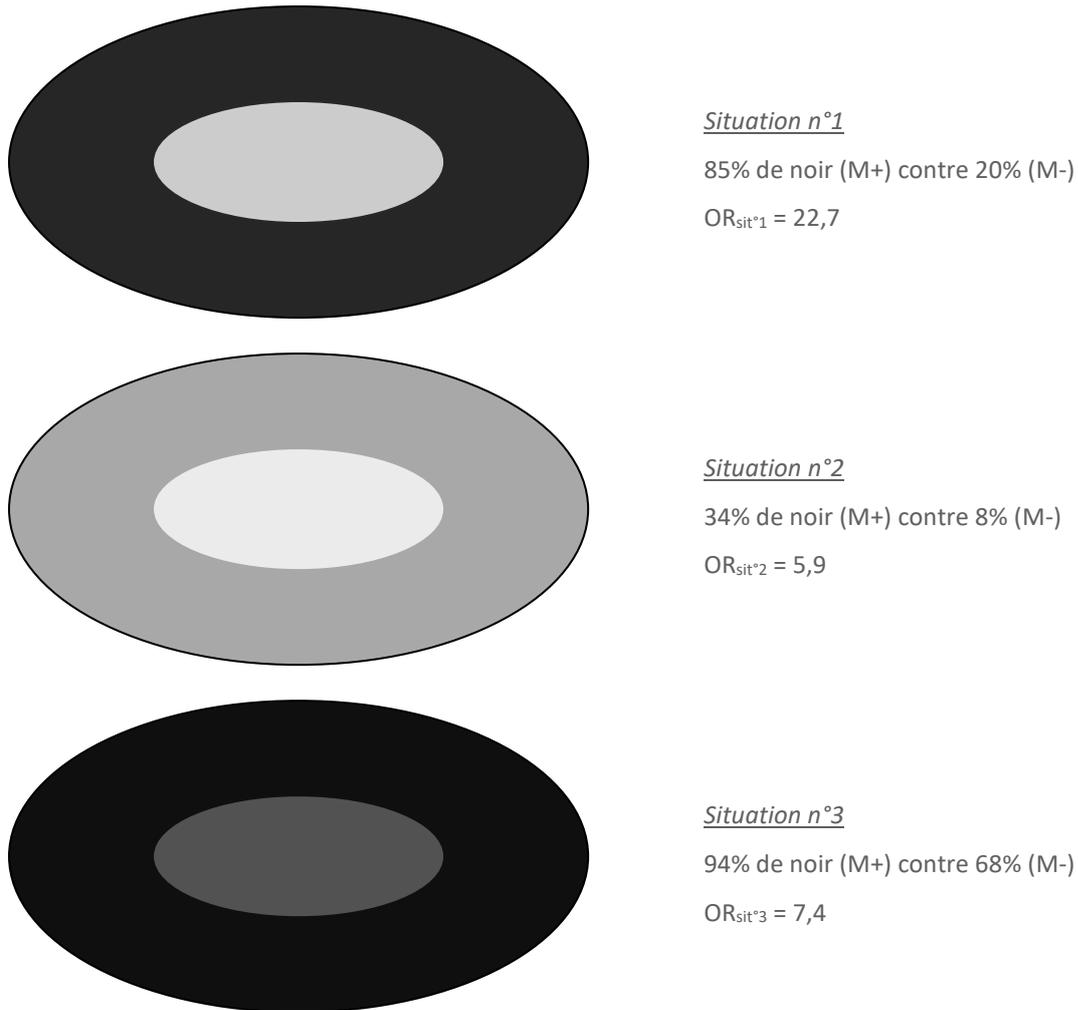


Figure 14. Mise en évidence de la diminution de la force d'une association lors d'un biais de classement non différentiel. Le % de gris représente le % d'animaux (dans l'exemple du texte, les vaches) exposés ; le grand cercle représente les animaux malades (M+), et le petit cercle représente les animaux non malades (M-). La situation n°1 correspond à la situation sans erreur de classement sur l'exposition ; les situations n°2 et n°3 correspondent aux situations où il y a respectivement une sous-estimation globale et une sur-estimation globale de l'exposition, indépendamment donc de la présence de la maladie.

C. Démonstration de « l'égalité des indicateurs comparés est synonyme d'absence d'association statistique » (rappels de biostatistique)

Si A et B sont deux expositions binaires (A+/A- et B+/B-), et si le % de A+ parmi les animaux B+ est égal au % de A+ parmi les animaux B-, alors A n'est pas du tout statistiquement associée à B. Par exemple, soit A le statut castré/stérilisé d'un chien (A+ = chien castré/stérilisé, et A- = chien entier), et B la présence d'une maladie rénale chronique (MRC) (B+ = présence de MRC, et B- = absence de MRC). Si, dans l'échantillon, le % de chiens castrés/stérilisés parmi les chiens avec MRC est égal au % de chiens castrés/stérilisés parmi les chiens sans MRC, alors il n'existe pas d'association statistique du tout entre le statut castré/stérilisé et la MRC parmi les chiens de l'échantillon.

Si A est une exposition binaire (A+/A-) et B est une exposition quantitative, et si la moyenne (ou la médiane) de B parmi les animaux A+ est égale à la moyenne (ou la médiane) de B parmi les animaux A-, alors A n'est pas du tout statistiquement associée à B. Par exemple, soit A le

statut castré/stérilisé d'un chien et B l'âge d'un chien. Si, dans l'échantillon, la moyenne (ou la médiane) de l'âge des chiens castrés/stérilisés est égale à la moyenne (ou la médiane) de l'âge des chiens entiers, alors il n'existe pas d'association statistique du tout entre le statut castré/stérilisé et l'âge chez les chiens de l'échantillon.