



HAL
open science

**Approches statistiques en pharmacoépidémiologie pour
la prise en compte des facteurs de confusion
indirectement mesurés dans les bases de données
médico-administratives : Application aux médicaments
pris au cours de la grossesse**

Chi-Hong Duong

► **To cite this version:**

Chi-Hong Duong. Approches statistiques en pharmacoépidémiologie pour la prise en compte des facteurs de confusion indirectement mesurés dans les bases de données médico-administratives : Application aux médicaments pris au cours de la grossesse. Méthodologie [stat.ME]. Université Paris-Saclay, 2024. Français. NNT : 2024UPASR028 . tel-04906721

HAL Id: tel-04906721

<https://theses.hal.science/tel-04906721v1>

Submitted on 22 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approches statistiques en
pharmacoépidémiologie pour la prise en
compte des facteurs de confusion
indirectement mesurés dans les bases de
données médico-administratives :
Application aux médicaments pris au cours de
la grossesse

*Statistical approaches in pharmacoepidemiology to account for indirectly
measured confounders in healthcare administrative databases:
Application to medications used during pregnancy*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 570, Santé Publique (EDSP)
Spécialité de doctorat : Epidémiologie
Graduate School : Santé Publique. Référent : Faculté de Médecine

Thèse préparée dans l'unité de recherche CESP (Université Paris-Saclay, UVSQ, Inserm),
sous la direction de **Pascale TUBERT-BITTER**, directrice de recherche,
du co-encadrement d'**Ismail AHMED**, chargé de recherche

Thèse soutenue à Paris-Saclay, le 20 Décembre 2024, par

Chi-Hong DUONG

Composition du Jury

Membres du jury avec voix délibérative

Mahmoud ZUREIK PU-PH, Université de Versailles-Saint- Quentin-en-Yvelines	Président
Antoine PARIENTE PU-PH, Université de Bordeaux	Rapporteur & Examineur
Anne-Sophie JANNOT MCU-PH, HDR, Université Paris Cité	Rapporteuse & Examinatrice
David HAJAGE PU-PH, Sorbonne Université	Examineur
Laurent CHOUCHANA PH, Université Paris Cité	Examineur

Titre : Approches statistiques en pharmacoépidémiologie pour la prise en compte des facteurs de confusion indirectement mesurés dans les bases de données médico-administratives : Application aux médicaments pris au cours de la grossesse

Mots clés : Facteurs de confusion indirectement mesurés, Système National des Données de Santé (SNDS), Score de propension en grande dimension, G-computation, Estimation Ciblée par Maximum Vraisemblance, Grossesse

Résumé : Les bases de données médico-administratives sont de plus en plus utilisées en pharmacoépidémiologie. Néanmoins, l'existence de facteurs de confusion (FDC) non mesurés et non pris en compte peut biaiser les analyses. Dans ce travail, nous explorons l'intérêt d'exploiter la richesse des données avec la sélection à large échelle d'un grand nombre de covariables mesurées corrélées avec d'éventuels facteurs manquants pour les ajuster indirectement. Ce concept est à la base du score de propension en grande dimension (hdPS), et nous appliquons la même démarche à la G-computation (GC) et l'estimation Ciblée par Maximum de Vraisemblance (TMLE). Bien que ces méthodes aient été évaluées dans certaines études de simulation, leurs performances sur de grandes bases de données réelles restent peu étudiées. Cette thèse vise à évaluer leurs contributions à l'atténuation de l'effet de FDC directement ou indirectement mesurés dans le système national des données de santé (SNDS) pour des études de pharmacoépidémiologie chez la femme enceinte.

Dans le chapitre 2, nous avons utilisé un ensemble de médicaments de référence en lien avec la prématurité pour comparer les performances des trois méthodes. Toutes ont diminué le biais de confusion, la GC donnant les meilleures performances. Dans le chapitre 3, nous avons réalisé une analyse par hdPS dans un contexte de modélisation plus complexe pour étudier le lien controversé entre les anti-inflammatoires non stéroïdiens (AINS) et la fausse couche spontanée (FCS). Nous avons implémenté un modèle de Cox avec variable dépendant du temps et l'approche "lag-time" visant à corriger d'autres biais (biais de temps immortel et biais protopathique). Nous avons comparé des analyses basées sur les facteurs d'ajustement choisis selon la littérature actuelle ou avec le hdPS. Dans ces deux types d'analyse, les AINS étaient associés à un surrisque de FCS, les différences observées dans les risques estimés pouvant s'expliquer en partie par la différence entre les estimands théoriques ciblés par les approches. Nos travaux permettent de confirmer la contribution des méthodes statistiques à atténuer le biais de confusion. Ils soulignent aussi des difficultés majeures rencontrées lors de leur application en pratique en lien avec la complexité de la modélisation et du plan d'étude, ainsi qu'avec leur coût computationnel.

Title : Statistical approaches in pharmacoepidemiology to account for indirectly measured confounders in healthcare administrative databases: Application to medications used during pregnancy

Keywords : Unmeasured confounders, French administrative healthcare database (SNDS), High-dimensional propensity score, G-computation, Targeted Maximum Likelihood Estimation, Pregnancy

Abstract : Healthcare administrative databases are increasingly used in pharmacoepidemiology. However, the existence of unmeasured and uncontrolled confounders can bias analyses. In this work, we explore the value of leveraging the richness of data through large-scale selection of a large number of measured covariates correlated with unmeasured confounders to indirectly adjust for them. This concept is the cornerstone of the High-dimensional propensity score (hdPS), and we apply the same approach to G-computation (GC) and Targeted Maximum Likelihood Estimation (TMLE). Although these methods have been evaluated in some simulation studies, their performance on large real-world databases remains underexplored. This thesis aims to assess their contributions to mitigating the effect of directly or indirectly measured confounders in the French administrative health care database (SNDS) for pharmacoepidemiological studies in pregnant women.

In Chapter 2, we used a set of reference drugs related to prematurity to compare the performance of the three methods. All reduced confounding bias, with GC showing the best performance. In Chapter 3, we conducted an hdPS analysis in a more complex modeling setting to investigate the controversial association between non-steroidal anti-inflammatory drugs (NSAIDs) and miscarriage. We implemented a Cox model with time-dependent variables and the “lag-time” approach to address other biases (immortal time bias and protopathic bias). We compared analyses adjusted for factors chosen according to the current literature with those chosen by the hdPS algorithm. In both types of analysis, NSAIDs were associated with an increased risk of miscarriage, and the observed differences in estimated risks could partly be explained by the difference between the causal estimands targeted by the approaches. Our work confirms the contribution of statistical methods to reducing confounding bias. It also highlights major challenges encountered during their practical application, related to the complexity of modeling and study design, as well as their computational cost.

REMERCIEMENTS

Je souhaite tout d'abord adresser mes plus sincères remerciements à ma directrice de thèse, Pascale Tubert-Bitter, pour sa bienveillance, sa patience, et son soutien constant depuis mon master jusqu'à aujourd'hui. Ses conseils éclairés et sa rigueur scientifique m'ont permis de progresser avec confiance et de surmonter les obstacles rencontrés au fil de cette thèse.

Je tiens également à remercier chaleureusement mon co-encadrant de thèse, Ismaïl Ahmed, pour toutes les expériences partagées et la confiance qu'il m'a accordée tout au long de ce parcours. Tu es pour moi à la fois un enseignant, un collègue et aussi un ami. Merci pour tous les échanges passionnants sur le plan scientifique, mais aussi pour les conversations agréables sur la course, les jeux de société, et tant d'autres sujets captivants.

A vous deux, je suis profondément reconnaissant pour votre guidance bienveillante et vos encouragements dans les moments les plus difficiles. Travailler à vos côtés a été un immense honneur et une expérience exceptionnelle que j'apprécie profondément.

Je voudrais remercier M. Antoine Pariente et Mme. Anne-Sophie Jannot d'avoir accepté d'être les rapporteurs de cette thèse. Leur expertise et leurs remarques précieuses ont grandement contribué à enrichir ce travail. Je souhaite également remercier M. Mahmoud Zureik, M. David Hajage, et M. Laurent

Chouchana pour leur présence et pour le temps qu'ils ont consacré, ainsi que pour l'expertise qu'ils ont apportée à ce jury. Je vous remercie tous pour l'intérêt et l'attention que vous avez portés à ce travail de thèse.

Je souhaite exprimer ma reconnaissance à l'EPIPHARE, dont le soutien financier et la collaboration ont été essentiels à la réalisation de cette thèse. Je remercie tout particulièrement Jérémie Botton et François Cuenot pour leur réactivité et pour leur soutien précieux tout au long de ce projet.

Merci à Sylvie, ma voisine de bureau, qui m'a choisi parmi tant de candidatures pour rejoindre cette équipe chaleureuse. Merci pour toutes les questions sur la culture, le vocabulaire et les expressions français que tu m'as bien expliqués.

Merci également à mes autres collègues : Émeline, Yannis, Elise, Sidonie, Florence, Matthieu, Ana-Maria, Lucas, Etienne, Astelle, Anne, Hervé, Juliette, Anne-Louise, et Phillippe, pour leur bienveillance et pour tous les discussions intéressantes. Grâce à vous, cette aventure a été bien plus agréable.

Merci à mes amis, Thu-Huong, Huong-Chi, Mai-Hien, Quang-Khai, Anh-Thu, Ngoc-Anh, Quentin, Thai-Anh, Anh-Tho, Tra-My, Quynh-Anh, Tung-Son pour leur soutien indéfectible, leurs encouragements constants. Leur écoute, leurs conseils ont été des sources précieuses de motivation. Merci de m'avoir aidé à organiser un pot de thèse mémorable, pour lequel les idées ont été réfléchies même plusieurs mois à l'avance. A chacun d'entre eux, je suis profondément reconnaissant pour leur amitié et leur générosité de cœur.

Je tiens à adresser un immense merci à ma famille, ma sœur, mon beau-frère, ma nièce et notamment mes parents. Sans vous, je ne serais pas la personne que je suis aujourd'hui. Merci pour vos encouragements, votre amour et la

confiance que vous m'avez toujours accordée. Cela a été une source de motivation inestimable, qui m'ont encouragé à persévérer même lorsque les défis semblaient insurmontables. « Con cảm ơn cả nhà đã luôn là nguồn động viên và là điểm tựa của con trong suốt hành trình này. Con cảm ơn ba mẹ đã sinh ra con và cho con có được ngày hôm nay. Con yêu cả nhà rất nhiều.»

Enfin, je tiens à remercier Tu-Hai, mon compagnon. Aucun mot ne saurait exprimer ma gratitude pour tout ce que tu as fait pour moi. Merci infiniment d'avoir toujours été là, de m'avoir aidé à devenir chaque jour une meilleure version de moi-même. Merci pour ton dévouement, tes encouragements et ta confiance sans faille. Merci de m'avoir accompagné dans cette aventure et d'être une part essentielle de ma vie.

TABLE DES MATIERES

Remerciements.....	i
Table des matières	iv
Valorisation scientifique.....	viii
Publications.....	viii
Communications orales.....	viii
Liste des acronymes	ix
Liste des tableaux.....	xi
Liste des figures	xii
CHAPITRE 1	1
<i>Introduction générale</i>	1
1.1 Causalité, études observationnelles et bases de données médico-administratives.....	1
1.1.1 Bases de données utilisées dans la pharmacoépidémiologie.....	4
1.1.2 Estimation causale et cadre contrefactuel	14
1.1.3 Méthodes d'estimation causale	16

1.1.4	Conditions d'identifiabilité	31
1.1.5	Outils d'évaluation concernant la présence des facteurs de confusion non mesurés	34
1.1.6	Données en grande dimension et facteurs de confusion non mesurés	40
1.2	Objectif de la thèse	44
CHAPITRE 2	47
	<i>Évaluation des performances des méthodes d'inférence causale pour corriger le biais de confusion : comparaison fondée sur un ensemble de référence de médicaments en lien avec la prématurité</i>	<i>47</i>
2.1	Introduction	47
2.2	Méthodologie	50
2.2.1	Base de données	50
2.2.2	Réduction de dimension	52
2.2.3	La régression LASSO (Least Absolute Shrinkage and Selection Operator)	53
2.2.4	Score de propension en grande dimension (hdPS)	55
2.2.5	G-computation	57
2.2.6	Estimation Ciblée par Maximum Vraisemblance (TMLE)	58
2.2.7	Médicaments de référence.....	60
2.2.8	Mesure de performances.....	65
2.3	Résultats.....	66
2.4	Discussion.....	74

2.5	Conclusion	79
CHAPITRE 3	81
	<i>Score de propension en grande dimension dans le cadre de modèles complexes : étude des anti-inflammatoires non stéroïdiens et du risque de fausse couche spontanée</i>	81
3.1	Introduction	81
3.2	Méthodologie et résultats par analyses.....	86
3.2.1	Base de données	86
3.2.2	Analyse avec sélection experte des facteurs de confusion.....	89
3.2.3	Analyse enrichie par une sélection automatisée des covariables d'ajustement	107
3.3	Discussion.....	111
3.4	Conclusion.....	120
CHAPITRE 4	121
	<i>Conclusion générale et perspectives</i>	121
4.1	Conclusion générale	121
4.2	Perspectives.....	126
REFERENCES	132
ANNEXES A	<i>: Documents supplémentaires des analyses du chapitre 2</i>	i
	Annexe A.1 : Algorithme de sélection des grossesses dans le PMSI	ii
	Annexe A.2 : Liste des codes CIM10 utilisés pour le repérage des pathologies observées au cours de la grossesse	iv
ANNEXES B	<i>: Documents Supplémentaires des Analyses du chapitre 3</i>	xx

Annexe B.1 : Liste des médicaments foetotoxiques	xxi
Annexe B.2 : Liste de médicaments tératogènes.....	xxiv
Annexe B.3 : Distribution des hdPS au début de chaque semaine de grossesse (SG)	xxvi
Annexe B.4 : Distribution des délais de première prescription de paracétamol et d'AINS après le début de grossesse.....	xxxii
ANNEXE C : Publications et Soumissions	xxxiii
Annexe C.1 : Article soumis	xxxiv

VALORISATION SCIENTIFIQUE

Publications

CH. Duong, S. Ecolano, R. Demailly, A. Thiebaut, J. Cottenet, C. Quantin, P. Tubert-Bitter, I. Ahmed, Evaluation of Methods Adjusting for Indirectly Measured Confounding in Pharmacoepidemiology Using Healthcare Databases: An Empirical Study Concerning Drugs Inducing Prematurity. (Soumis)

CH. Duong, E. Quentin, R. Demailly, S. Escolano, C. Quantin, J. Botton, P. Tubert-Bitter, I. Ahmed, Non-steroidal anti-inflammatory drugs and the risk of miscarriage: A French nationwide study 2013 – 2019. (Rédaction du manuscrit en cours)

Communications orales

CH. Duong, S. Ecolano, R. Demailly, A. Thiebaut, J. Cottenet, C. Quantin, P. Tubert-Bitter, I. Ahmed, Evaluation of Methods Adjusting for Indirectly Measured Confounding in Pharmacoepidemiology Using Healthcare Databases: An Empirical Study Concerning Drugs Inducing Prematurity, 38th International conference in pharmacoepidemiology (ICPE 2022), Copenhagen, Denmark, August 24–26th, 2022

CH. Duong, S. Ecolano, R. Demailly, A. Thiebaut, J. Cottenet, C. Quantin, P. Tubert-Bitter, I. Ahmed, Approches statistiques en pharmacoépidémiologie pour la prise en compte de facteurs de confusion indirectement mesurés dans les bases médico-administratives, Séminaire Scientifique de l'EPI-PHARE, 26 janvier 2023

LISTE DES ACRONYMES

AINS	Anti-inflammatoires Non Stéroïdiens
AIPTW	Augmented Inverse Propensity Weighted
ATC	Anatomical Therapeutic Chemical
ATE	Average Treatment Effect
ATT	Average Treatment effect for the Treated
ATO	Average Treatment effect for the Overlap population
ATU	Average Treatment effect for the Untreated
BIC	Bayesian Information Criterion
CCAM	Classification Commune des Actes Médicaux
CIM-10	Classification Internationale des Maladies, 10e révision
CMU-C	Couverture Maladie Universelle Complémentaire
DCIR	Datamart de Consommation Inter-régime
ECP	Exposition Cumulative Pondérée
ECR	Essai Contrôlé Randomisé
FCS	Fausse Couche Spontanée
FDC	Facteur de confusion
FDep	French Deprivation Index
GC	G-computation
hdPS	High-dimensional Propensity Score
HR	Hazard Ratio
IC95 %	Intervalle de Confiance à 95 %
IPTW	Inverse Probability of Treatment Weight
IPTWs	IPTW stabilisée
LASSO	Least Absolute Shrinkage and Selection Operator
MSM	Marginal Structural Model
MW	Matching Weight
OAL	Outcome-adaptive LASSO
OR	Odds ratio

OW	Overlap weight
PMSI	Programme de Médicalisation des Systèmes d'Information
SNDS	Système National des Données de Santé
SNIRAM	Système national d'information inter-régimes de l'Assurance maladie
SP	Score de Propension
TMLE	Targeted Maximum Likelihood Estimation

LISTE DES TABLEAUX

Tableau 1. Biais présentés dans les études en pharmacoépidémiologie utilisant les bases de données de santé	14
Tableau 2. Estimands théoriques courants en inférence causale.....	16
Tableau 3. Méthodes de pondération et leurs estimands théoriques correspondants.....	21
Tableau 4. Estimateurs doublement robustes.....	29
Tableau 5. Formule de l'E-value pour certaines mesures d'effet	38
Tableau 6. Liste des médicaments de référence négatifs	63
Tableau 7. Liste des médicaments de référence positifs.....	64
Tableau 8. Catégories des résultats obtenus par les méthodes	66
Tableau 9. Caractéristiques des grossesses exposées et non exposées aux AINS	101
Tableau 10. Résultats de l'analyse intégrant trois jours de latence avec l'ajustement sur 10 covariables choisies selon la littérature	104
Tableau 11. Résultats des analyses de sensibilité	106
Tableau 12. Résultats des analyses intégrant le hdPS au modèle de Cox avec variable dépendant du temps et une période de latence de trois jours.	111

LISTE DES FIGURES

Figure 1. G-computation pour un événement binaire	26
Figure 2. Etapes d'implémentation de la TMLE pour un résultat binaire	30
Figure 3. Exemples de témoins négatifs.....	35
Figure 4. Etapes d'implémentation d'hdPS	42
Figure 5. Mise en œuvre des méthodes hdPS, GC et TMLE	60
Figure 6. Proportion de références négatives dans chaque catégorie : <i>Vrai négatif (protecteur)</i> : blanc ; <i>Vrai négatif (nul)</i> : gris ; <i>Faux positif</i> : noir.....	67
Figure 7. Diagramme de Venn des faux positifs dans les médicaments de référence négatifs	68
Figure 8. Proportion de références positives dans chaque catégorie. <i>Faux négatif (protecteur)</i> : blanc ; <i>Faux négatif (nul)</i> : gris ; <i>Vrai positif</i> : noir.....	69
Figure 9. Diagramme de Venn des vrais positifs dans les médicaments de référence positifs.....	70
Figure 10. Distribution des log(OR) estimés des références négatives et des références positives en fonction des méthodes	71
Figure 11. Log(OR) estimé des médicaments de référence négatifs	72
Figure 12. Log(OR) estimé des médicaments de référence positifs.....	73
Figure 13. Biais de temps immortel	90
Figure 14. Construction de la table des données dépendant du temps	92
Figure 15. Application de l'approche "lag-time"	96
Figure 16. Biais protopathique et approche "lag-time"	98
Figure 17. Flow-chart : sélection des grossesses pour les analyses portant sur la FCS	100

Figure 18. Étapes pour la sélection des variables incluses dans le hdPS.....	108
Figure 19. Estimation des hdPS dépendants du temps.....	109

CHAPITRE 1

Introduction générale

1.1 Causalité, études observationnelles et bases de données médico-administratives

Dans la médecine fondée sur des données probantes, les essais cliniques randomisés (ECR) sont normalement considérés comme les études de référence qui permettent de conclure sur le lien causal entre une exposition (le traitement) et un événement de santé d'intérêt. Dans les ECR, étant donné que l'allocation de traitement est aléatoire grâce à la randomisation, les distributions de toutes les caractéristiques mesurées et non mesurées sont similaires entre les groupes. En conséquence, une différence observée dans le risque d'événement entre les groupes est une preuve de l'effet causal du traitement. Cependant, les ECR présentent certains inconvénients. Premièrement, la petite taille des échantillons et les durées de suivi assez courtes des ECR posent souvent des difficultés quand on veut évaluer les effets indésirables peu fréquents ou rares des médicaments. Deuxièmement, de

nombreux critères d'inclusion dans les ECR peuvent restreindre leurs échantillons à des sous-populations particulières qui ne sont pas représentatives de la population générale et diminuer la généralisabilité des résultats (Sanson-Fisher et al. 2007). Enfin, pour des raisons d'éthique, certains individus vulnérables tels que les personnes âgées, les femmes enceintes, etc. ne sont pas inclus ou sous-représentés dans les ECR ce qui aboutit à un manque d'informations de bonne qualité sur l'efficacité et la sécurité des médicaments pour ces populations (Thake et Lowry 2017; Z, Aa, et Ac 2021).

Bien qu'il soit recommandé que les femmes enceintes limitent leur consommation médicamenteuse, plus de 80 % d'entre elles ont pris au moins un médicament au cours de la grossesse pour soigner des conditions préexistantes ou survenues au cours de la grossesse (Lupattelli et al. 2014). En France, une étude concernant les grossesses sur la période 2011-2014 a relevé que le nombre médian de médicaments pris au cours de la grossesse était de 9 (Demailly et al. 2017). Pourtant, la majorité des médicaments sont utilisés chez les femmes enceintes en « off-label » parce que 90 % des ECR excluent les femmes enceintes pour des raisons éthiques, et alors que des changements physiologiques de la grossesse peuvent avoir pour conséquence des effets médicamenteux imprévus par rapport aux sujets non enceintes, chez la femme enceinte elle-même, et sur l'issue de la grossesse, le développement des organes du fœtus et le devenir néonatal et à plus long terme de l'enfant. Ainsi, les informations sur l'efficacité et la sécurité des médicaments sont souvent insuffisantes ou basées sur des études dans d'autres populations.

Ces dernières années, la tendance croissante au développement et à la mise à disposition de bases de données de santé à des fins de recherche facilite de

plus en plus la mise en œuvre des études en pharmacoépidémiologie chez les femmes enceintes (Benevent et al. 2019). Ces bases de données offrent de grands échantillons représentatifs de la population générale et des sources d'information riches sur l'état de santé des individus. Néanmoins, il s'agit d'études observationnelles, l'intervention n'étant pas assignée au hasard, sujettes à différents types de biais tels que le biais d'indication, le biais protopathique, ou encore le biais causé par les facteurs de confusion (FDC). Un FDC est défini comme un facteur qui influence à la fois la probabilité d'être exposé et le risque d'événement. Autrement dit, il est une cause commune de l'exposition et de l'événement. En réalité, si le traitement est alloué différemment aux individus en fonction de caractéristiques qui impactent également le risque d'événement (les FDC), la différence observée dans le taux d'événement pourrait refléter en partie la distribution différentielle des FDC, ce qui biaise l'effet observé entre le traitement et l'événement. Bien que l'effet des FDC puisse être atténué lors de l'analyse de données à l'aide de techniques telles que la régression classique, le score de propension, la standardisation, ou la stratification, etc. (Kahlert et al. 2017), cette tâche est complexe en pratique. Cela est souvent dû à (1) l'omission de FDC mesurés parmi les variables observées et à (2) l'existence de FDC non mesurés dans la base de données. En effet, l'identification de FDC parmi les variables mesurées dépend des connaissances des chercheurs concernant l'association d'intérêt. Par conséquent, certains FDC peuvent être omis dans les analyses en raison d'une compréhension partielle de la question étudiée. Par ailleurs, même dans le cas de FDC bien identifiés, ils ne sont pas toujours mesurés dans les bases de données, notamment celles construites pour des finalités administratives comme les bases de données médico-administratives.

Différentes approches statistiques et épidémiologiques ont été développées pour estimer le lien causal dans les études observationnelles en contrôlant les effets des FDC non mesurés (Uddin et al. 2016; Guertin, Rahme, et LeLorier 2016). En particulier, l'élargissement des accès aux bases de données de santé en grande dimension pour la recherche, par la richesse des informations présentes, ouvre la possibilité d'atténuer le biais causé par les FDC non mesurés et de compléter les FDC des experts par les FDC potentiels identifiés par des algorithmes automatiques.

Les sections suivantes sont dédiées à la description générale de l'utilisation des bases de données de santé dans les études en pharmacoépidémiologie, du cadre contrefactuel dans l'inférence causale, des méthodes d'estimation causale étudiées dans le cadre de cette thèse : Score de propension (SP), G-computation (GC), et Estimation Ciblée par Maximum de Vraisemblance (TMLE pour « Targeted Maximum Likelihood Estimation »), et des conditions nécessaires pour l'estimation causale. Elles abordent ensuite l'apport des données en grande dimension à la prise en compte des FDC non mesurés, et les approches évaluant le biais résiduel causé par les FDC non mesurés.

1.1.1 Bases de données utilisées dans la pharmacoépidémiologie

1.1.1.1 Bases de données médico-administratives et dossiers médicaux électroniques

Les grandes bases de données de santé (« Healthcare databases ») sont de plus en plus souvent utilisées dans les études en pharmacoépidémiologie ces dernières années. Leur exploitation offre différents avantages : la richesse des

données couvrant divers aspects de la santé des individus, la disponibilité rapide et à faible coût des données, et la représentativité de la population générale et des pratiques de soins. Cette dernière permet d'étudier l'effet d'un traitement dans un contexte réel. De plus, la grande taille et la stabilité des populations suivies dans ces bases de données permettent également d'évaluer l'effet des médicaments à long terme et de détecter les effets indésirables médicamenteux rares.

En général, les bases de données de santé peuvent être catégorisées en deux grandes groupes : les bases de données médico-administratives et les dossiers de santé électroniques (« Electronic healthcare records ») (Hennessy 2006). Chacun a ses propres avantages et inconvénients à considérer lors de la sélection des données utilisées dans les études en pharmacoépidémiologie.

Les bases de données médico-administratives enregistrent les informations sur le recours aux soins des individus à partir de diverses sources de données : les séjours hospitaliers, les délivrances de médicaments en pharmacie, les résultats d'examens biologiques effectués en laboratoire, etc. Ces données servent principalement à calculer les remboursements des dépenses de santé ou à d'autres fins administratives. Les dossiers de santé électroniques, quant à eux, sont alimentés par les informations fournies par les professionnels de santé concernant la prise en charge des patients. Ces informations servent à suivre le parcours de soins des individus, et, dans certaines bases de données, à des fins de recherche, comme le CPRD (« Clinical Practice Research Datalink ») au Royaume-Uni (Wolf et al. 2019) ou la base IPCI (« Integrated Primary Care Information ») aux Pays-Bas (de Ridder et al. 2022).

Contrairement aux dossiers médicaux électroniques, les bases de données médico-administratives ne se limitent pas à des établissements de santé particuliers. Elles couvrent généralement une population plus large, voire la totalité de la population d'un pays, comme c'est le cas en France.

Cela facilite la mise en œuvre de tout type d'étude épidémiologique, comme des études de cohorte, des études cas-témoins, etc. ce qui est plus difficile avec d'autres types de données. Par exemple, les études cas-témoins basées sur les dossiers médicaux électroniques rencontrent souvent des difficultés à constituer un groupe contrôle provenant de la même population source que les cas. La comparaison entre des personnes en bonne santé issues de la population générale et des patients hospitalisés est susceptible de présenter un biais de sélection.

Pour certaines bases de données médico-administratives, dont le SNDS (Système National des Données de Santé) en France, un autre avantage réside dans la possibilité de chaîner ces données avec d'autres bases de données non administratives, telles que des données de cohortes ou de registres, etc. Ce croisement permet non seulement de compléter les informations contenues dans les bases de données administratives, mais aussi de valider les algorithmes de d'identification d'une exposition ou d'une maladie.

Néanmoins, les bases de données médico-administratives présentent des inconvénients qui doivent être pris en compte lors de leur utilisation. Un défaut important de ces bases est l'absence de facteurs liés au mode de vie, de mesures cliniques et biologiques, qui sont potentiellement des FDC, telles que la consommation de tabac ou d'alcool, l'indice de masse corporelle (IMC), la

glycémie, etc. En revanche, ces informations sont souvent enregistrées dans les dossiers médicaux électroniques par les professionnels de santé.

Une autre limite de l'utilisation de ces données à des fins d'études pharmacoépidémiologiques est qu'elles ne contiennent aucune information concernant des médicaments non remboursés, ce qui peut entraîner une mauvaise spécification de l'exposition.

Enfin, une autre faiblesse des bases de données médico-administratives réside dans la validité de certaines informations, en particulier celles qui ne servent pas à déterminer le remboursement ou à d'autres fins administratives, comme les diagnostics de maladies chroniques dans les bases de données des soins ambulatoires.

1.1.1.2 Système National des Données de Santé en France

En France, le SNDS (Système National des Données de Santé) est une grande base de données qui contient les informations de soins d'environ 99% de la population française (Tuppin et al. 2017). Issu du SNIIRAM (le Système National d'Information Inter-régimes de l'Assurance Maladie), qui contient les informations sur les remboursements de l'assurance maladie, le SNDS a été créé en 2016 en consolidant le SNIIRAM avec d'autres bases de données dans le but de faciliter l'exploitation des données de santé à des fins de recherche.

Aujourd'hui, le SNDS se compose de trois sources principales de données (de Germay et al. 2023) :

- *DCIR (données de consommation inter-régimes)* : comprend les variables sociodémographiques : l'âge, le sexe, la date de naissance, le

département de naissance, l'indicateur des bénéficiaires d'une assurance spécifique aux personnes à faible revenu (la CMU), la FDEP, un indice de territoire de désavantage social, ainsi que les données de remboursement des soins, des médicaments, des consultations, etc.

- *PMSI (Programme de Médicalisation des Systèmes d'Information)* : inclut les informations sur les séjours hospitaliers en établissements publics et privés.
- *CépiDC (Centre d'épidémiologie sur les causes médicales de décès)* : contient les informations sur les causes médicales des décès.

Ces dernières années, l'exploitation du SNDS pour mener études en pharmacoépidémiologie a considérablement augmenté. Une revue systématique a montré que 96 % des études utilisant le SNDS sur la période 2012 - 2018 se sont intéressées à l'utilisation, l'efficacité et la sécurité des médicaments (Maillard et al. 2024).

L'exhaustivité est l'un des grands avantages du SNDS. Couvrant 99 % de la population française, cette base de données permet de réaliser des études à grande échelle incluant des groupes souvent sous-représentés dans les ECR, tels que les personnes âgées, les enfants ou les femmes enceintes. Cependant, étant donné que les données du SNDS sont collectées à des fins administratives, il manque des informations précises sur les indications des prescriptions, les médicaments utilisés au cours des séjours hospitaliers, les mesures cliniques ou les résultats d'analyses biologiques telles que la tension artérielle, la glycémie, des scores diagnostiques, etc. Ce manque d'informations peut compliquer le contrôle des différents types de biais. Néanmoins, le SNDS est riche en données concernant divers aspects des soins telles que la

comorbidité, le recours aux soins via les consultations chez les spécialistes par exemple, les expositions médicamenteuses, les hospitalisations, etc. Cette richesse en données peut aider à combler les informations manquantes par des informations substitut. Par ailleurs, le chaînage du SNDS avec d'autres cohortes ou bases de données externes ouvre aussi la perspective de pouvoir combler ces lacunes. En outre, le suivi des données des individus sur plusieurs années permet d'avoir un recul suffisant pour étudier l'effet à long terme des médicaments. Cependant, la richesse des données du SNDS s'accompagne d'une structure complexe qui nécessite des compétences avancées en gestion de base de données et en statistique.

1.1.1.3 Biais dans les études utilisant les bases de données de santé

Une revue systématique portant sur les études en pharmacoépidémiologie s'appuyant sur l'utilisation secondaire des bases de données de santé, telles que les bases médico-administratives ou les dossiers médicaux électroniques, réalisées sur la période 2000 – 2018, a identifié les biais couramment rencontrés dans ce type d'étude (Prada-Ramallal, Takkouche, et Figueiras 2019). Ces biais ont été regroupés en quatre grandes catégories : les biais causés par les FDC, les biais de sélection, les biais de mesure et les biais liés au temps. Le Tableau 1 résume ces biais et présente certaines méthodes utilisées pour les corriger.

Biais	Définition	Méthodes
<i>Biais causé par les FDC</i>		
FDC non mesurés	Certains FDC ne sont pas mesurés dans les bases de données en raison de leurs finalités administratives	<ul style="list-style-type: none"> • Score de propension en grande dimension • Analyse de sensibilité • Etude de cas autocontrôlée
Biais d'indication	Les facteurs cliniques qui déterminent l'allocation du traitement d'intérêt influencent également le risque d'événement (par exemple, le médicament A est donné préférentiellement aux personnes qui présentent un état de maladie plus sévère)	<ul style="list-style-type: none"> • Score de propension classique • Stratification • Ajustement dans le modèle de régression
FDC dépendants du temps	Un FDC varie au fil du temps et dépend du traitement précédent	<ul style="list-style-type: none"> • G-computation • Nested structural model • Modèle structural marginal
Healthy user	Le traitement est alloué préférentiellement aux individus qui tendent à adopter des comportements « favorables » à la santé, comme une meilleure hygiène de vie ou un accès plus fréquent aux	<ul style="list-style-type: none"> • Variable instrumentale • Recours à un comparateur actif plutôt qu'aux personnes non exposées • Score de propension en

	ressources de soins de santé	grande dimension pour capturer les proxys de la tendance à rechercher des soins
<i>Biais de sélection</i>		
Biais protopathique	Le traitement est lié aux manifestations précoces de la maladie, pas encore diagnostiquée	<ul style="list-style-type: none"> • Approche « lag-time » • Restreindre aux individus non concernés par les indications liées aux symptômes précoces de la maladie
Censure informative	Le facteur qui entraîne l'interruption du suivi influence également le risque d'événement	<ul style="list-style-type: none"> • Pondération par l'inverse de la probabilité d'être censuré
Déplétion des susceptibles	Le taux d'un effet délétère est sous-estimé chez les exposés prévalents par rapport aux exposés incidents si la survenue de l'effet au début du traitement entraîne l'interruption du traitement	<ul style="list-style-type: none"> • Restreindre le groupe exposé aux nouveaux utilisateurs
<i>Biais de mesure</i>		

Biais de classification de l'exposition	L'état d'exposition mesuré ne correspond pas à l'état réel (non-observance du traitement, absence d'enregistrement de médicaments non remboursés, etc.)	<ul style="list-style-type: none"> • Validation de l'algorithme de détection + analyses de sensibilité
Biais de classification de l'événement	Erreur de codage	
<i>Biais liés au temps</i>		
Biais de temps immortel	La période entre l'entrée de l'étude et l'exposition est classée incorrectement comme « exposé » alors qu'elle est « non exposée »	<ul style="list-style-type: none"> • Modèle de Cox avec variable dépendant du temps
Biais de temps non mesurable	Une période où les informations sur l'exposition sont ignorées et considéré à tort comme « non exposée » (par exemple : l'absence d'information sur des médicaments utilisés au cours de séjours hospitaliers)	<ul style="list-style-type: none"> • Restreindre l'analyse aux individus non hospitalisés • Pondération des individus basée sur le temps mesurable
Biais de fenêtre temporelle (« time-window bias»)	Dans une étude cas-témoins, la période d'observation pour l'exposition est différente entre les cas et les témoins en raison de durées différentes du	<ul style="list-style-type: none"> • Appariement sur la durée du suivi • Analyse dépendante du temps

	suivi, ce qui génère une différence pour la probabilité d'être exposé. (les personnes saines ont un suivi plus long et ont ainsi une probabilité plus élevée d'être exposées)	
Biais de temps de latence (« time-lag bias »)	La comparaison des traitements à différents stades de la maladie, ce qui entraîne un biais lié à la progression de la maladie	<ul style="list-style-type: none"> • Comparer les patients ayant la même durée ou le même stade de la maladie

Tableau 1. Biais présentés dans les études en pharmacoépidémiologie utilisant les bases de données de santé d'après la publication de Prada-Ramallal et al.

1.1.2 Estimation causale et cadre contrefactuel

La théorie de la causalité repose souvent sur le concept des issues potentielles ou contrefactuelles (Little et Rubin 2000; Höfler 2005). Dans le cadre des études en pharmacoépidémiologie, on s'intéresse au lien entre une exposition médicamenteuse et un événement de santé tel que le décès, un effet indésirable, la guérison d'une maladie etc. Un résultat contrefactuel désigne ce qui pourrait se produire si un individu avait un état d'exposition différent de son état réel. Par exemple, dans le cas d'une exposition binaire ($A = 0$ pour non exposé, $A = 1$ pour exposé) et d'un événement lui aussi binaire ($Y = 0$ pour l'absence de l'événement et $Y = 1$ pour sa présence), si un patient prend le médicament ($A = 1$) et observe l'événement ($Y_{obs} = 1$) dans la réalité, le statut de l'événement (présent ou absent) de ce patient s'il n'avait pas pris le médicament est un résultat contrefactuel et inconnu dans la réalité ($Y^{A=0}$ n'est

pas observé). Dans cet exemple, Y_{obs} désigne le résultat observé qui correspond à l'état d'exposition réel et $Y^{A=a}$ représente celui contrefactuel correspondant à la valeur a de l'exposition. Ces notations seront utilisées dans les autres sections. En outre, l'estimation causale de l'effet d'une exposition requiert la prise en compte d'autres covariables concernant des caractéristiques des individus qui seront notées \mathbf{X}_t pour un vecteur des covariables mesurées au temps t , X_{jt} ($X_{1t}, X_{2t}, \dots, X_{jt}$) considérées pour l'ajustement. X_0 désigne l'ensemble de covariables mesuré à l'inclusion.

L'exposition a un effet causal sur l'événement lorsque $Y^{A=0} \neq Y^{A=1}$. La magnitude de l'effet causal peut être calculée par différentes mesures d'effet : la différence absolue, le risque relatif, l'odds ratio etc. Puisqu'il est impossible de connaître $Y^{A=0}$ et $Y^{A=1}$ d'un même sujet, mais seulement l'un d'entre eux, les analyses de causalité ne ciblent donc pas l'effet de l'exposition au niveau individuel, mais plutôt l'effet au niveau populationnel qui est mesuré en tant qu'effet moyen du traitement. Dans les analyses de causalité, il existe différents estimands théoriques correspondants à une quantité causale et à une population ciblée à partir de laquelle l'on cherche à estimer l'effet de l'exposition. Il est important de distinguer l'objet ciblé à estimer (estimand) de l'outil d'estimation (estimateur ou estimand empirique) et des valeurs calculées de l'estimateur (estimations). Les trois estimands théoriques les plus courants en inférence causale correspondent à une différence absolue mesurée dans différentes populations. Ils sont présentés dans le Tableau 2.

Dénomination	Population cible	Estimand théorique
ATE (Average Treatment Effect)	Toute la population	$E(Y^1 - Y^0)$
ATT (Average Treatment Effect for the Treated)	Individus traités ($A_{obs} = 1$)	$E(Y^1 - Y^0 A_{obs} = 1)$
ATU (Average Treatment Effect for the Untreated)	Individus non traités ($A_{obs} = 0$)	$E(Y^1 - Y^0 A_{obs} = 0)$

Tableau 2. Estimands théoriques courants en inférence causale

1.1.3 Méthodes d'estimation causale

1.1.3.1 Score de propension (SP)

Le SP a été défini par Rosenbaum et Rubin (1983) comme la probabilité d'être exposé au médicament d'un individu conditionnellement à ses caractéristiques observées à l'inclusion, $P(A = 1 | \mathbf{X}_0)$. Le SP est un score d'équilibrage dans le sens où, conditionnellement au SP, la distribution de \mathbf{X}_0 est similaire entre les groupes exposés et non exposés. Ainsi, le SP aide à rééquilibrer les groupes d'exposition et à atténuer le biais causé par \mathbf{X}_0 .

Dans les ECR, le SP est connu car le mécanisme d'allocation de traitement est connu. Dans les études observationnelles, le SP n'est pas connu et il faut l'estimer à partir des données grâce à un modèle statistique $\hat{g}(\mathbf{X}_0) = \hat{P}(A = 1 | \mathbf{X}_0)$. L'estimation du score repose sur deux questions principales : (1) la sélection des variables \mathbf{X} et (2) le choix du modèle de score.

Concernant la sélection de variables pour construire le SP, Brookhart et al. (2006) ont évalué les performances du SP estimé à partir de différentes combinaisons issues d'un ensemble de trois variables : un facteur de risque (associé seulement à l'événement), une variable instrumentale (associée seulement à l'exposition), et un facteur de confusion (associé à l'événement et à l'exposition). Ils ont conclu que l'inclusion de variables instrumentales ne diminuait pas le biais mais réduisait la précision de l'estimateur. En revanche, l'inclusion de variables associées à l'événement (facteurs de risque et de confusion) favorisait les performances du SP en minimisant les conséquences de l'omission des facteurs de confusion ou du surajustement. Notamment, l'inclusion de facteurs de risque pour l'estimation du SP permet d'augmenter la précision sans augmenter le biais de l'estimateur et était donc recommandée par les auteurs.

Après avoir sélectionné les variables, le SP est souvent estimé à l'aide d'un modèle de régression logistique. Cependant, ce modèle suppose certaines hypothèses sur la forme de la relation entre les prédicteurs et l'exposition. Si ces hypothèses ne sont pas vérifiées, l'ajout du SP dans les analyses peut biaiser les estimations. Certaines études récentes ont montré que l'utilisation de modèles statistiques non paramétriques issues de l'apprentissage statistique pourrait améliorer les performances du SP (Lee, Lessler, et Stuart 2010; Cannas et Arpino 2019). Cependant, l'utilisation de ces modèles peut être limitée par leur coût computationnel important et leur plus grande difficulté d'implémentation.

Une fois estimé, le SP peut être intégré dans les analyses de quatre façons : ajustement, appariement, pondération, et stratification.

1.1.3.1.1 Ajustement

L'ajustement consiste à inclure directement le SP en tant que variable continue dans un modèle qui régresse l'événement sur le traitement. Le type de modèle dépend du type d'événement d'intérêt : la régression linéaire pour un événement continu, la régression logistique pour un événement binaire, et la régression de Cox pour le temps de survie. Les modèles de régression reposent sur certaines hypothèses concernant la relation entre le traitement, le score et l'événement (linéarité, homoscedasticité, etc.). Pour que les estimations soient non biaisées, il est nécessaire que ces hypothèses soient vérifiées.

1.1.3.1.2 Pondération

La pondération consiste à assigner à chaque individu un poids qui correspond à une fonction du SP pour créer une pseudo-population dans laquelle l'allocation de traitement est indépendante des variables à l'inclusion. A partir de cette pseudo-population, un modèle de régression de l'événement qui inclut uniquement l'exposition en tant que prédicteur pour estimer l'effet causal de l'exposition. Ce modèle, avec sa structure particulière, est appelé « modèle structural marginal » (MSM). Différents types de pondération créent différentes pseudo-populations qui déterminent l'estimand théorique ciblé par le MSM. Le Tableau 3 montre certains poids couramment utilisés et leur estimand théorique d'intérêt correspondant.

Parmi les pondérations, l'inverse de la probabilité d'être traité (IPTW pour « Inverse Probability of Treatment Weighting ») est la plus courante. L'IPTW produit une pseudo-population dans laquelle tous les individus au sein de

chaque sous-population constituée par les combinaisons des variables utilisées pour estimer le score ont 50 % de chance d'être exposés.

Dans certains cas, les poids d'IPTW extrêmes peuvent être présents, par exemple lorsque l'exposition est rare ou qu'un grand nombre de variables sont utilisées pour estimer le SP. Ils sont notamment observés chez les exposés ayant un SP très proche de 0 ou les non exposés ayant un SP très proche de 1. Intégrer les poids extrêmes dans des analyses peut induire un biais ou une variance excessive dans les estimations (Crump et al. 2009). Dans cette situation, il est possible d'utiliser une troncature pour traiter les poids extrêmes et améliorer la validité des estimations. La troncature consiste à exclure les patients ayant le SP en dehors d'un intervalle prédéfini, par exemple $[0,1 ; 0,9]$ ou $[2,5^{\text{e}} \text{ percentile} ; 97,5^{\text{e}} \text{ percentile}]$ (Crump et al. 2009; Stürmer et al. 2010). Par ailleurs, d'autres pondérations plus robustes aux poids extrêmes peuvent être considérées, telle que l'IPTW stabilisée (IPTWs), le poids de chevauchement (OW pour « overlap weight »), ou le poids d'appariement (MW pour « matching weight »).

L'IPTWs stabilise les poids d'IPTW en les multipliant par la prévalence de l'intervention réellement reçue par l'individu. L'objectif de L'IPTWs est de diminuer le poids des individus exposés ayant un SP faible et ceux non exposés ayant un SP élevé. A la différence de l'IPTW classique, qui crée une pseudo-population plus grande que la population originale, l'IPTWs génère une pseudo-population de même taille que la population d'origine. Par conséquent, cette technique permet d'éviter l'augmentation de la variance liée à une pseudo-population gonflée par des SP extrêmes (Xu et al. 2010).

L'OW assigne un poids égal à $1 - \hat{g}(X)$ aux exposés et à $\hat{g}(X)$ aux non exposés (F. Li, Thomas, et Li 2019). Ainsi, plus le SP est proche de 50 %, plus le poids est grand. Cette pondération privilégie les individus dont la probabilité de recevoir l'une ou l'autre intervention est proche. De ce fait, l'OW cible l'effet causal du traitement au sein de la population des individus dont le SP est dans la région de chevauchement du score.

Le MW assigne un poids égal au ratio entre la probabilité la plus faible de $\hat{g}(X)$ et $1 - \hat{g}(X)$ et la probabilité de recevoir le traitement reçu réellement (L. Li et Greene 2013). Ainsi, le poids varie seulement de 0 à 1 par conception, ce qui permet d'éviter la présence de poids extrêmes. A condition d'un chevauchement adéquat entre les SP des groupes d'exposition, le MW cible l'ATE lorsque les groupes sont de même taille et l'effet causal au sein de la population représentée par le groupe le plus petit (l'ATT ou l'ATU) lorsque la taille des groupes diffère.

Afin d'estimer la variance de l'estimation obtenue par la pondération, il est important de prendre en compte non seulement la variabilité de l'estimation de l'effet de l'exposition, mais aussi celle des scores estimés. A côté de certains estimateurs asymptotiques proposés, le bootstrap peut également être utilisé pour estimer la variance des estimations (Austin 2022).

Pondération	Poids		Estimand ciblé	Interprétation
	Exposé	Non exposé		
IPTW	$\frac{1}{SP}$	$\frac{1}{1-SP}$	ATE	L'effet de l'exposition attendu quand toute la population est exposée versus non exposée
IPTW stabilisé	$\frac{Pr(A=1)}{SP}$	$\frac{Pr(A=0)}{1-SP}$	ATE	Le même estimand que celui de l'IPTW mais de variance inférieure
Poids de chevauchement (overlap weight)	1 - SP	SP	ATO (Average Treatment for the Overlap population)	L'effet causal de l'exposition mesuré chez la population dont le SP est dans la région de chevauchement
Poids d'appariement (matching weight)	$\frac{\text{Min}(SP, 1 - SP)}{SP}$	$\frac{\text{Min}(SP, 1 - SP)}{1-SP}$	ATE d'une sous population	L'effet causal de l'exposition de toute la population quand les groupes sont de mêmes tailles (ATE). L'effet causal de l'exposition du groupe le plus petit quand la taille des groupes diffère (ATT ou ATU).

Tableau 3. Méthodes de pondération et leurs estimands théoriques correspondants

1.1.3.1.3 Appariement

Cette méthode apparie les exposés et les non exposés ayant des SP similaires. Un sujet exposé peut être apparié avec un ou plusieurs sujets non exposés (Austin 2010). L'appariement avec plusieurs sujets non exposés peut être employé pour éviter une exclusion importante des sujets non exposés lorsque le groupe exposé est de petite taille. Par ailleurs, un sujet non exposé peut être apparié à un ou plusieurs sujets exposés, selon que l'appariement se fasse sans ou avec remise. L'appariement sans remise indique qu'un sujet non exposé ne peut être apparié qu'une seule fois à un sujet exposé. En revanche, dans l'appariement avec remise, un sujet non exposé peut apparaître plusieurs fois dans différents ensembles de risque et être apparié à plusieurs exposés.

Les deux méthodes principales pour évaluer la proximité des SP des sujets afin de former les ensembles sont l'appariement par plus proche voisin et l'appariement avec caliper (Rosenbaum et Rubin 1985). Le premier apparie un sujet exposé avec celui non exposé qui a le SP le plus proche de son SP. Dans le second, le sujet non exposé avec le SP le plus proche est choisi au hasard à partir d'un ensemble de sujets non exposés dont le SP tombe dans un intervalle prédéfini (caliper) autour du SP du sujet exposé à apparié.

Une fois les ensembles, l'effet du traitement chez les traités (ATT) est estimé dans le nouvel échantillon des individus appariés. Plusieurs méthodes d'estimation de la variance ont été proposées. Alors que Schafer et Kang (2008) ont proposé de considérer les individus au sein d'un ensemble comme des observations indépendantes, Imbens (2004) a suggéré qu'il faudrait tenir compte de la dépendance liée à l'appariement, car les individus d'un même ensemble ont tendance à avoir des SP similaires, et donc des caractéristiques

similaires. Si le risque d'événement dépend de ces caractéristiques, ces individus auront aussi des risques d'événement plus proches que ceux des sujets des autres ensembles.

1.1.3.1.4 Stratification

La stratification divise les sujets en sous-groupes (les strates) en fonction de leur PS. Le nombre de strates influence la réduction de biais et la précision de l'estimation. En effet, lorsqu'il y a peu de strates, la taille de celles-ci est grande, ce qui diminue la proximité des PS des individus dans chaque strate. Cela nuit donc à l'équilibre des covariables au niveau des strates et induit un biais dans les estimations. En revanche, lorsqu'il y a beaucoup de strates, leur taille diminue ainsi que la précision des estimations. Des travaux ont montré que la stratification fondée sur les quintiles des PS réduisait considérablement le biais (Rosenbaum et Rubin 1984; Cochran 1968). L'effet de l'exposition est estimé au sein de chaque strate, soit par la comparaison directe entre les exposés et les non exposés, soit par un modèle qui régresse le risque d'événement sur l'exposition et les covariables restant déséquilibrées dans les strates (Rosenbaum et Rubin 1983; Lunceford et Davidian 2004).

L'effet de l'exposition est ensuite calculé comme une moyenne pondérée des effets spécifiques de chaque strate. Différents estimands théoriques peuvent être calculés en fonction de la manière de pondérer les effets des strates. Pour le calcul de l'ATE, les strates sont pondérées par la proportion des individus de chaque strate par rapport à la taille totale de l'échantillon de l'étude. Pour le calcul de l'ATT, les strates sont pondérées par la proportion de personnes exposées dans chaque strate par rapport au nombre total de personnes

exposées de l'étude. De la même façon, la variance globale de l'estimation est estimée comme la moyenne pondérée des variances estimées dans chacune des strates.

1.1.3.2 G-computation

La GC fait partie d'une famille de trois méthodes appelées « g méthodes » (g pour généralisée) proposées par James Robins (1986) pour répondre à la question sur la présence de facteurs de confusion dépendant à la fois du temps et de l'exposition précédente. Ces trois méthodes sont le modèle structurel marginal, les modèles emboîtés structurels, et la g formule ou GC.

Bien que ces méthodes aient été développées initialement pour résoudre le problème lié aux FDC dépendant à la fois du temps et de l'exposition précédente, il est possible de les appliquer aux analyses où les FDC sont mesurés à l'inclusion et restent constants tout au long du suivi. Dans ce cas, la GC est équivalente à la standardisation couramment utilisée dans les études épidémiologiques (Vansteelandt et Keiding 2011; Snowden, Rose, et Mortimer 2011). Cette dernière consiste à calculer le risque d'événement Y observé comme

$$E(Y) = \sum_{a,c} E(Y|A = a, C = c)P(A = a|C = c)P(C = c)$$

Équation 1

Où C est le vecteur de facteurs de confusion à prendre en compte.

Dans l'inférence causale, on cherche à calculer une quantité théorique $E(Y^a)$ correspondant au risque de Y lorsque l'exposition A est fixée à a pour toute la population, ce qui implique $P(A = a|X = x) = 1$. D'où

$$E(Y^a) = \sum_c E(Y|a, C = c)P(c)$$

Équation 2

Lorsque C inclut un grand nombre de variables ou des variables continues, il est impossible de calculer $E(Y|A = a, C = c)$ pour chaque valeur de c et a. Au lieu de cela, on peut ajuster un modèle pour estimer $\hat{Q}(a, c) = \hat{E}(Y|A = a, C = c)$, ce qui est également la première étape de la GC. Afin de fixer $P(A = a|X = x) = 1$ pour transformer l'équation 1 en 2, la GC assigne a à la valeur d'exposition de tous les individus. $E(Y^a)$ est ensuite estimée en appliquant le modèle Q à une pseudo-base de données où on fixe $A = a$. La Figure 1 présente la GC appliquée au cas d'un événement binaire dans une étude qui n'implique pas de variation temporelle de variables.

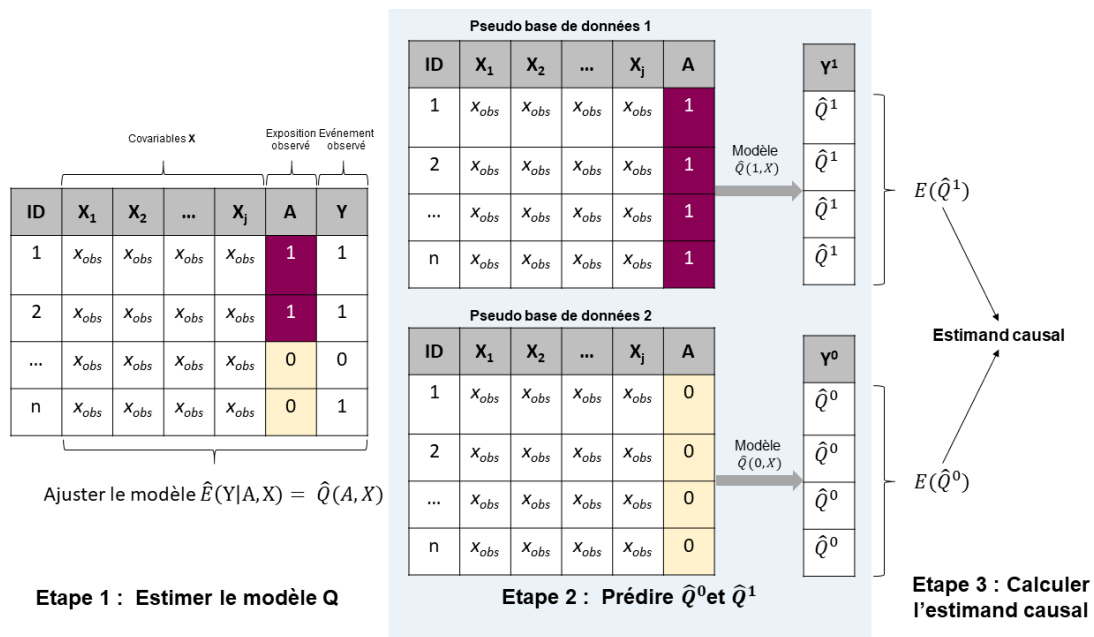


Figure 1. G-computation pour un événement binaire

Dans le cas d'une exposition binaire, chaque individu a deux issues potentielles à prédire $\hat{E}(Y|1, c)$ et $\hat{E}(Y|0, c)$, correspondant respectivement aux scénarios où il est exposé ou non exposé. Ces prédictions sont obtenues en fixant $A = 1$ et $A = 0$ pour tous les individus dans le modèle \hat{Q} et toutes les autres covariables à leur valeur observée. L'effet causal est estimé à partir de deux vecteurs de prédictions de différentes manières.

Lorsque la variable réponse est continue, l'effet causal peut être mesuré par la différence des moyennes des deux prédictions en considérant l'ensemble de la population exposée ou non exposée : $E(\hat{Y}^{A=1}) - E(\hat{Y}^{A=0})$. Pour un événement binaire, les risques d'événement lorsque toute la population est exposée $\hat{\pi}^{A=1}$ ou non exposée $\hat{\pi}^{A=0}$ peuvent être estimés par la moyenne des probabilités prédites par le modèle $\hat{Q}(A, X)$. L'odds ratio est estimé

comme $\frac{\hat{\pi}^{A=1} \times (1 - \hat{\pi}^{A=0})}{(1 - \hat{\pi}^{A=1}) \times \hat{\pi}^{A=0}}$. La variance de l'estimation peut être calculée à l'aide du bootstrap.

Un premier avantage de la GC est sa capacité à évaluer des interventions complexes telles que des interventions dynamiques dans lesquelles l'allocation de traitement dépend de d'autres variables biologiques variant au cours du temps, ou des interventions conjointes impliquant différentes sous-interventions. Un deuxième avantage est qu'il est simple de cibler divers estimands théoriques grâce à la liberté de choisir la population ciblée. Par exemple, l'ATE est calculé à partir des prédictions de tous les sujets, tandis que le calcul de l'ATT se base uniquement sur les prédictions des sujets exposés réels dans la base de données originale.

Néanmoins, la GC présente certains inconvénients. Dans le contexte d'origine de la méthode, où des variables dépendantes du temps existent, la nécessité de construire plusieurs modèles pour les variables et la dépendance entre les modèles via leurs prédictions rendent la GC plus susceptible au biais causé par la mauvaise spécification des modèles. Le développement croissant des méthodes d'apprentissage statistique offre l'opportunité de se servir d'algorithmes plus flexibles pour éviter la mauvaise spécification (Le Borgne et al. 2021). Cependant, plus la modélisation est complexe, plus le coût computationnel de la GC est important ainsi que le risque de surajustement. En outre, l'utilisation du bootstrap pour le calcul de la variance alourdit encore le coût computationnel de la méthode.

1.1.3.3 Estimation Ciblée par Maximum de Vraisemblance (TMLE)

Un estimateur doublement robuste, noté $\hat{\theta}_{DR}$, combine les estimations du modèle d'événement ($\hat{Q}(X)$) et celles du modèle d'exposition ($\hat{g}(X)$) via une fonction spécifique à chaque estimateur. Grâce à cette combinaison, leur biais est égal à la multiplication des biais $\hat{Q}(X) - Q$ et $\hat{g}(X) - g$. Par conséquent, il faut qu'au moins $\hat{Q}(X)$ ou $\hat{g}(X)$ soit non biaisé pour que $\hat{\theta}_{DR}$ soit consistant. Le Tableau 4 présente les deux estimateurs doublement robustes les plus connus dans la littérature à savoir l'AIPTW (pour Augmented Inverse Propensity Weighted) et la TMLE (pour Targeted Maximum Likelihood Estimation).

Méthode	Formule de combinaison
AIPTW	$\hat{Q}_{AIPTW}^1 = \frac{AY_{obs}}{\hat{g}} - \frac{A - \hat{g}}{\hat{g}} \hat{Q}_{GC}^1$ $\hat{Q}_{AIPTW}^0 = \frac{(1 - A)Y_{obs}}{1 - \hat{g}} - \frac{A - \hat{g}}{1 - \hat{g}} \hat{Q}_{GC}^0$
TMLE	$\hat{Q}_{TMLE}^1 = \text{expit} \left\{ \text{logit}(\hat{Q}_{GC}^1) + \hat{\varepsilon}_1 \frac{A}{\hat{g}} \right\}$ $\hat{Q}_{TMLE}^0 = \text{expit} \left\{ \text{logit}(\hat{Q}_{GC}^0) + \hat{\varepsilon}_0 \frac{1 - A}{1 - \hat{g}} \right\}$ <p>Avec $\hat{\varepsilon}_i$ le coefficient estimé à partir d'un modèle de régression logistique (voir la Figure 2 ci-dessous)</p>

Tableau 4. Estimateurs doublement robustes

Proposée par van der Laan et al. (2006), la TMLE suscite un intérêt croissant dans la littérature. Cette méthode se compose de deux étapes principales : (1) obtention de prédictions initiales de la variable réponse, (2) mise à jour de ces prédictions initiales par les informations obtenues à partir d'une modélisation du mécanisme d'allocation de traitement (le modèle du SP). L'application de la TMLE à un événement binaire est présentée dans la Figure 2.

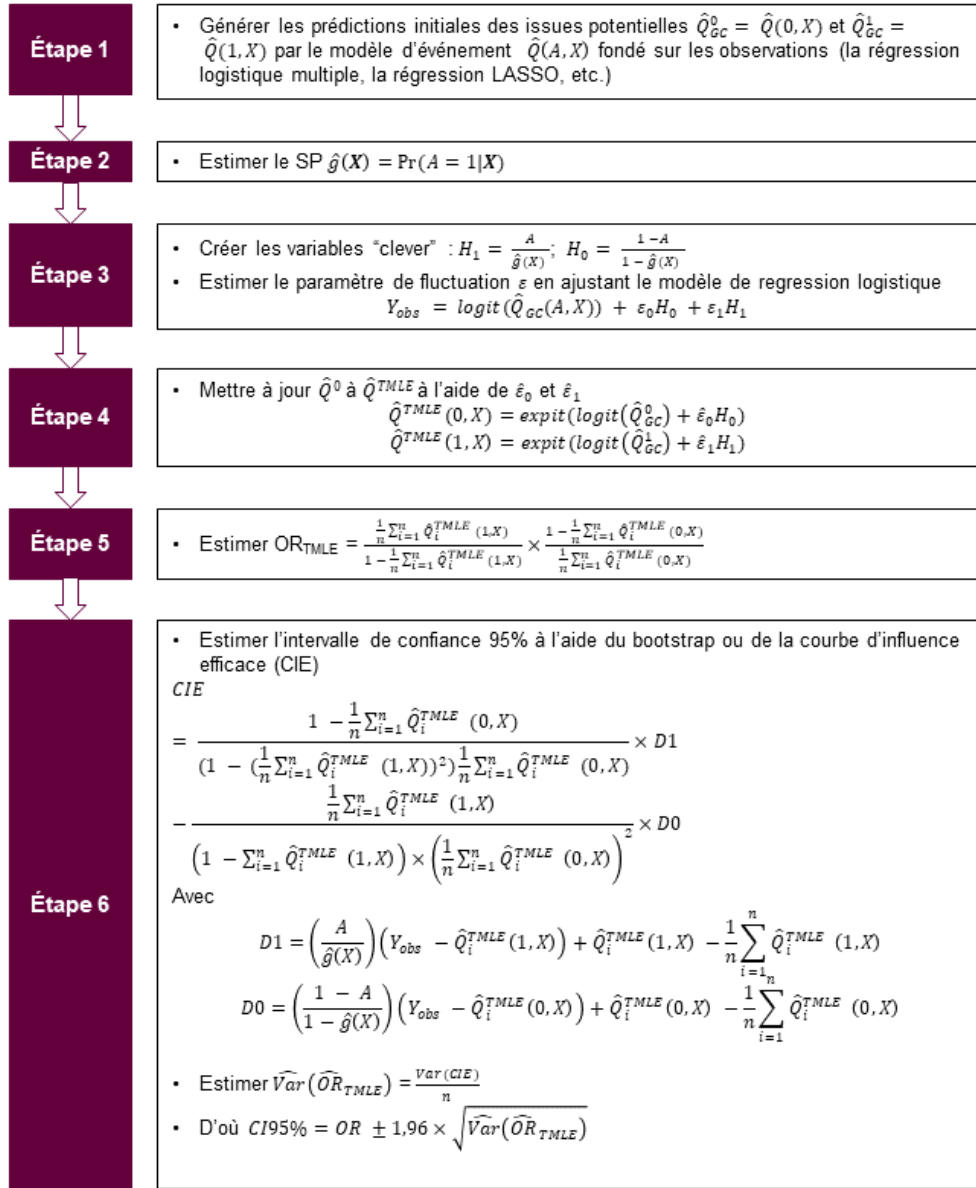


Figure 2. Etapes d'implémentation de la TMLE pour un résultat binaire

Comme les autres estimateurs doublement robustes, la TMLE nécessite qu'au moins un des deux modèles soit correctement spécifié pour produire une estimation consistante. Malgré cette caractéristique attrayante et un intérêt croissant (Smith et al. 2023), la complexité de cette méthode, couplée à un coût computationnel élevé, freine l'utilisation de la TMLE en pharmacoépidémiologie. Des publications didactiques ont été proposées récemment, mais elles se limitent nécessairement à des plans d'études simples (Schuler et Rose 2017; Luque-Fernandez et al. 2018).

Depuis quelques années, plusieurs extensions sont développées pour améliorer les performances ou réduire le coût computationnel de la TMLE traditionnelle. La « Collaborative Double Robust TMLE » intègre les informations du SP estimé par divers algorithmes. Le meilleur modèle du SP est choisi à l'aide de la validation croisée et d'une fonction de perte spécifique au paramètre d'intérêt (Laan et Gruber 2010). Cette méthode aide à améliorer la robustesse et l'efficacité de la TMLE traditionnelle. « Long-format TMLE » est une version adaptée aux bases de données de santé qui concernent souvent les mesures répétées pour chaque individu. Cette méthode permettrait de réduire le coût computationnel en utilisant une fonction choisie par les investigateurs afin de réduire le nombre de covariables prises en compte dans les modèles de prédiction (Sofrygin et al. 2019).

1.1.4 Conditions d'identifiabilité

En réalité, les estimands théoriques ne peuvent être directement estimés à partir d'une base de données en raison de l'ignorance des issues contrefactuelles. Il faut donc utiliser les estimands empiriques, qui sont les

quantités calculées à partir des données pour estimer l'effet causal de l'exposition. Un estimand théorique est dit « identifiable » à partir des données s'il correspond à l'estimand empirique. Par exemple, l'estimand ATE est identifiable dans la base de données lorsque $E(Y^1 - Y^0) = E(Y|A = 1) - E(Y|A = 0)$. Pour qu'un estimand théorique soit *identifiable*, il faut que certaines conditions sur les données et sur l'association d'intérêt soient satisfaites. Ces conditions, appelées « conditions d'identifiabilité », comprennent la positivité, la cohérence, la non-interférence, et l'échangeabilité.

La *positivité* exige que tous les sujets aient une probabilité non nulle de recevoir toutes les valeurs de l'exposition. Lorsque les analyses sont conditionnelles à des covariables, cette condition doit être respectée au sein de chaque strate construite par les combinaisons de covariables. En pratique, il faut que toutes les strates contiennent à la fois des personnes exposées et non exposées. La violation de la positivité est appelée « structurelle » lorsqu'en réalité le traitement n'est jamais ou toujours donné aux sujets de certains groupes. Elle est dite « aléatoire » quand l'absence des exposés ou non exposés résulte de la variation d'échantillon aléatoire. Alors que la violation structurelle augmente à la fois le biais et la variance des estimations, la violation aléatoire augmente principalement la variance sans augmenter le biais quand elle n'est pas trop importante. Bien que l'augmentation de la taille de l'échantillon puisse limiter le risque de violation aléatoire, elle ne peut résoudre le problème causé par la violation structurelle qui nécessite de réévaluer la pertinence de la question d'étude (Petersen et al. 2012; Zhu et al. 2021).

La *cohérence* demande une bonne définition de l'exposition d'intérêt, de telle sorte que chacun n'ait qu'un seul résultat potentiel pour chaque valeur de l'exposition (VanderWeele 2009). En effet, lorsque l'exposition est mal définie, une valeur de l'exposition peut correspondre à différents résultats potentiels. Dans ce cas, le résultat observé reflète la moyenne de tous les résultats potentiels concernés par cette valeur de l'exposition. En d'autres termes, cette condition garantit que le résultat observé pour chaque individu est égal à celui potentiel et unique pour la valeur d'intérêt de l'exposition.

La *non-interférence* postule que les résultats potentiels sont indépendants les uns des autres. Cette hypothèse est souvent violée dans les études concernant la vaccination ou l'infection, dans lesquelles l'état d'exposition d'une personne influence le risque de maladie d'autres personnes.

L'hypothèse d'*échangeabilité* ou d'absence de FDC non mesurés postule que les résultats potentiels sont indépendants du traitement : $Y^A \perp A$. Dans les ECR, cette condition est généralement valide grâce à la randomisation qui répartit équitablement toutes les caractéristiques entre les groupes. Dans les études observationnelles, l'échangeabilité est rarement atteinte parce que l'allocation du traitement dépend de certaines caractéristiques qui pourraient influencer la survenue de l'événement (FDC). Cependant, il est possible d'obtenir l'échangeabilité en conditionnant l'estimation de l'effet sur tous les FDC : $Y^A \perp A | X$. Cette forme d'échangeabilité est appelée « échangeabilité conditionnelle ». Malgré son importance pour estimer l'effet causal, il est impossible de vérifier cette hypothèse à partir des observations. En réalité, sa validité repose sur le plan d'étude et les connaissances disponibles sur le sujet d'étude. Néanmoins, certaines méthodes ont été récemment développées

pour évaluer la vraisemblance de la présence de FDC non mesurés et la robustesse des résultats face à cette présence (Zhang et al. 2018; VanderWeele et Ding 2017). La section ci-dessous présente deux de ces méthodes : l'utilisation de témoins négatifs et l'E-value.

1.1.5 Outils d'évaluation concernant la présence des facteurs de confusion non mesurés

1.1.5.1 Témoins négatifs

Bien que l'hypothèse sur l'absence de FDC non mesurés ne puisse être vérifiée à partir des données, la vraisemblance de la présence de ces facteurs peut être évaluée par l'utilisation de témoins négatifs dans les analyses. L'utilisation de témoins négatifs a été discutée à plusieurs reprises dans la littérature sous différents concepts : événement de falsification, (Prasad et Jena 2013) variable « indicateur », (Flanders et al. 2011) et témoin négatif (Lipsitch, Tchetgen Tchetgen, et Cohen 2010). D'après Lipsitch et al., il y a deux types de témoin négatif : le témoin d'exposition négatif (N_A) et le témoin d'événement négatif (N_Y) (Lipsitch, Tchetgen Tchetgen, et Cohen 2010). Un témoin d'événement négatif N_Y a le même ensemble de causes avec celui de l'association $A - Y$, mais il n'y a pas de lien causal entre A et N_Y . Un témoin d'exposition négatif est une exposition qui possède le même ensemble de causes que l'association d'intérêt $A - Y$, mais qui n'a aucun lien causal avec Y . La Figure 3 présente un exemple pour un témoin d'événement négatif (Figure 3A) et pour un témoin d'exposition négatif (Figure 3B) issus d'études publiées (Jackson et al. 2006; Surén et al. 2013).

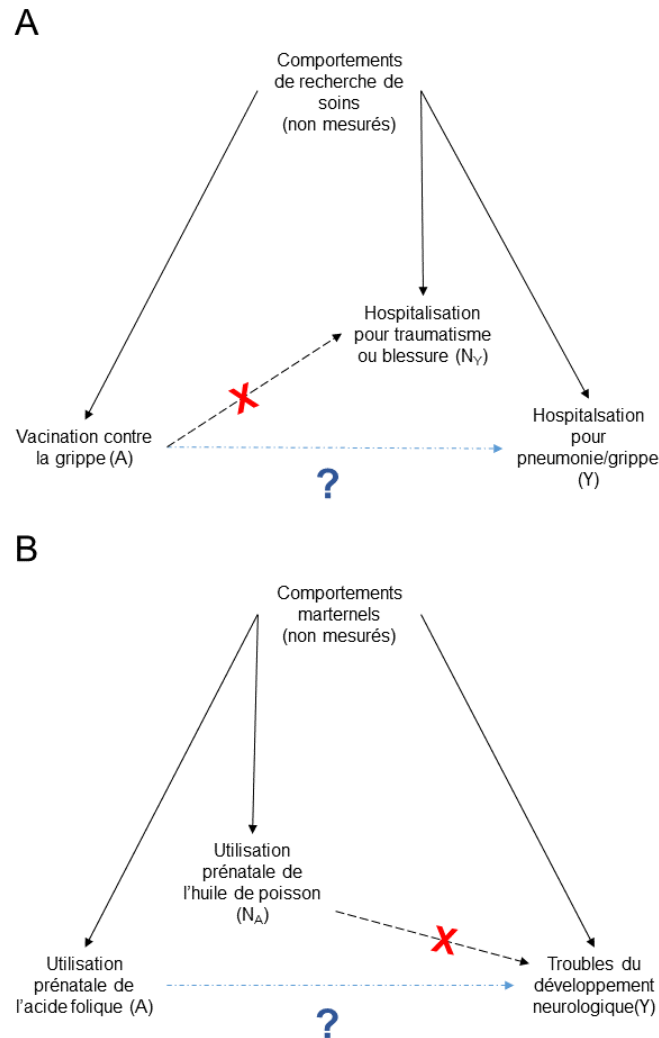


Figure 3. Exemples de témoins négatifs. La Figure 3A présente une étude qui s'intéresse à l'effet du vaccin contre la grippe A sur l'hospitalisation pour pneumonie ou grippe. L'hospitalisation pour d'autres raisons qui n'ont pas de lien causal avec le vaccin telles qu'une blessure ou un traumatisme peut être utilisée comme témoin d'événement négatif. La Figure 3B présente une étude qui s'intéresse au lien entre la prise d'acide folique entre la 4ème semaine avant la conception et la 8ème semaine de grossesse, et les troubles du développement neurologique. L'utilisation prénatale d'huile poisson peut être utilisée comme témoin d'exposition négatif.

Lorsque l'ensemble des causes communes de $N_A - Y$ (ou $N_Y - A$), noté \mathbf{L}_N , chevauche complètement celui d' $A - Y$, noté \mathbf{L} , le témoin négatif est dit « U-comparable » avec A pour N_A ou avec Y pour N_Y . Dans ce cas, il est un bon indicateur de la vraisemblance de la présence de FDC non mesurés. En effet, supposons que N_A soit un témoin d'exposition U-comparable avec A et que l'association observée entre $A - Y$ soit significative dans les analyses. Si les mêmes analyses appliquées à $A - Y$ aboutissent aussi à une association $N_A - Y$ significative, il est possible que l'association observée $A - Y$ soit biaisée par des FDC non mesurés. Une association nulle entre N_A et Y implique que l'association $A - Y$ ne semble pas biaisée, au moins par les voies examinées par N_A .

En réalité, les témoins négatifs sont seulement U-comparables approximativement et le degré de similarité entre \mathbf{L} et \mathbf{L}_N détermine la qualité et la validité du témoin négatif. Par exemple, s'il existe un FDC non mesuré U_1 qui fait partie de \mathbf{L}_N , mais pas de \mathbf{L} , l'association $N_A - Y$ pourrait être significative à cause du biais causé par U_1 , bien que l'association $A - Y$ ne soit pas biaisée. En revanche, un FDC non mesuré U_2 appartient à \mathbf{L} , mais pas à \mathbf{L}_N pourrait biaiser l'association $A - Y$ malgré une association $N_A - Y$ nulle (Lipsitch, Tchetgen Tchetgen, et Cohen 2010; Groenwold 2013). Donc, lors de la sélection de témoins négatifs, il faut que \mathbf{L} et \mathbf{L}_N soient les plus similaires possibles pour que le témoin possède une bonne validité.

1.1.5.2 E-value

L'E-value est un outil récemment proposé pour évaluer la force des preuves de causalité dans les études observationnelles (VanderWeele et Ding 2017). Cette mesure est simple à calculer et ne repose sur aucune hypothèse concernant la

nature ou la distribution des FDC non mesurés. L'E-value quantifie la force minimale de l'association qu'un FDC non mesuré doit avoir simultanément avec l'exposition A et l'événement Y pour invalider l'association observée A – Y après avoir ajusté sur des covariables prédéfinies. A la différence des témoins négatifs, l'E-value aide à évaluer la robustesse des résultats face à la présence éventuelle de FDC non mesurés, mais ne permet pas de vérifier leur présence dans les analyses. Un autre avantage de l'E-value est la possibilité d'appliquer cette méthode à différentes mesures d'effet (VanderWeele et Ding 2017). Les formules de calcul de l'E-value pour certaines mesures d'effet sont présentées dans Tableau 5.

Mesure d'effet	Formule de calcul
Risque relatif (RR)	<ul style="list-style-type: none"> • $\widehat{RR} > 1$ <ul style="list-style-type: none"> ○ Estimation ponctuelle E-value = $\widehat{RR} + \sqrt{\widehat{RR} \times (\widehat{RR} - 1)}$ ○ Borne inférieure de CI95% (BI > 1) E-value = $BI + \sqrt{BI \times (BI - 1)}$ • $\widehat{RR} \leq 1$ <ul style="list-style-type: none"> ○ Estimation ponctuelle E-value = $\frac{1}{\widehat{RR}} + \sqrt{\frac{1}{\widehat{RR}} \times (\frac{1}{\widehat{RR}} - 1)}$ ○ Borne supérieure de CI95% (BS < 1) E-value = $\frac{1}{BS} + \sqrt{\frac{1}{BS} \times (\frac{1}{BS} - 1)}$
	<ul style="list-style-type: none"> • Événement rare (prévalence < 15%) : Remplacer le \widehat{RR} dans les formules ci-dessus par \widehat{OR} • Événement rare (prévalence \geq 15%) : Remplacer le \widehat{RR} dans les formules ci-dessus par $\sqrt{\widehat{OR}}$
Hazard ratio (HR)	<ul style="list-style-type: none"> • Événement rare (prévalence < 15%) : Remplacer le \widehat{RR} dans les formules ci-dessus par \widehat{HR} • Événement rare (prévalence \geq 15%) : Remplacer le \widehat{RR} dans les formules ci-dessus par $\frac{1 - 0.5\sqrt{\widehat{HR}}}{1 - 0.5\sqrt{\frac{1}{\widehat{HR}}}}$

Tableau 5. Formule de l'E-value pour certaines mesures d'effet

Dans les études observationnelles, lorsque les échantillons sont de petites tailles, l'effet de l'exposition doit être suffisamment grand pour que l'association soit significative, ce qui produit nécessairement une E-value élevée. Appuyer le calcul uniquement sur l'estimation ponctuelle pourrait donc

entraîner une surestimation de la robustesse des résultats. Ainsi, une étude évaluant l'E-value de 100 études épidémiologiques sur la nutrition et la pollution de l'air relevait que les études ayant un petit échantillon tendaient à avoir les E-values les plus élevées (Trinquart et al. 2019). Il est donc recommandé de calculer l'E-value non seulement pour l'estimation ponctuelle, mais aussi pour les bornes de l'intervalle de confiance, afin d'évaluer la robustesse de la signification statistique observée (VanderWeele et Ding 2017).

L'interprétation de l'E-value dépend substantiellement du domaine de l'étude, de la nature de l'événement ou de l'exposition, et des covariables mesurées. Étant donné que l'E-value reflète la robustesse conditionnelle des résultats sur les covariables prédéfinies, l'omission de FDC importants dans les analyses pourrait remettre en cause la fiabilité de l'E-value calculée. En fonction du domaine d'étude, la conclusion concernant la force des preuves de causalité tirée d'une même E-value peut être différente. Une E-value égale à 2 signifie que, pour annuler l'association observée entre une exposition A et un événement Y, un FDC non mesuré U doit simultanément avoir un RR_{AU} (risque relatif entre U et A) ≥ 2 et un RR_{UY} (risque relatif entre U et Y) ≥ 2 . Cette valeur peut être importante et indiquer une preuve solide de causalité dans un domaine ou pour un événement donné, comme la mortalité toute cause confondue, mais paraît modeste dans d'autres domaines ou pour d'autres événements, comme la suicide (VanderWeele, Ding, et Mathur 2019). De nos jours, il n'existe pas encore de seuil consensus pour déterminer de « bonnes » E-values. L'interprétation de l'E-value varie ainsi entre les études. Une standardisation d'application et d'interprétation serait utile pour favoriser l'application de cette méthode (Blum, Tan, et Ioannidis 2020).

Il convient de noter qu'en se concentrant sur les FDC non mesurés, L'E-value ne reflète pas la robustesse des résultats face à d'autres types de biais. Par ailleurs, une E-value élevée indique seulement une association forte qu'un FDC non mesuré doit avoir pour infirmer le lien causal estimé, mais ne prouve pas que ces facteurs sont absents ou que l'effet observé est réel (VanderWeele et Ding 2017).

Bien qu'il existe encore aujourd'hui des controverses autour de l'application systématique de l'E-value dans les études observationnelles, (Ioannidis, Tan, et Blum 2019; Fox, Arah, et Stuart 2020; Greenland 2020) cette méthode pourrait être utile en complément d'autres analyses de sensibilité lors de l'évaluation de la force des preuves de causalité (VanderWeele et Mathur 2020).

1.1.6 Données en grande dimension et facteurs de confusion non mesurés

La présence de FDC non mesurés, tels que les résultats biologiques ou liés au mode de vie, est l'un des problèmes inhérents à l'exploitation des bases de données médico-administratives (Schneeweiss et Avorn 2005). Certaines méthodes épidémiologiques ou statistiques ont été développées pour faire face à ce défi, dont le score de propension en grande dimension (« hdPS » pour « High-dimensional Propensity Score ») proposé par Schneeweiss et al. (2009). Le hdPS repose sur l'idée d'utiliser un grand nombre de covariables mesurées pour compenser les informations manquantes concernant les facteurs non mesurés. Ces covariables de substitution sont souvent désignées sous le terme « unmeasured confounders proxies » dans la littérature. En effet, l'ajustement indirect sur les FDC non mesurés à travers des combinaisons de covariables mesurées corrélées peut atténuer le biais causé par les FDC manquants.

Certaines études ont proposé la condition de « repérabilité » (« pinpointability ») pour désigner le fait qu'une variable non mesurée U peut s'exprimer par une fonction déterministe des variables mesurées. Zhang et al. (2022) ont ainsi montré que lorsque U était repérable par un ensemble de covariables \mathbf{X} , l'ajustement sur \mathbf{X} pouvait donner des résultats similaires avec celui sur U .

Le hdPS est une méthode semi-automatique parce que le score est estimé à partir des covariables prédéfinies par les experts et de celles sélectionnées par un algorithme automatique adaptatif aux données. Ce dernier fait une sélection de variables fondée sur leur prévalence et leur potentiel de confusion évalué à l'aide de la formule de Bross :

$$\frac{P_{C1}(RR_{CD} - 1) + 1}{P_{C0}(RR_{CD} - 1) + 1}$$

Équation 3

Avec : P_{C1} : Prévalence de la covariable chez les exposés

P_{C0} : Prévalence de la covariable chez les non exposés

RR_{CD} : Risque relatif entre la variable X et l'événement

La procédure d'estimation du hdPS est présentée dans la Figure 4.

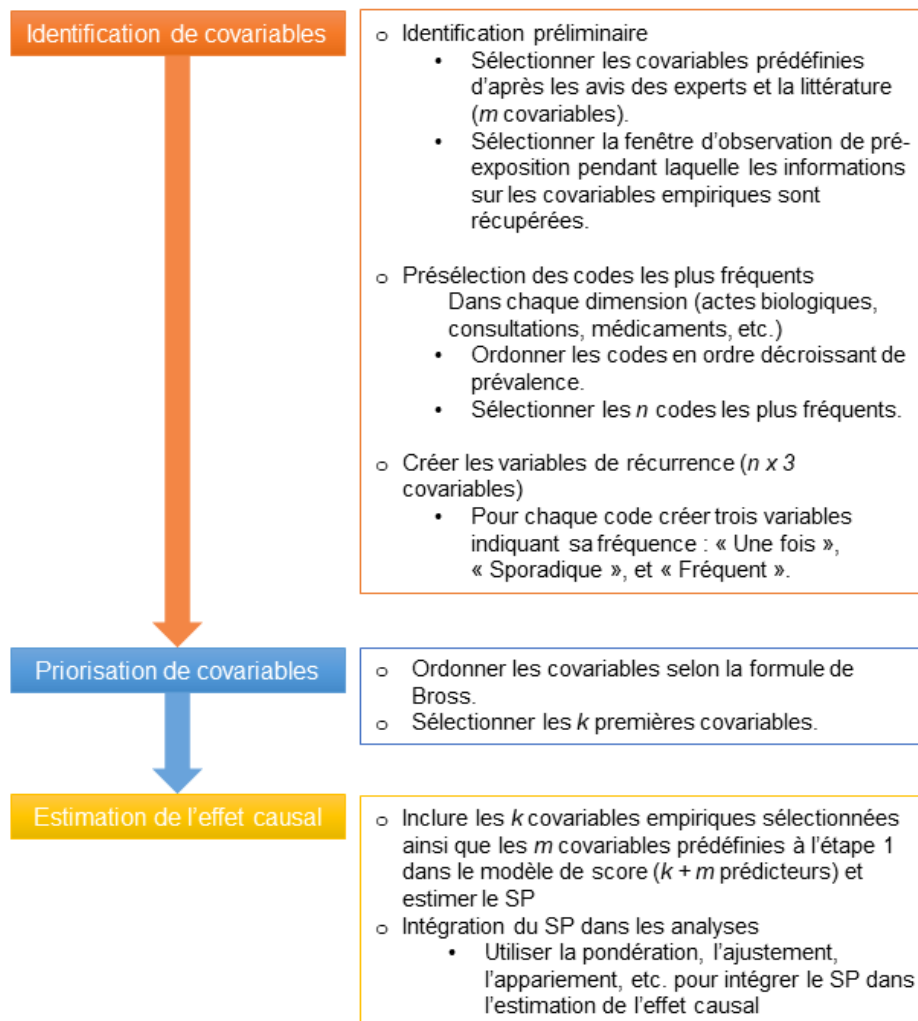


Figure 4. Etapes d'implémentation d'hdPS

Plusieurs extensions de hdPS ont été proposées dans la littérature. Le score de propension à grande échelle (LSPS pour Large-Scale Propensity Score, (Tian, Schuemie, et Suchard 2018)) propose par exemple d'inclure toutes les covariables pré-exposées dans un modèle LASSO (Least Absolute Shrinkage and Selection Operator, (Tibshirani 1996)) pour estimer le score. Il est possible d'appliquer d'autres méthodes d'apprentissage automatique plus flexibles que le modèle LASSO, telles que les forêts d'arbres décisionnelles ou les arbres de régression boostés (L. Zhang et al. 2022). Cependant, l'utilisation de modèles complexes et l'inclusion d'un grand nombre de covariables augmentent considérablement le coût computationnel du LSPS. Dans l'étude de Tian et al. (2018), le hdPS présente un coût computationnel nettement plus favorable que celui du LSPS (quelques minutes pour le hdPS contre plusieurs heures pour le LSPS, sur une base de données contenant 100 000 observations et 100 000 covariables).

Il convient de noter qu'à part le hdPS, qui a été conçu spécifiquement pour l'exploitation des données en grande dimension, il est possible d'adapter l'idée d'ajuster indirectement les FDC non mesurés à travers un grand nombre de variables corrélées mesurées à la GC et à la TMLE. L'intégration de l'algorithme du hdPS à la TMLE a été comparée au hdPS dans une étude de Pang et al. (2016), qui a relevé des divergences entre les résultats obtenus par ces deux méthodes. Plus récemment, Ju et al. (2019) ont mis en œuvre la TMLE et la GC au contexte des données en grande dimension dans le cadre du développement d'une nouvelle approche, la « Scalable Collaborative TMLE ».

1.2 Objectif de la thèse

Comme mentionné précédemment, il est important d'évaluer l'efficacité et la sécurité des médicaments chez les femmes enceintes, une population souvent sous-représentée dans les essais cliniques. Depuis quelques années, grâce à l'exhaustivité des bases de données médico-administratives, la mise en place d'études dans cette population est facilitée. Néanmoins, seule hdPS s'est spécifiquement intéressée à la prise en compte de FDC non mesurés par le biais de sélection de variables semi-automatisée. L'intégration de cette sélection à d'autres méthodes causales est peu étudiée dans les études de simulation et encore moins à partir de grandes bases de données médico-administratives.

L'objectif général de cette thèse est d'évaluer l'apport de méthodes tirant partie de la richesse des données présentes dans le SNDS pour la conduite d'études en pharmacoépidémiologie chez la femme enceinte.

Dans la première partie, nous avons comparé les performances de trois méthodes adaptées au cadre de la grande dimension : le score de propension, la G-computation, et l'Estimation Ciblée par le Maximum de Vraisemblance dans une base de données médico-administrative concernant les grossesses en France pendant la période 2011-2014. Afin de comparer leurs performances, nous avons utilisé un ensemble de médicaments de référence établi dans une étude précédente de Demailly et al. (2020) qui s'intéressait au lien entre médicaments et risque d'accouchement prématuré.

Dans la deuxième partie de ce travail, nous avons étudié l'association entre anti-inflammatoires non stéroïdiens et la fausse couche spontanée. Nous nous

sommes notamment penchés sur l'implémentation d'hdPS dans le cadre d'un plan d'étude plus complexe visant notamment à limiter les risques de biais de temps immortel ou protopathique. Ce travail s'est appuyé sur le registre EPI-MERES une extraction du SNDS concernant 10 millions de grossesses pendant la période 2010-2021.

CHAPITRE 2

Évaluation des performances des méthodes d'inférence causale pour corriger le biais de confusion : comparaison fondée sur un ensemble de référence de médicaments en lien avec la prématurité

2.1 Introduction

Les bases de données médico-administratives sont de plus en plus utilisées dans les études en pharmacoépidémiologie. Cependant, les estimations causales dans les études observationnelles sont susceptibles d'être biaisées. L'une des principales sources de biais réside dans le fait que l'allocation du traitement

dépend souvent des caractéristiques qui influencent le risque d'événement, appelées « facteurs de confusion ». La prise en compte de ces facteurs est donc nécessaire pour garantir la validité des estimations causales. Différentes méthodes statistiques peuvent être mises en œuvre pour atténuer l'effet de ces facteurs, lorsqu'ils sont identifiés et mesurés, telles que l'ajustement par un modèle de régression, le score de propension, la stratification, etc. Le score de propension a été étendu à la grande dimension avec hdPS dans une version automatisée à large échelle, avec l'objectif de réduire le biais de confusion, qu'il soit issu de facteurs mesurés ou non. Ce n'est que plus récemment que de telles approches exploitant la richesse des données, pilotées par les données, ont été proposées pour d'autres méthodes causales, mais restent encore peu mises en œuvre. Dans cette partie, nous avons porté notre attention sur trois méthodes de plus en plus utilisées dans le domaine de l'inférence causale : l'Estimation Ciblée par le Maximum de Vraisemblance (TMLE pour « Targeted Maximum Likelihood Estimation »), hdPS, et la G-Computation (GC) étendue au contexte des données en grande dimension.

La GC consiste à estimer, pour chaque individu, les issues potentielles correspondant aux valeurs d'intérêt de l'exposition (Robins 1986). Ces issues sont prédites à l'aide d'un modèle d'événement, appelé « modèle Q ». Par ailleurs, le score de propension estime la probabilité d'être exposé, conditionnée par certaines caractéristiques observées chez les individus à l'inclusion (Rosenbaum et Rubin 1983). Ce score permet d'équilibrer les groupes en termes de FDC potentiels et, par conséquent, de réduire ainsi le biais de confusion dans l'estimation de l'effet de l'exposition. Cependant, le SP traditionnel repose sur l'hypothèse d'absence de FDC non mesurés, une condition rarement remplie en pratique,

notamment dans les études exploitant les bases de données de santé (Schneeweiss et Avorn 2005). Avec l'idée d'exploiter la richesse des données disponibles, Schneeweiss et al. (2009) ont ainsi développé le hdPS, une méthode semi-automatique qui emploie un algorithme pour identifier de manière automatisée des FDC potentiels ou des proxys de FDC non mesurés (voir section 1.1.6, page 40). Le SP est ensuite estimé à partir de cet ensemble, ainsi que les covariables prédéfinies par les investigateurs. Inspirés par le même concept, nous avons estimé le modèle Q de la GC à l'aide d'une régression LASSO appliquée à un grand ensemble de variables présélectionnées par la formule de Bross (voir Équation 3, page 41). La dernière méthode étudiée est la TMLE, une méthode qui combine les prédictions du modèle Q avec les SP estimés (Laan et Rubin 2006) pour aboutir à une estimation finale ciblant un estimand théorique d'intérêt. Cette méthode suscite un intérêt croissant dans les études récentes en raison de sa double robustesse, qui permet d'obtenir des estimations consistantes, même si l'un des deux modèles, $\hat{Q}(X)$ (le modèle d'événement) ou $\hat{g}(X)$ (le modèle d'exposition), est incorrectement spécifié (Smith et al. 2023).

Différentes études ont comparé les performances de ces méthodes en terme de biais, mais elles ont abouti à des résultats divergents (Chatton et al. 2020; Kang et Schafer 2007; Kreif et al. 2016; Lendle, Fireman, et Laan 2013; Schuler et Rose 2017). Par exemple, alors que les études de Chatton et al. (2020) ainsi que de Kang et al. (2007) ont montré que la GC avait de meilleures performances, Schuler et al. (2017) ont constaté que la TMLE surpassait la GC et l'IPTW en cas de spécification incorrecte du modèle. Par ailleurs, ces études ont évalué les méthodes uniquement par simulations ou sur de petites bases de données. Ces études ne

rendent pas complètement compte de la complexité des bases de données médico-administratives, qui présentent des relations plus complexes entre les variables ou incluent des expositions rares.

L'objectif de cette partie est de comparer les performances des méthodes susmentionnées sur une grande base de données médico-administrative réelle extraite du SNDS concernant les grossesses entre 2011 et 2014. Cette comparaison se fonde sur un ensemble de médicaments de référence récemment publié dans le contexte du risque d'accouchement prématuré, comprenant des médicaments associés à la prématurité (témoins positifs), et des médicaments non impliqués dans la prématurité (témoins négatifs) (Demailly et al. 2020).

2.2 Méthodologie

2.2.1 Base de données

Nous avons utilisé des données extraites du SNDS concernant les femmes enceintes entre janvier 2011 et décembre 2014. Les grossesses ont été identifiées à l'aide des codes de la Classification internationale des maladies, 10^e révision (CIM-10) et des codes de la Classification commune des actes médicaux (CCAM) liés aux accouchements et aux grossesses dans les bases de données hospitalières PMSI. La liste des codes utilisés se trouve en annexe (voir *Annexe A.1 : Algorithme de sélection des grossesses dans le PMSI*). Les grossesses sans information sur l'âge gestationnel et le nombre de jours depuis les dernières règles ont été exclues. En outre, puisque les accouchements prématurés sont définis comme ceux survenant

entre la 22^e et la 37^e semaine d'aménorrhée, les grossesses de moins de 22 semaines d'aménorrhée ont été exclues.

Les données extraites contiennent les variables suivantes : les délivrances de médicaments en code d'ATC à 5 digits (Code du Système de classification anatomique, thérapeutique et chimique), 16 maladies et conditions survenant au cours de la grossesse, spécifiques ou non de la grossesse enregistrées dans la base de données hospitalière PMSI en diagnostic principale, diagnostic relié ou diagnostic associé (voir la liste des codes CIM-10 des pathologies utilisés dans l'*Annexe A.2 : Liste des codes CIM10 utilisés pour le repérage des pathologies observées au cours de la grossesse*): grossesse multiple, diabète gestationnel, oligohydramnios, polyhydramnios, rupture prématurée des membranes, macrosomie, pré-éclampsie, maladies hémorragiques, hyperemesis gravidarum, cholestase intra-hépatique de la grossesse, insuffisance rénale aiguë, insuffisance hépatique aiguë, pancréatite aiguë, troubles thrombotiques, colique néphrétique, infections ; les variables sociodémographiques : l'âge, le statut de CMU-C (couverture maladie universelle complémentaire), et l'indice territorial de déprivation social (FDep). En France, la CMU-C est un dispositif qui alloue une aide financière aux personnes ayant de faibles revenus, en remboursant la partie complémentaire non couverte par la couverture de base de la Sécurité Sociale. Le FDep est un indicateur de la déprivation socio-économique au niveau communal. Il a été construit à partir de quatre variables sociodémographiques : le revenu médian des foyers, le pourcentage de bacheliers, le pourcentage d'ouvriers parmi les actifs et le taux de chômage (Rey et al. 2009). Pour une commune donnée, un FDep élevé est associé à une déprivation socio-économique plus importante.

En ce qui concerne les expositions médicamenteuses, nous avons considéré, pour chaque grossesse, trois périodes d'exposition médicamenteuse : 3 mois avant la conception (T_0), le premier trimestre (T_1), et 3 mois avant l'accouchement (T_{BD}). Utiliser la même durée pour toutes les périodes d'exposition permet d'éviter le biais causé par le fait que les grossesses à terme ont une probabilité d'exposition à des médicaments plus élevée, car elles ont un suivi plus long que celles qui finissent prématurément (Demailly et al. 2020). Ainsi, à partir de k médicaments, $k \times 3$ variables d'exposition aux médicaments spécifiques à chaque période ont été créées. La période d'exposition d'intérêt dans les analyses est T_{BD} . Une grossesse était considérée comme exposée à un médicament au cours d'une période donnée s'il y avait au moins une délivrance de ce médicament enregistrée pendant cette période. Les expositions médicamenteuses concernant moins de 100 femmes pendant chaque période ont été exclues des analyses.

2.2.2 Réduction de dimension

Il y a potentiellement plusieurs milliers d'expositions médicamenteuses, donc de covariables possibles lors de l'analyse d'une exposition d'intérêt donnée (l'un des médicaments de l'ensemble de référence, à T_{BD}). Afin de réduire le coût computationnel, pour chaque médicament de référence, nous avons utilisé la formule de Bross (voir Équation 3, page 41) pour sélectionner un ensemble de 500 covariables présentant le potentiel de confusion le plus élevé, à partir de 16 complications et maladies, et des covariables d'exposition médicamenteuse. Cet ensemble final de 500 covariables, enrichi avec les 3 covariables

sociodémographiques ont été utilisés en tant que prédicteurs pour toutes les méthodes étudiées : hdPS, GC, et TMLE (Ju et al. 2019).

2.2.3 La régression LASSO (Least Absolute Shrinkage and Selection Operator)

La régression LASSO est une méthode qui permet de construire un modèle de régression dans lequel un grand nombre de variables explicatives peut être inclus. Dans le cas d'un événement binaire, à la différence d'une la régression logistique classique, qui cherche à estimer le coefficient $\hat{\beta}$ maximisant la log-vraisemblance $L = \sum_{i=1}^n y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)$, la régression LASSO cherche à estimer un coefficient $\hat{\beta}$ minimisant la somme $-L + \lambda \sum_{j=1}^p |\beta_j|$ (Tibshirani 1996). Le terme $\lambda \sum_{j=1}^p |\beta_j|$ permet au LASSO de pénaliser la complexité du modèle et de contracter les coefficients de covariables vers 0, et donc d'éviter le sur ajustement lorsque davantage de variables sont incluses dans le modèle. Le degré de la contraction dépend du paramètre de régularisation λ . Plus λ est élevé, plus la contraction est forte, et moins de variables restent dans le modèle. Cette propriété de sélection de variables rend la régression LASSO particulièrement attractive pour l'analyse des données de grande dimension.

Pour déterminer la valeur λ optimale, on fait souvent appel à la validation croisée à k blocs. Cette technique consiste à diviser la base de données en k blocs de taille identique. Le modèle est ajusté à partir de k - 1 blocs (ensemble d'entraînement) tandis que le bloc restant (ensemble de test) sert à évaluer la qualité de l'ajustement (goodness-of-fit) du modèle. L'algorithme est itéré k fois pour

chaque candidat de λ , et à chaque fois, un nouveau bloc est choisi comme l'ensemble de test. Le λ associé à la meilleure qualité moyenne sur les k itérations est choisi. La qualité de l'ajustement peut être évaluée par différentes mesures, telles que l'erreur quadratique moyenne ($\frac{1}{n} \sum_n (\hat{y} - y_{obs})^2$), ou la déviance. Ces mesures priorisent principalement la capacité prédictive du modèle plutôt que la fonction de sélection de variables. Dans notre étude, nous avons cherché à construire un modèle qui retient les variables qui sont potentiellement des FDC ou des proxy des FDC non mesurés plutôt que les variables associées à une bonne prédiction. Ainsi, nous avons utilisé le BIC (Bayesian information criterion) au lieu de la validation croisée pour choisir la valeur optimale du λ (Sabourin, Valdar, et Nobel 2015; Shmueli 2010). Nous avons ajusté une suite de modèles LASSO correspondant à une grille de valeurs de λ . Les covariables retenues par chaque modèle LASSO ont été incluses dans un modèle de régression logistique classique non pénalisé, d'où le BIC_λ est calculé :

$$BIC_\lambda = -2l_\lambda + df(\lambda) \ln(N)$$

Équation 4

Avec l_λ : la log-vraisemblance du modèle de régression logistique classique intégrant les variables sélectionnées par le modèle LASSO avec la pénalité λ .

$df(\lambda)$: le nombre de covariables retenues par le modèle LASSO

$\ln(N)$: le logarithme de la taille de l'échantillon

Le λ associé au BIC_λ le plus faible a été choisi. Le terme $df(\lambda)\ln(N)$ permet de pénaliser la complexité du modèle en terme de nombre de variables explicatives par rapport à la taille de l'échantillon et de prioriser ainsi un modèle relativement parcimonieux. Dans notre étude, ce modèle LASSO-BIC a été utilisé pour sélectionner les variables incluses dans le modèle du SP, en complément de la présélection basée sur la formule de Bross (voir Équation 3, page 41), ainsi que pour prédire les issues potentielles pour la GC.

2.2.4 Score de propension en grande dimension (hdPS)

Comme décrit dans l'introduction (voir la section 1.1.6, page 40), Le hdPS est une méthode qui permet de profiter de la richesse en informations des bases de données de santé pour atténuer le biais causé par les FDC mesurés ou indirectement mesurés (Schneeweiss et al. 2009). Le hdPS repose sur l'idée d'utiliser des combinaisons de variables mesurées pour capturer des informations sur les facteurs non mesurés. L'ensemble de covariables est sélectionné par un algorithme automatique qui s'appuie sur la formule de Bross pour quantifier le potentiel de confusion de chaque covariable (voir Équation 3 page 41). Différentes techniques de sélection des variables ont été développées ces dernières années. Lorsque l'échantillon est petit et l'événement d'intérêt est rare, une priorisation des covariables en fonction de leur association avec l'exposition, $\frac{\Pr(X=1|A=1)}{\Pr(X=1|A=0)}$, peut être utilisée à la place de la quantité calculée par la formule de Bross (Rassen et al. 2011). Des approches alternatives qui recourent à des algorithmes plus complexes, tels que la régression LASSO, la forêt des arbres décisionnels, la régression logistique Bayésienne, etc. pour sélectionner les variables ou pour

estimer le score, ont été étudiées dans la littérature (Karim, Pang, et Platt 2018; Schneeweiss et al. 2017). Dans notre travail, nous avons utilisé l'une de ces approches, la régression LASSO pour construire le SP, avec une modification supplémentaire pour sélectionner les variables pertinentes à intégrer dans cette construction.

En effet, concernant la sélection de variables pour estimer le hdPS, les mêmes recommandations que pour le SP classique (voir la section 1.1.3.1, page 16) sont à suivre. Celles-ci préconisent l'exclusion des variables instrumentales de l'estimation, qui sont associées à l'exposition uniquement mais non à l'événement, du modèle du score (Schneeweiss 2018; Schuster, Lowe, et Platt 2016). Afin d'exclure ces dernières parmi les covariables présélectionnées par la formule de Bross et d'ajuster un modèle plus parcimonieux, nous avons employé un modèle LASSO-BIC régressant le risque d'accouchement prématuré à partir des 3 covariables sociodémographiques et des 500 covariables retenues par la formule de Bross. Les variables sélectionnées par ce modèle LASSO-BIC ont ensuite été utilisées pour l'estimation du SP.

Nous avons utilisé deux méthodes de pondération pour intégrer le SP à l'estimation de l'effet des médicaments de référence : la probabilité de traitement inverse stabilisée (IPTWs) et le poids de chevauchement (OW pour « Overlap weight »). L'OW attribue un poids de $1 - \hat{g}(X)$ aux personnes exposées et de $\hat{g}(X)$ aux personnes non exposées. L'IPTWs, quant à lui, multiplie les poids de l'IPTW classique par la prévalence du traitement réellement reçu par les individus et attribue un poids de $\frac{P(A=1)}{\hat{g}(X)}$ aux exposés et $\frac{P(A=0)}{1-\hat{g}(X)}$ aux non exposés. Ces deux

approches ont été développées pour résoudre le problème lié à la présence de scores extrêmes (F. Li, Thomas, et Li 2019; Xu et al. 2010). En effet, tandis que l'OW assigne un poids variant entre 0 et 1 à chaque individu, éliminant ainsi complètement le risque de poids extrêmes, l'IPTW réduit les poids en les ajustant par la prévalence du traitement, sans toutefois éliminer totalement la présence potentielle des poids extrêmes. La variance des estimations a été calculée en utilisant le bootstrap et les intervalles de confiance à 95 % (IC95 %) sont calculés grâce à la formule suivante :

$$IC95 \% = \exp\left[\frac{\sum_{b=1}^B \log(\widehat{OR}_b)}{B} \pm 1,96 \times sd(\log(\widehat{OR}))\right]$$

Équation 5

Avec \widehat{OR}_b : OR estimé dans une itération de bootstrap b
B : nombre total d'itérations de bootstrap (B = 500)

2.2.5 G-computation

La GC consiste à prédire les résultats potentiels de tous les individus à l'aide d'un modèle d'événement $\widehat{Q}^a(Y|X)$ (Robins 1986). Cette méthode a été initialement proposée pour traiter les FDC dépendant du temps dans des études qui impliquent des expositions variant elles aussi au fil du temps. Cependant, il est possible d'appliquer la méthode aux études dans lesquelles l'exposition reste constante. L'algorithme de la GC a été présenté dans la Figure 1, section 1.1.3.2, page 26.

A l'instar du SP classique, la GC nécessite l'hypothèse d'absence de FDC non mesurés pour que l'estimateur soit consistant. Il a été démontré que l'omission de FDC pouvait détériorer les performances de la GC (Schuler et Rose 2017). Afin de pallier cette limitation, nous nous sommes appuyés sur le concept de proxy des FDC non mesurés et avons mis en œuvre un modèle Q incluant un grand nombre de covariables, dans le but de capturer indirectement les informations sur les FDC non mesurés via leur proxy mesuré dans les bases de données. Pour construire le modèle d'événement et estimer \hat{Q}^1 et \hat{Q}^0 (voir la Figure 1, section 1.1.3.2, page 26), nous avons utilisé le même modèle LASSO-BIC que celui utilisé pour affiner la sélection de variables pour le hdPS (voir la section 2.2.4, page 55). L'odds ratio d'accouchement prématuré du médicament d'intérêt était calculé en utilisant la formule : $\frac{\frac{\hat{\pi}_1}{1-\hat{\pi}_1}}{\frac{\hat{\pi}_0}{1-\hat{\pi}_0}}$, où $\hat{\pi}_1$ et $\hat{\pi}_0$ sont les moyennes des probabilités prédites \hat{Q}^1 et \hat{Q}^0 , respectivement.

Nous avons utilisé le bootstrap avec 500 itérations pour estimer la variance des estimations. L'IC95 % des médicaments de référence a été calculé à l'aide de l'Équation 5.

2.2.6 Estimation Ciblée par Maximum Vraisemblance (TMLE)

La TMLE fait partie des méthodes doublement robustes qui s'appuient sur le principe de combinaison de différents modèles d'estimation pour produire une estimation finale consistante, à condition qu'au moins un des modèles soit correct. La TMLE consiste à calculer \hat{Q}^0 et \hat{Q}^1 via le modèle \hat{Q} de la GC, puis à les réestimer en intégrant les \widehat{SP} pour obtenir \hat{Q}^{0*} et \hat{Q}^{1*} . Le détail de la TMLE est présenté à la

Figure 2, section 1.1.3.3, page 30. Pour réduire le risque de mauvaise spécification des modèles, il est souvent recommandé d'utiliser des algorithmes d'apprentissage statistique, en combinant éventuellement leurs prédictions via la méthode du Super Learner (Naimi, Mishler, et Kennedy 2021). Cependant, l'utilisation de ces algorithmes s'accompagne souvent d'un coût computationnel élevé, ce qui peut poser des difficultés avec de grandes bases de données. Dans notre étude, en raison de la taille importante des données, nous n'avons pas pu recourir au Super Learner et les estimations initiales \hat{Q}^0 et \hat{Q}^1 ont été obtenues à partir du modèle BIC-LASSO.

A l'instar du hdPS et de la GC, nous avons eu recours au bootstrap avec 500 itérations pour estimer la variance des estimations de la TMLE. L'IC95 % a été construit en utilisant l'Équation 5, page 57.

La Figure 5 résume la mise en œuvre des méthodes, avec deux remarques à noter. Premièrement, pour chaque médicament de référence, le modèle LASSO n'a été construit qu'une seule fois. Il a servi à la sélection de variables pour le hdPS et à la prédiction des issues potentielles pour la GC. Deuxièmement, afin de réduire le coût computationnel, nous avons utilisé l'algorithme du hdPS pour présélectionner 500 covariables pour chaque médicament de référence. Chaque itération du bootstrap a été réalisée sur une base de données comprenant les observations bootstrapées, 503 variables (500 covariables hdPS, 3 covariables sociodémographiques), ainsi que le statut d'événement et d'exposition. Chaque itération de bootstrap a conduit à quatre estimations correspondant aux quatre algorithmes étudiés : hdPS-OW, hdPS-IPTWs, GC, et TMLE.

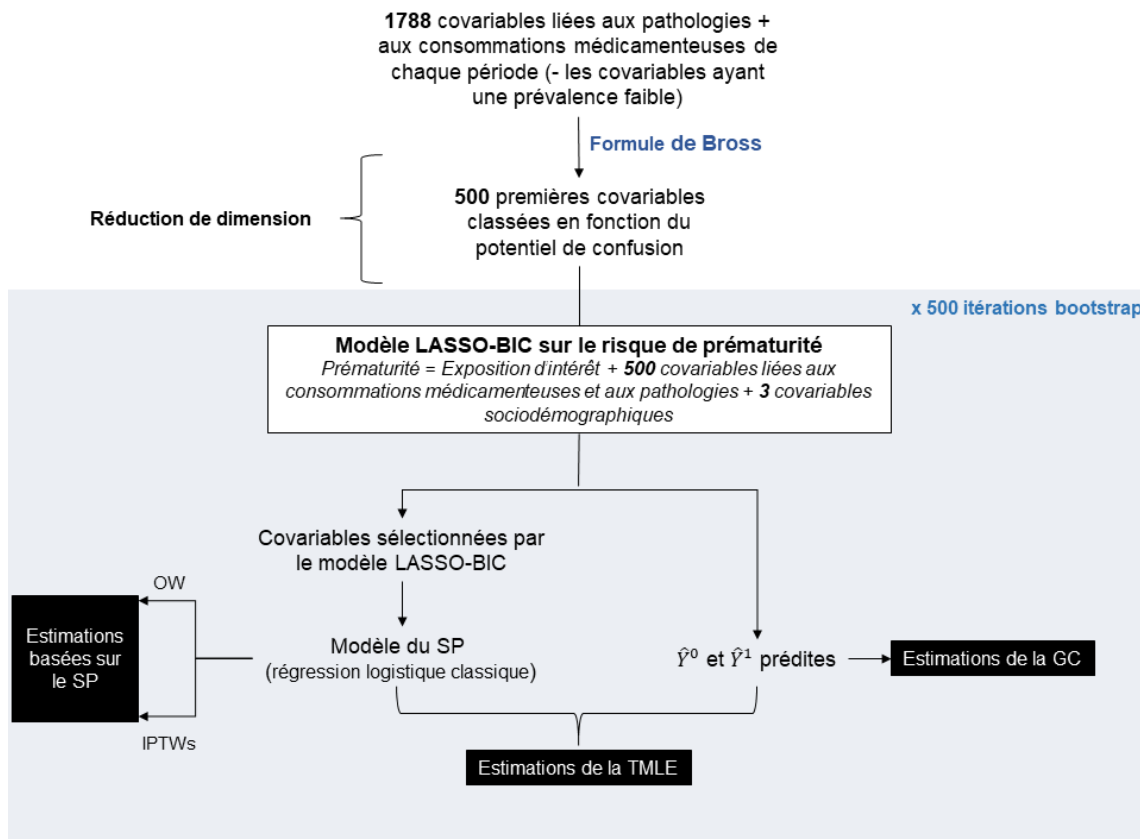


Figure 5. Mise en œuvre des méthodes hdPS, GC et TMLE

2.2.7 Médicaments de référence

Toutes les méthodes ont été appliquées à un ensemble de référence composé de 42 médicaments de référence négatifs et de 13 médicaments de référence positifs concernant le risque d'accouchement prématuré, établi dans une étude méthodologique de détection de signal en pharmacovigilance visant à identifier des médicaments potentiellement associés à un risque d'accouchement prématuré (Demailly et al. 2020). Dans cette dernière, 81 signaux ont été générés à partir de 400 médicaments criblés par la détection automatisée. Ces signaux ont

ensuite été expertisés et classés en trois groupes selon la littérature scientifique publiée et les avis d'une experte en pharmacovigilance : « *Aberrant* » (42 médicaments), « *Potentiel* » (13 médicaments), et « *Incertain* » (26 médicaments). Dans notre étude, les médicaments de référence négatifs correspondent au groupe « *Aberrant* » de Demailly et al., qui regroupe les médicaments non associés à un surrisque d'accouchement prématuré selon les connaissances médicales de l'époque (ceux qui ont un effet nul ou protecteur sur le risque d'accouchement prématuré). Nos médicaments de référence positifs correspondent au groupe « *Potentiel* » de Demailly et al., qui désigne les médicaments connus pour augmenter le risque d'accouchement prématuré. Pour chaque médicament de l'ensemble de référence, nous avons estimé l'odds ratio d'accouchement prématuré entre les femmes exposées au médicament d'intérêt au moins une fois pendant la période T_{BD} et celles non exposées à ce médicament au cours de la même période. La liste des médicaments de référence et leurs codes ATC-5 se retrouvent dans le Tableau 6 (médicaments de référence négatifs) et le Tableau 7 (médicaments de référence positifs).

	ATC	Nom du médicament	n (%)
1	N02BE01	Paracétamol	635,788 (29.26%)
2	A03AX12	Phloroglucinol	545,027 (25.09%)
3	J01CA04	Amoxicilline	206,694 (9.51%)
4	B03BB01	Acide folique	85,889 (3.95%)
5	J01CR02	Amoxicilline et inhibiteur d'enzyme	50,638 (2.33%)
6	C08CA04	Nicardipine	48,462 (2.23%)
7	A02BC05	Esoméprazole	46,850 (2.16%)
8	R03AC02	Salbutamol	45,965 (2.12%)
9	J01DD08	Céfixime	41,147 (1.89%)
10	C08CA05	Nifédipine	40,765 (1.88%)
11	N05BB01	Hydroxyzine	32,379 (1.49%)
12	A10AB05	Insuline asparte	31,250 (1.44%)
13	A03FA03	Dompéridone	27,265 (1.25%)
14	N02BA01	Acide acétylsalicylique	26,565 (1.22%)
15	N02AA59	Codéine, combinaisons excluant les psycholéptiques	22,678 (1.04%)
16	G01AF01	Métronidazole	22,460 (1.03%)
17	A10AC01	Insuline (humaine)	21,476 (0.99%)
18	H02AB01	Bétaméthasone	21,141 (0.97%)
19	R06AE07	Cétirizine	20,560 (0.95%)
20	B03AA02	Fumarate ferreux	20,161 (0.93%)
21	J01FA02	Spiramycine	17,834 (0.82%)
22	B01AB05	Enoxaparine	17,404 (0.8%)
23	J01XE01	Nitrofurantoïne	14,415 (0.66%)
24	D07AB08	Désonide	10,247 (0.47%)

25	J01FA07	Josamycine	7,992 (0.37%)
26	G03DA04	Progestérone	6,889 (0.32%)
27	A05AA02	Acide ursodésoxycholique	6,219 (0.29%)
28	N07BA01	Nicotine	5,558 (0.26%)
29	A02BC02	Pantoprazole	5,385 (0.25%)
30	G03DA03	Hydroxyprogestérone	3,638 (0.17%)
31	J01FA06	Roxithromycine	3,458 (0.16%)
32	R03AK08	Formotérol et béclométasone	2,122 (0.1%)
33	R05DA20	Alcaloïdes d'opium et dérivés, antitussifs combinés	2,060 (0.09%)
34	D06BB10	Imiquimod	1,601 (0.07%)
35	J01MA01	Ofloxacine	1,235 (0.06%)
36	R06AX26	Fexofénadine	780 (0.04%)
37	N02AA01	Morphine	587 (0.03%)
38	G03DB01	Dydrogestérone	280 (0.01%)
39	N02CC06	Elétriptan	260 (0.01%)
40	J07AM01	Toxine tétanique	218 (0.01%)
41	R03AL01	Fénotérol et bromure d'ipratropium	179 (0.01%)
42	B01AB06	Nadroparine	167 (0.01%)

Tableau 6. Liste des médicaments de référence négatifs. La colonne n(%) représente le nombre de femmes exposées (n) et la prévalence de l'exposition (%)

	ATC	Nom du médicament	n (%)
1	N05BA04	Oxazépam	5,015 (0.23%)
2	N03AX09	Lamotrigine	3,805 (0.18%)
3	N05BA12	Alprazolam	3,710 (0.17%)
4	N05BA08	Bromazépam	3,610 (0.17%)
5	N06AB10	Escitalopram	2,669 (0.12%)
6	A07EC02	Mésalazine	2,291 (0.11%)
7	N05CF02	Zolpidem	2,227 (0.1%)
8	P01BA02	Hydroxychloroquine	1,957 (0.09%)
9	N05BA11	Prazépam	1,308 (0.06%)
10	N05AA01	Chlorpromazine	1,056 (0.05%)
11	N05BA01	Diazépam	993 (0.05%)
12	G03AA07	Lévonorgestrel et éthinyloestradiol	592 (0.03%)
13	N06AX03	Miansérine	157 (0.01%)

Tableau 7. Liste des médicaments de référence positifs. La colonne n(%) représente le nombre de femmes exposées (n) et la prévalence de l'exposition (%)

2.2.8 Mesure de performances

Les méthodes basées sur les PS, GC et TMLE ont été comparées à un modèle de régression logistique classique, appelé « modèle brut », incluant l'exposition d'intérêt (à T_{BD}) ainsi que les expositions au même médicament à T_0 , T_1 et les trois covariables sociodémographiques.

Étant donné que les méthodes ciblent différents estimands théoriques et que les vraies valeurs de l'effet sont inconnues, nous ne comparons pas leurs performances en termes de valeurs d'estimations ponctuelles, mais plutôt leur capacité à classer correctement les médicaments selon leur potentiel d'augmenter le risque d'accouchement prématuré. Pour chacun des 55 médicaments de référence, nous avons estimé l'intervalle de confiance à 95 % (IC95 %) des OR avec toutes les méthodes. Nous avons ensuite calculé la proportion de médicaments dans chaque catégorie (voir le Tableau 8). Les méthodes ont été comparées en fonction de leur proportion de faux positifs et de vrais positifs. Ces derniers sont définis, respectivement, comme des associations dont les bornes inférieures de l'IC95 % sont supérieures à un dans l'ensemble de référence négatif et dans l'ensemble de référence positif.

		Type de médicament de référence	
		Médicaments de référence négatifs	Médicaments de référence positifs
Résultat des méthodes	Borne supérieure du IC95 % < 1 (effet protecteur)	Vrais négatifs	Faux négatifs
	L'IC95 % inclut 1 (effet nul)	Vrais négatifs	Faux négatifs
	La borne inférieure du IC95 % > 1 (surrisque)	Faux positifs	Vrais positifs

Tableau 8. Catégories des résultats obtenus par les méthodes

En outre, pour chaque méthode, nous avons comparé la distribution des $\log(\text{OR})$ estimés séparément chez les références négatives et positives. Toutes les analyses statistiques ont été réalisées avec le logiciel R (v3.4.4) en utilisant les packages `glmnet` (v2.0-16) et `adapt4pv` (v0.2-2) pour les régressions LASSO-BIC.

2.3 Résultats

La base de données finale concernait 2 172 702 grossesses, après exclusion de 41 grossesses sans information sur l'âge gestationnel et le nombre de jours depuis les dernières règles. L'âge médian des mères était de 30 ans (écart interquartile : 26-33 ans). La prévalence des accouchements prématurés était de 7,74 %. Il y avait 1 778 variables liées aux expositions médicamenteuses spécifiques à chaque

période, et 16 variables liées aux maladies et complications survenant au cours de la grossesse. Parmi les médicaments de référence négatifs, le paracétamol avait la plus forte prévalence (n = 635 788, 29,26 %), tandis que la nadroparine était la moins courante (0,01 %, n = 167). Vingt-six des 42 médicaments de référence négatifs concernaient moins de 1 % des grossesses. Pour les références positives, l'oxazépam était le plus fréquent (0,23 %, n = 5 015) et la miansérine la moins fréquente (0,01 %, n = 157). La prévalence de tous les médicaments de référence est détaillée dans le Tableau 6 et Tableau 7, page 63-64.

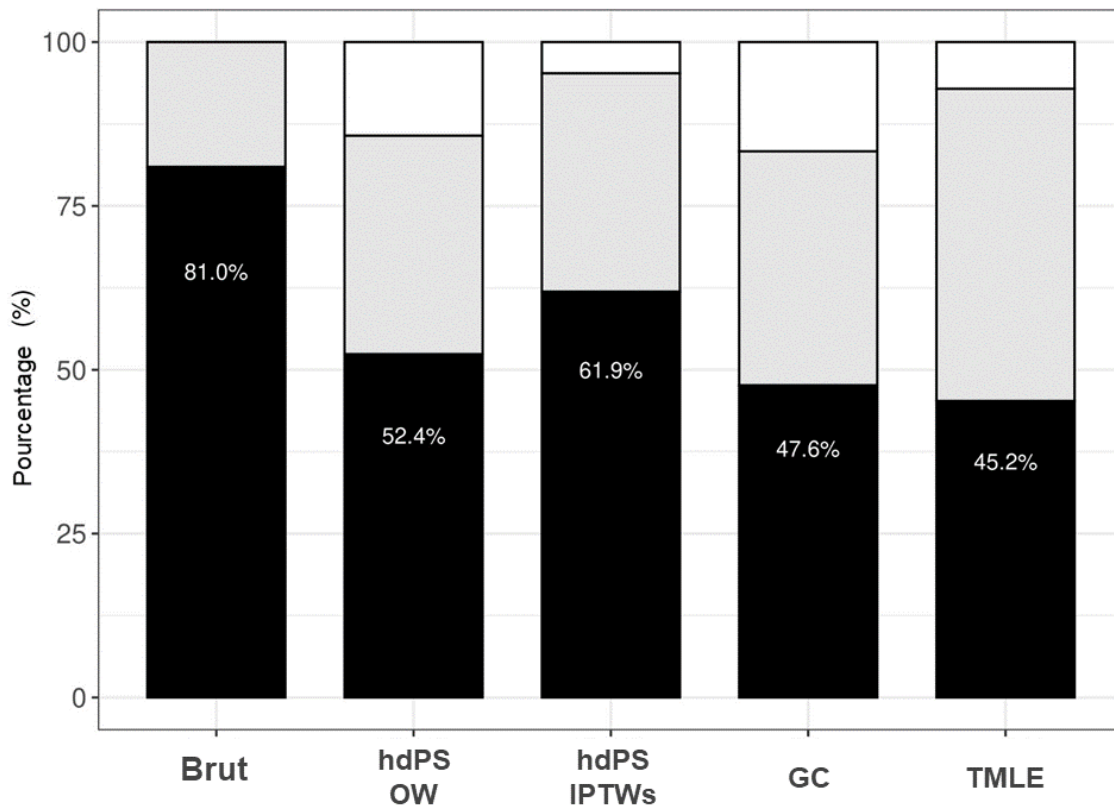


Figure 6. Proportion de références négatives dans chaque catégorie : *Vrai négatif* (protecteur) : blanc ; *Vrai négatif (nul)* : gris ; *Faux positif* : noir

La Figure 6 montre que dans les analyses des références négatives, le modèle brut a généré la proportion la plus élevée de faux positifs (81,0 %), suivi par la SP-IPTWs (61,9 %). L'intégration des informations sur l'exposition dans les estimations initiales $\hat{Q}^0(a, \mathbf{X})$ via une estimation doublement robuste a légèrement réduit la proportion de faux positifs (47,6 % pour la GC contre 45,2 % pour la TMLE). Toutes les méthodes ont identifié 14 faux positifs communs (Figure 7). Par ailleurs, l'ensemble des faux positifs détectés par le modèle brut (38 sur 42) incluait ceux produits par les autres méthodes (Figure 7).

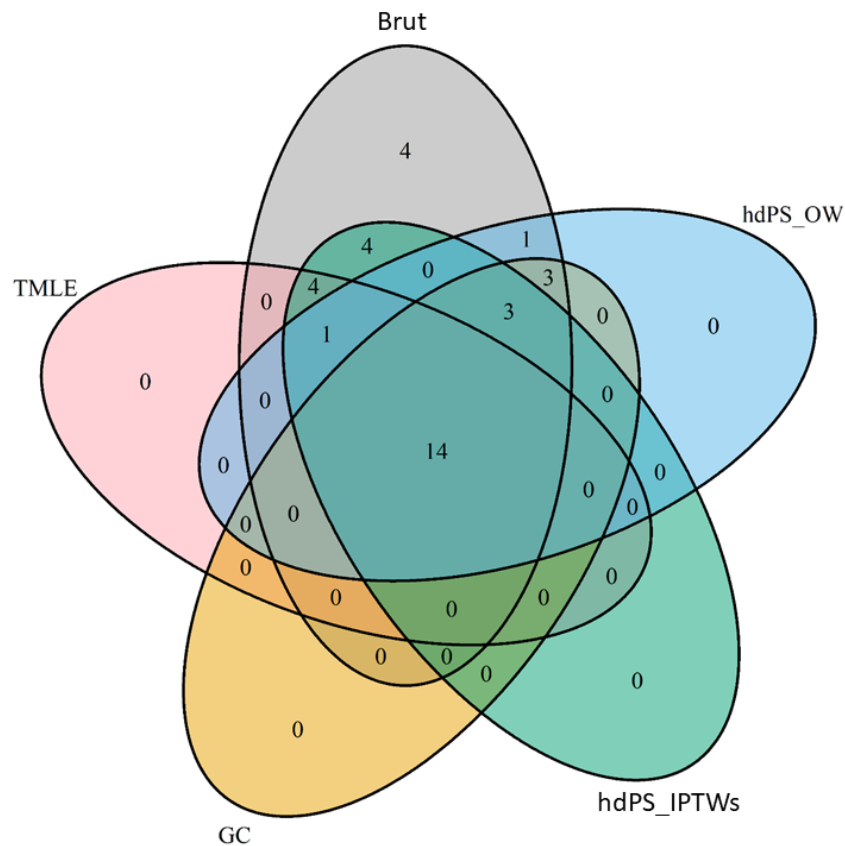


Figure 7. Diagramme de Venn des faux positifs dans les médicaments de référence négatifs

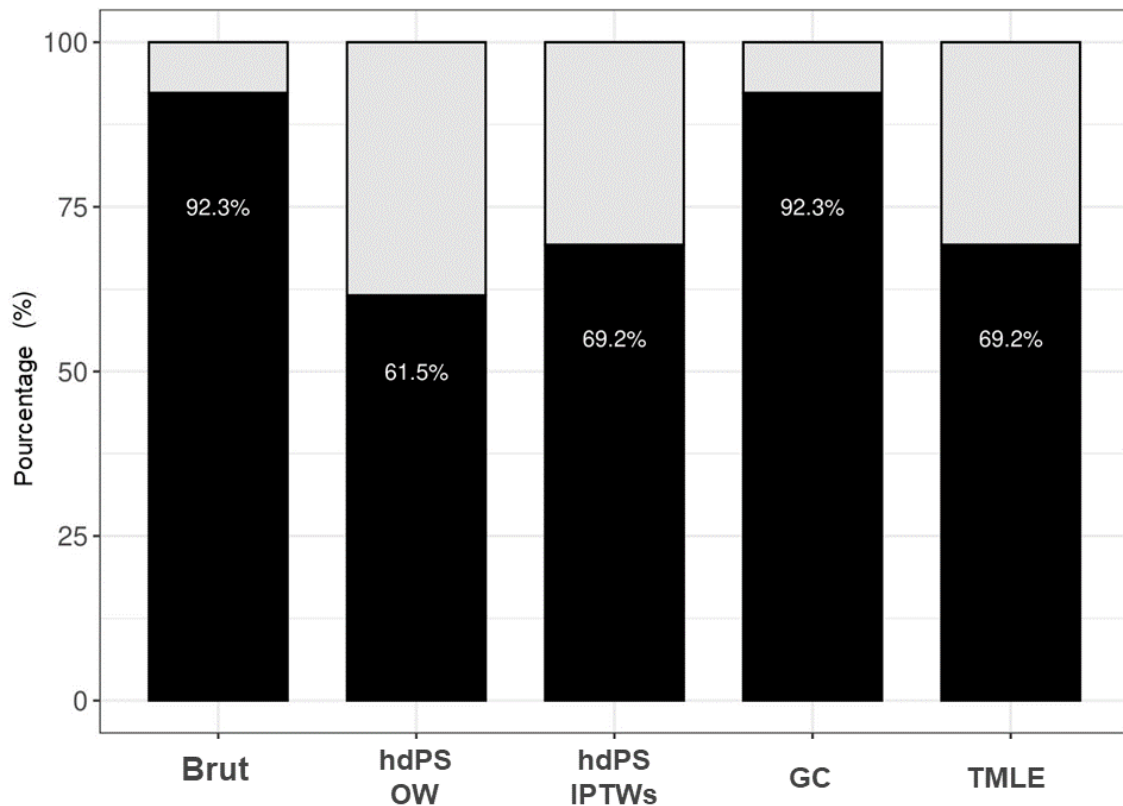


Figure 8. Proportion de références positives dans chaque catégorie. *Faux négatif (protecteur)* : blanc ; *Faux négatif (nul)* : gris ; *Vrai positif* : noir

La Figure 8 montre que, dans les analyses des références positives, le modèle brut et la GC ont produit la proportion la plus élevée de vrais positifs (92,3 %). La SP-OW a conduit à la proportion la plus faible (61,5 %), suivie par les méthodes SP-IPTWs (69,2 %) et TMLE (69,2 %). La GC et le modèle brut ont identifié le même ensemble de vrais positifs. Ce dernier incluait tous les vrais positifs détectés par les autres méthodes (voir la Figure 9).

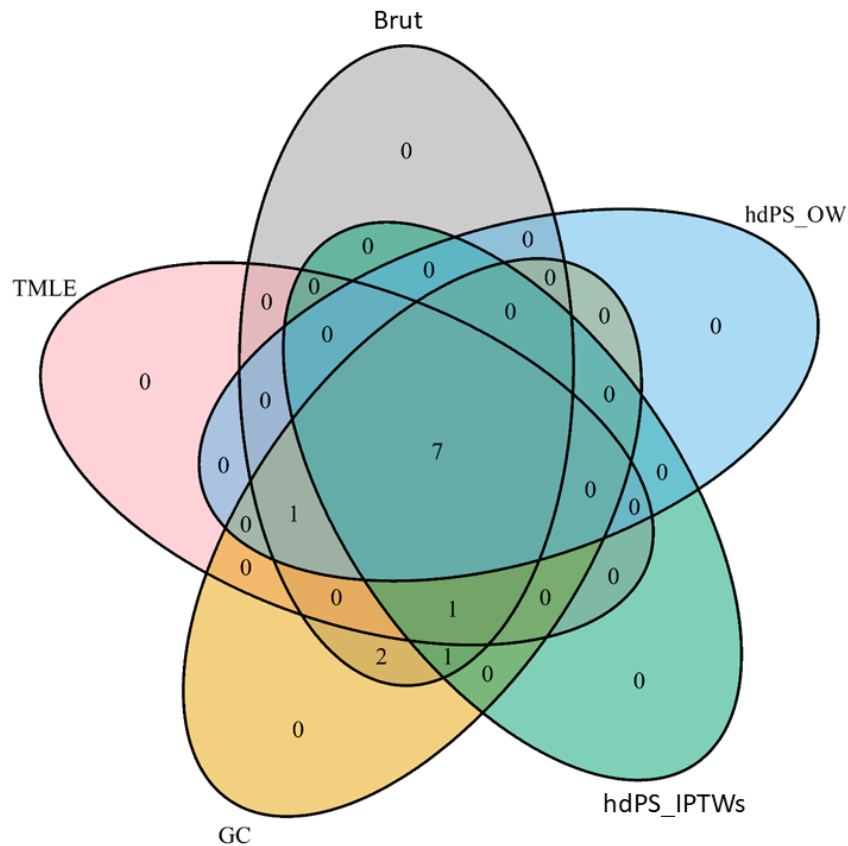


Figure 9. Diagramme de Venn des vrais positifs dans les médicaments de référence positifs

La Figure 10 illustre la distribution des $\log(\text{OR})$ estimés des références positives et des références négatives entre les méthodes. Toutes les méthodes étaient capables de discriminer les références négatives des références positives, mais la GC semble produire une distinction plus nette (p-value du test de Wilcoxon : Brut 0,018 ; TMLE 0,010 ; SP-IPTWs 0,010 ; SP-OW 0,014 ; GC 0,001).

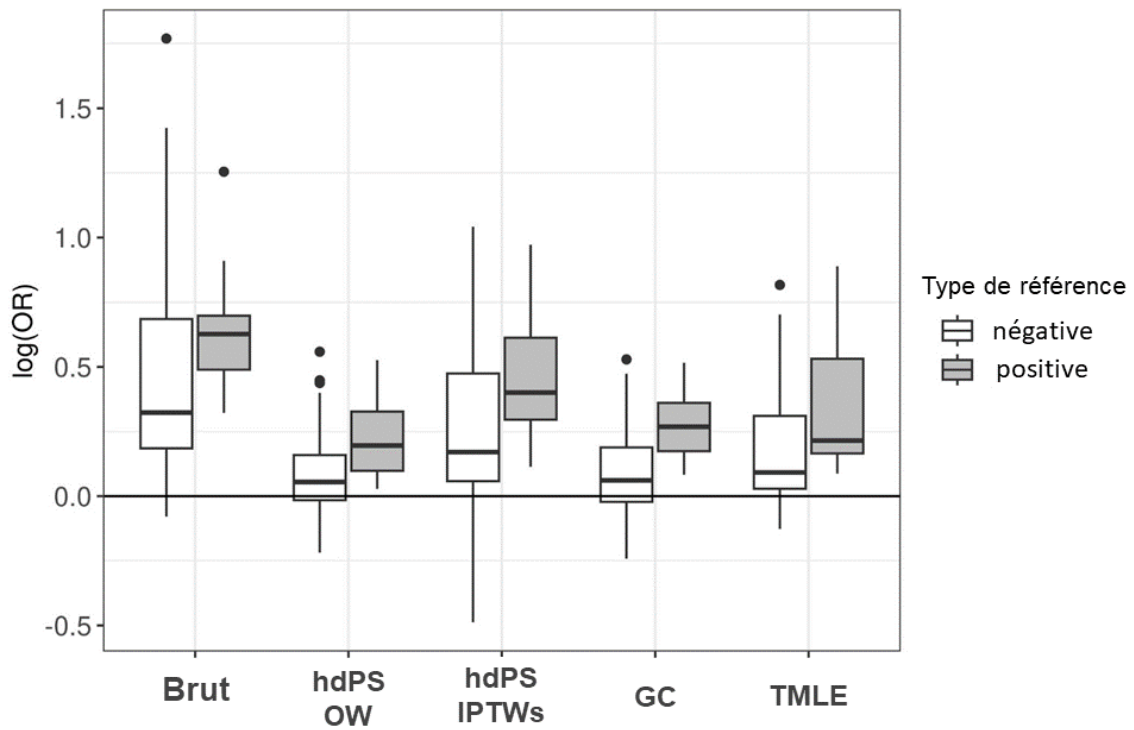


Figure 10. Distribution des $\log(\text{OR})$ estimés des références négatives et des références positives en fonction des méthodes

L'estimation ponctuelle et l'IC95 % de tous les médicaments de référence sont présentés dans la Figure 11 pour les références négatives et la Figure 12 pour les références positives.

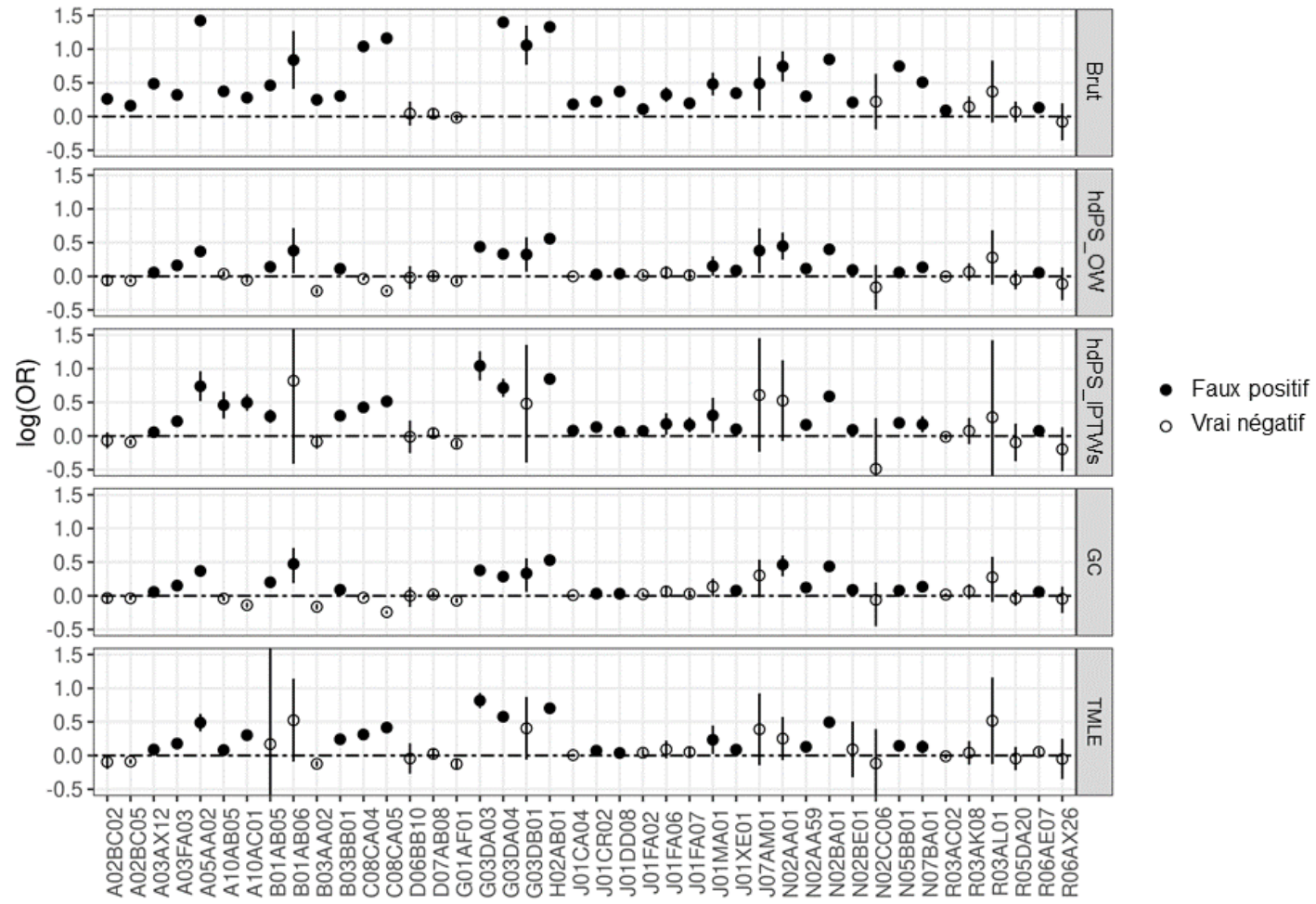


Figure 11. Log(OR) estimé des médicaments de référence négatifs

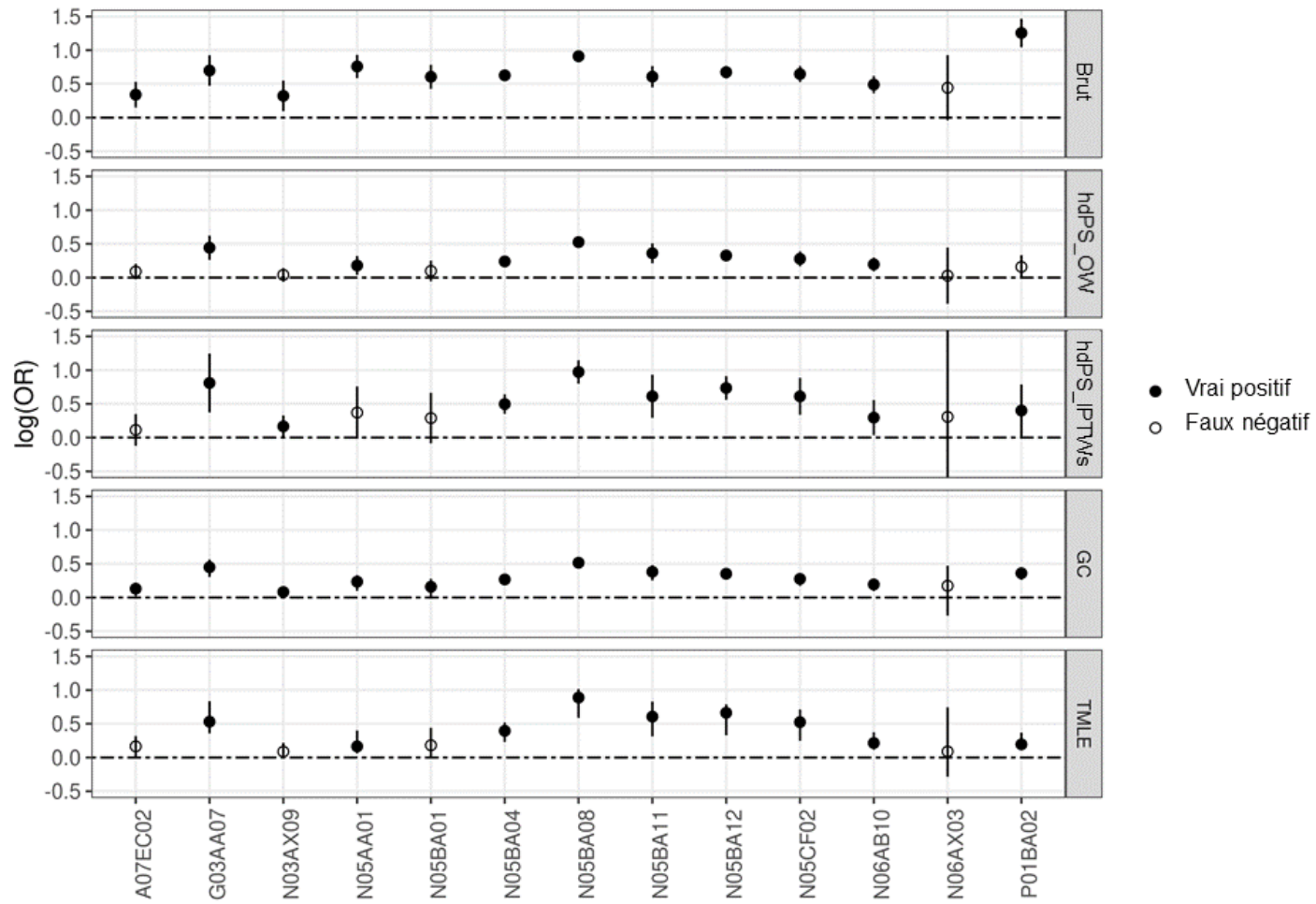


Figure 12. Log(OR) estimé des médicaments de référence positifs

2.4 Discussion

Malgré les avantages fournis par les bases de données médico-administratives, leur utilisation dans les études de pharmacoépidémiologie peut être entravée par la présence de FDC non mesurés. Pour traiter cette question, nous avons comparé les performances de trois méthodes : le hdPS (avec l'IPTW stabilisée et l'OW), la GC, et la TMLE, adaptées à l'exploitation des données de grande dimension du SNDS via l'utilisation du modèle LASSO, une technique permettant la sélection de variables. La comparaison a été mise en œuvre sur des données réelles concernant l'ensemble des grossesses enregistrées dans le SNDS sur la période 2011-2014. Elle s'appuyait sur un ensemble de médicaments de référence établi dans le cadre d'une étude de pharmacovigilance sur le risque d'accouchement prématuré (Demailly et al. 2020). Nous avons comparé la proportion de faux et de vrais positifs parmi les références négatives et les références positives. Nos résultats ont montré la supériorité des trois méthodes par rapport au modèle brut dans la réduction de biais, avec un avantage pour la GC. Bien que produisant une proportion de faux positifs légèrement plus élevée que la TMLE (Figure 6), sa proportion de vrais positifs était nettement supérieure aux autres méthodes (Figure 8).

Le nombre de faux positifs reste relativement élevé après l'application de toutes les méthodes (Figure 6). Cela pourrait s'expliquer en partie par la construction de l'ensemble des références négatives qui se compose de signaux fallacieux de pharmacovigilance (voir la section 2.2.7, page 60) (Demailly et al. 2020), d'où la

possibilité de sources de biais résiduelles non résolues par les méthodes. Par exemple, la progestérone et l'hydroxyprogestérone, qui ont été incorrectement classées comme positives par toutes les méthodes, sont souvent prescrites à des femmes qui présentent un risque élevé d'accouchement prématuré (antécédents d'accouchements prématurés spontanés, col de l'utérus court, etc.) pour prévenir cet événement (Jain et al. 2020; Norman 2020; Romero et al. 2014; Thangaraju et al. 2021). Une source du biais de sélection pourrait donc persister après avoir été partiellement traitée par les méthodes, comme en témoigne la diminution des estimations ponctuelles des OR. Bien que les méthodes ne parviennent pas à éliminer tous les biais, elles ont clairement contribué à l'amélioration du contrôle des biais, comme en témoignent les résultats (voir la Figure 6, page 67).

Concernant les médicaments identifiés comme négatifs par les méthodes, leur prévalence semble être l'un des facteurs impactant cette classification, car une faible prévalence est souvent associée à un large intervalle de confiance, donc à une association non statistiquement significative. Cela est particulièrement vrai pour les faux négatifs, comme l'illustre l'exemple de la miansérine, qui est la moins prévalente des références positives (0,01 %). Elle a été classée à tort comme négative par toutes les méthodes. Il convient de noter que la GC a produit des IC95 % plus étroits que les autres méthodes, même pour les médicaments rares. Ceci pourrait expliquer sa puissance supérieure dans la détection des références positives à faible prévalence, telles que le diazépam (0,05 %), qui a été correctement classé comme positif uniquement par la GC.

La moins bonne performance de la TMLE comparée à la GC n'est pas en accord avec certaines études antérieures (Ju et al. 2019; Naimi, Mishler, et Kennedy 2021; Schuler et Rose 2017; Zivich et Breskin 2021). Cependant, ces dernières n'ont pas comparé les méthodes sur des bases de données de grande dimension réelles et ne se sont pas intéressées à la question de leur capacité à atténuer le biais causé par les FDC non mesurés dans ce contexte. Elles ont plutôt cherché à démontrer la double robustesse de la TMLE. En effet, les meilleures performances de la TMLE ont souvent été mises en évidence dans des contextes où il y avait au moins un modèle correct, ou en absence de FDC non mesurés (Lendle, Fireman, et Laan 2013; Schuler et Rose 2017). En pratique, la validité des modèles n'est jamais connue dans les études observationnelles et les meilleures performances de la TMLE par rapport aux autres méthodes sont moins évidentes à démontrer sur données réelles (Lendle, Fireman, et Laan 2013). Une étude de simulation a montré que les estimateurs doublement robustes avec deux modèles incorrects étaient moins performants que la GC avec un modèle de bonne capacité prédictive (Kang et Schafer 2007). De même, Kreif et al. (2016) ont comparé les méthodes sur des données simulées et montré que combiner deux modèles Q et SP incorrects ne diminuait pas le biais par rapport à l'utilisation séparée de chaque modèle. Pour résoudre ce problème, certaines études ont suggéré l'utilisation de méthodes plus complexes, telles que le boosting, la forêt d'arbres décisionnels, le modèle LASSO, etc., souvent combinées via un estimateur ensembliste, appelé Super Learner (Naimi, Mishler, et Kennedy 2021; Pirracchio, Petersen, et van der Laan 2015; Schuler et Rose 2017). Ces méthodes plus flexibles ne posent aucune hypothèse paramétrique sur la relation entre les variables, ce qui permet de diminuer le risque d'ajuster incorrectement les modèles. Cependant, dans les études utilisant

de grandes bases de données, comme la nôtre, leur utilisation peut être limitée par leur coût computationnel élevé. Par ailleurs, la moindre-performance dans notre étude de la TMLE en comparaison de la GC semble être attribuable en partie aux difficultés rencontrées dans la construction des SP pour les expositions médicamenteuses rares, ce qui a conduit à des valeurs extrêmes du SP même avec la stabilisation des poids (Hajage et al. 2016). En effet, les méthodes non doublement robustes basées sur le SP étaient les moins performantes dans notre étude, notamment lorsque l'IPTWs a été utilisé. L'intégration du SP en présence de valeurs extrêmes a été décrite comme ayant un effet négatif sur les performances des estimateurs doublement robustes (Lendle, Fireman, et Laan 2013).

Les méthodes évaluées dans notre étude ont demandé des ressources computationnelles importantes. En effet, il nous a fallu plus de deux semaines de calcul parallélisé sur un serveur équipé de 100 unités de calculs et de 1 TB de mémoire vive. Ce coût computationnel était dû à la grande taille des données (2 172 702 individus et 3 375 covariables), au nombre de médicaments de référence testés tour à tour (55 médicaments), et au nombre de méthodes évaluées. Le nombre de 500 covariables retenues via la formule de Bross a été choisi comme un compromis entre la tentative de maximiser la capture des FDC potentiels et un coût computationnel raisonnable. Ce choix est cohérent avec l'utilisation classique du hdPS dans la littérature où le nombre de variables retenues se trouve généralement entre 200 et 500 (Tazare et al. 2022). En outre, il a été avancé que la variation du nombre de variables retenues à cette étape, entre 100 et 1 000, ne changeait pas de manière significative l'estimation de l'effet (Rassen et al. 2023).

Une part très importante du coût computationnel est due à la mise en œuvre du bootstrap pour estimer la variance des estimateurs. Bien qu'il existe des estimateurs de la variance qui requièrent moins de ressources computationnelles pour certaines méthodes, telles que la courbe d'influence efficiente pour la TMLE, ou l'estimateur sandwich pour les pondérations par le SP, nous avons réalisé le bootstrap pour garantir une comparabilité adéquate entre les méthodes. Le choix du LASSO comme méthode d'apprentissage a lui aussi été influencé par des questions de coût computationnel dans la mesure où cette méthode est relativement peu coûteuse. En revanche, le choix du BIC plutôt que de la validation croisée pour la détermination de l'hyperparamètre a été motivé par l'objectif de sélection de variables (Shmueli 2010).

L'évaluation d'un grand nombre de médicaments de référence nous a contraint à adopter un plan d'analyse standardisé et simple au regard de ce qui aurait été fait pour de véritables études de pharmacoépidémiologie s'intéressant séparément et uniquement à chacune des 55 associations. En particulier, le modèle brut incluait toujours le même ensemble de covariables plutôt que différents ensembles adaptés au médicament d'intérêt. Bien que ce modèle ne fournisse pas d'information sur l'efficacité des méthodes en comparaison avec une étude de pharmacoépidémiologie réelle, il sert à donner des valeurs de référence contre lesquelles les méthodes sont comparées. Par ailleurs, malgré l'utilisation d'un grand nombre de covariables, il n'a pas été possible d'éliminer tous les biais, comme en témoignent les faux positifs persistant après les analyses. Cela peut être dû d'une part au fait que ces méthodes ne peuvent récupérer de l'information que sur des FDC qui sont directement ou indirectement mesurés (Ishimaru, 2020;

Rassen et al., 2023). D'autre part, il existe éventuellement d'autres sources de biais qui ne peuvent être correctement pris en compte qu'au travers de la mise en œuvre de plan d'analyses spécifiques (voir le Tableau 1, section 1.1.1.2, page 14).

Néanmoins, ces limites, ayant trait au plan d'étude choisi et à la richesse de la base de source, affectent l'ensemble des méthodes comparées de manière identique et ne semblent donc pas pouvoir invalider les conclusions sur les performances relatives de ces dernières.

2.5 Conclusion

Dans ce travail, nous avons démontré que l'application des méthodes statistiques exploitant les informations riches dans les bases de données médico-administratives permettait de réduire les biais dans une étude observationnelle. La GC a donné de meilleures performances que celles de la TMLE et du hdPS. Cette méthode est également intéressante en raison de sa mise en œuvre simple dans le cadre d'un événement d'intérêt binaire. Cependant, malgré l'utilisation de ces méthodes, un biais résiduel subsiste dans les estimations, qui nécessiteraient d'autres plans d'étude plus raffinés en pharmacoépidémiologie.

CHAPITRE 3

Score de propension en grande dimension dans le cadre de modèles complexes : étude des anti-inflammatoires non stéroïdiens et du risque de fausse couche spontanée

3.1 Introduction

Les anti-inflammatoires non-stéroïdiens (AINS) sont des médicaments largement utilisés pour traiter la douleur, les réactions inflammatoires, et la fièvre. Ils figurent également parmi les médicaments les plus utilisés chez les femmes enceintes. Une étude multinationale a estimé que 2,2 % des femmes enceintes étaient exposées aux AINS au moins une fois au cours de leur grossesse entre octobre 2011 et février 2012 dans les pays d'Europe orientale (Lupattelli et al. 2014). En France, une

étude de Demailly et al. (2017) indiquait que 5 % de femmes enceintes étaient exposée aux AINS au cours du premier trimestre, 1,5 % durant le second trimestre, et 0,5 % durant le troisième trimestre entre l'année 2011 et 2014. Une tendance similaire a été trouvée dans une étude de Louchet et al. (2024) estimant que 1,28 % des femmes enceintes étaient exposées aux AINS pendant le deuxième et le troisième trimestre sur la période 2013-2019.

Plusieurs études ont démontré l'association entre l'utilisation des AINS au cours de la grossesse et différents effets néfastes sur le fœtus et le nouveau-né. L'un des mécanismes proposés pour expliquer ces effets serait l'inhibition de la synthèse des prostaglandines au niveau de l'utérus, où ces substances jouent un rôle crucial dans l'implantation du fœtus et son développement (Edwards et al. 2012; Leathers et Rogers 2023; Ying et al. 2022). Des études antérieures ont montré que l'exposition aux AINS au cours du deuxième et du troisième trimestre pouvait entraîner différents effets indésirables. Généralement, l'utilisation des AINS chez les femmes enceintes est fortement déconseillée à partir de la vingtième semaine de grossesse en raison d'un risque accru d'oligohydramnios et de dysfonction rénale chez le fœtus. Elle est contre-indiquée à partir du troisième trimestre en raison du risque de fermeture prématurée du canal artériel et de complications cardio-pulmonaires chez le fœtus (New Zealand Medicines and Medical Devices Safety Authority 2024; Medicines and Healthcare products Regulatory Agency 2024; FDA 2024). En France, l'utilisation des AINS est déconseillée à partir de la 14^e semaine de grossesse (4^e mois de grossesse) et contre-indiquée à partir de la 23^e semaine de grossesse (6^e mois de grossesse) (ANSM 2024). Récemment, Quantin et al. (2021) ont montré un lien entre l'exposition précoce aux AINS (entre

la conception et la 22^e semaine de grossesse) et l'accouchement prématuré. De plus, plusieurs études ont trouvé que l'exposition aux AINS pendant cette période pourrait également accroître le risque de malformations congénitales ou de fausses couches spontanées (FCS). Cependant, les conclusions sur ces effets demeurent controversées à ce jour. Concernant le risque de malformations congénitales, alors que Choi et al. (2023) ont trouvé un lien entre l'exposition aux AINS au cours du premier trimestre et des malformations congénitales majeures, Ericson et Källén (2001) ont trouvé un lien uniquement avec les malformations cardiaques. Ces associations n'ont pas été observées dans les études de van Gelder et al. (2011) et de Niesel et al. (2001). Par ailleurs, dans cette dernière, les auteurs ont trouvé un lien potentiel entre les AINS et la FCS dans un délai de 12 semaines suivant l'exposition. L'association entre l'exposition précoce aux AINS et la FCS a été également observée dans certaines études (D.-K. Li et al. 2018; Bérard et Kori 2012; Nakhai-Pour et al. 2011), mais pas dans d'autres (Dathe et al. 2018; Daniel et al. 2015b; Edwards et al. 2012). Le présent travail vise à évaluer l'association entre les AINS et le risque de FCS.

Concernant le lien entre AINS et FCS, l'une des causes possibles de ces résultats inconsistants réside dans les différentes fenêtres d'exposition utilisées. Par exemple, Daniel et al. (2015b) ont fait plusieurs analyses ignorant les délivrances d'AINS enregistrées dans les un à trois jours précédant la date de la FCS pour prendre en compte un biais d'indication. En revanche, d'autres études ont considéré toutes les délivrances d'AINS avant la date de FCS pour définir les individus exposés (D.-K. Li et al. 2018). Le risque de biais protopathique provient du fait que les AINS peuvent être prescrits pour traiter les symptômes précoces

d'une FCS débutant (par exemple, douleur abdominale, fièvre) (van Gelder, Lupattelli, et Nordeng 2023). Par conséquent, la fréquence de l'exposition aux AINS peut augmenter dans les jours précédant une FCS, ce qui aboutit à une surestimation de son effet. Pour éviter ce type de biais, nous avons appliqué l'approche « lag-time » qui consiste à reporter tous les temps d'exposition d'un certain nombre de jours (période de latence). Cette méthode repose sur l'hypothèse qu'une FCS qui survient peu après l'exposition ne peut pas être imputée à cette dernière en raison de son délai trop court (Tamim, Monfared, et LeLorier 2007). La méthode sera décrite plus en détail dans la section 3.2.2.2 ci-dessous.

L'incohérence des résultats de la littérature peut également s'expliquer, en partie, par des différences de choix de modèles utilisés pour estimer l'effet de l'exposition. En effet, divers types de modèles ont été employés pour estimer l'effet des AINS, tels que la régression logistique classique (Nakhai-Pour et al. 2011; Nielsen et al. 2001), la régression de Cox classique (D.-K. Li et al. 2018), et la régression de Cox avec variable dépendant du temps (Daniel et al. 2015a). Il a été montré que l'absence de prise en compte des variations temporelles du statut d'exposition peut introduire un biais du temps immortel dans l'estimation de l'effet (Suissa 2008; Daniel et al. 2015a; Jones et Fowler 2016). En effet, si le statut d'exposition de toute la période de suivi est défini de manière fixe après un certain nombre de délivrances du médicament d'intérêt, une personne doit survivre assez longtemps afin de recevoir le nombre requis de délivrances et être ainsi considérée comme exposée. Si la période non exposée initiale n'est pas prise en compte correctement dans les analyses, par exemple en l'excluant ou en la

classant à tort comme « exposée », le risque est de surestimer l'effet du médicament en sous-estimant la durée de survie du groupe non exposé ou de surestimer celle du groupe exposé (Suissa 2008). L'utilisation d'un modèle de Cox avec variables dépendant du temps permet d'atténuer ce biais (Jones et Fowler 2016).

La source des données pourrait être une autre raison qui explique le désaccord entre les études. Les études de cohorte peuvent probablement mieux capturer les FCS, notamment celles très précoces, comme l'ont suggéré Li et al. (2018). Néanmoins, l'échantillon dans ces études était plus petit et probablement moins représentatif de la population générale comparé aux études utilisant les bases de données médico-administratives comme l'étude de Nakhai Pour et al. (2011). Par exemple, dans leur cohorte, Dathe et al. (2018) ont constaté que le niveau d'éducation des femmes incluses dans leur étude était plus élevé que la population générale.

En France, le SNDS couvre 99 % de la population. L'utilisation de cette base nous permet d'étudier l'effet des AINS sur une large population de femmes enceintes, quasi-exhaustive. Cependant, comme pour toutes les autres bases de données médico-administratives, la présence de FDC non mesurés dans le SNDS est une question importante à résoudre lors de son exploitation.

L'objectif de ce chapitre est d'étudier le lien entre AINS et FCS. Il est aussi d'implémenter le hdPS dans un plan d'étude plus complexe que celui du chapitre précédent, à savoir un modèle de Cox avec variable dépendant du temps et différentes périodes de latence. Nous avons comparé les résultats entre les

analyses avec et sans hdPS pour examiner les apports de l'intégration des données de grande dimension aux résultats des analyses. Ce travail s'appuie sur le registre EPI-MERES, une extraction du SNDS, concernant les femmes enceintes en France pendant la période de janvier 2013 à décembre 2019.

3.2 Méthodologie et résultats par analyses

3.2.1 Base de données

Construit par le GIS EPI-PHARE, le registre EPI-MERES inclut toutes les grossesses en France enregistrée dans le SNDS sur une période de 12 années, de 2010 à 2020 (Miranda et al. 2023). Dans cette partie, nous nous sommes intéressés à un sous-échantillon du registre EPI-MERES, constitué des grossesses ayant abouti à une naissance vivante, à un mort-né ou à une FCS spontanée sur la période de janvier 2013 à décembre 2019. Ce choix a été fait dans le but de simplifier les analyses en évitant la présence de risques compétitifs (les femmes qui ont fait une interruption volontaire de grossesse ou une interruption médicalisée de grossesse ne peuvent avoir une FCS) et des particularités liées à la pandémie COVID en 2020. Nous avons exclu les femmes exposées à des médicaments tératogènes et foetotoxiques entre deux semaines avant la date de conception et la fin de la grossesse. Ces médicaments figurent dans la liste établie dans une étude antérieure de Louchet et al. (2024) concernant les consommations médicamenteuses des femmes enceintes en France entre 2013 et 2019 (Voir *Annexe B.1 : Liste des médicaments foetotoxiques* et *Annexe B.2 : Liste de médicaments tératogènes*). Dans notre étude,

une FCS était définie comme un avortement spontané (code CIM-10 O03) survenant avant la 20^e semaine de grossesse. De plus, dans l'analyse principale, nous avons considéré uniquement les FCS de plus de 6 semaines de grossesse, car les FCS survenant très tôt peuvent ne pas être diagnostiquées, et être mal enregistrées dans les bases de données médico-administratives (San Lazaro Campillo et al. 2021).

Notre base de données contient des informations sur les variables sociodémographiques, les comorbidités, les consultations chez les généralistes et les spécialistes, les actes biologiques, et les consommations médicamenteuses sur une année précédant la grossesse et au cours de celle-ci. Les covariables sociodémographiques comprennent l'âge de la mère, l'indice de déprivation sociale, FDep, et le statut de bénéficiaire de la CMU-C. Les comorbidités sont les pathologies détectées pendant une année précédant la grossesse dans les tables de cartographie. Dans le SNDS, les tables de cartographie contiennent les indicateurs de 58 maladies construits à partir des algorithmes développés par la Cnam. Ces algorithmes se basent sur diverses informations, telles que les codes diagnostiques, les médicaments, etc. pour repérer les maladies (Rachas et al. 2022). Les 58 maladies peuvent être regroupées en grandes catégories ci-dessous :

- Maladies cardio-vasculaire et les traitements du risque vasculaire
- Diabète
- Cancers
- Maladies psychiatriques et les traitements psychotropes,

- Maladies neurologiques,
- Maladies respiratoires chroniques,
- Maladies inflammatoires, les maladies rares, ou l'infection par le virus de l'immunodéficience humaine (VIH),
- Insuffisance rénale chronique terminale,
- Maladies du foie ou du pancréas,
- Affections de longue durée (dont ALD 31 et 32) pour d'autres causes, la maternité, les hospitalisations hors pathologies repérées,
- Hospitalisations pour Covid-19,
- Traitements antalgiques ou anti-inflammatoires.

En plus de ces dernières, l'état d'hypercoagulabilité et l'hypothyroïdie sont deux FDC potentiels souvent prises en compte dans les études antérieures (Daniel et al. 2014; 2015b; Nakhai-Pour et al. 2011). Etant donné qu'elles ne figurent pas dans les tables de cartographie et sont mal codées dans les données hospitalières, nous avons construit leur indicateur à partir des médicaments associés. Une personne était considérée comme atteinte d'hypothyroïdisme lorsqu'au moins deux prescriptions d'hormones thyroïdiennes (code ATC « H03AA ») avaient été délivrées une année avant la conception. De même, l'état d'hypercoagulabilité était identifié par la délivrance d'au moins deux prescriptions d'anticoagulants (code ATC « B01 ») durant la même période.

3.2.2 Analyse avec sélection experte des facteurs de confusion

3.2.2.1 Modèle de Cox avec variable dépendant du temps

Dans les études observationnelles, il est possible que l'exposition ne survienne pas toujours à la date d'entrée des individus, mais au cours du suivi chez certains individus. Une analyse ayant une définition fixe de l'exposition peut considérer la période avant l'exposition comme exposée, et entraîner ainsi un biais de temps immortel dans l'estimation de l'effet. La Figure 13 présente un exemple d'une analyse biaisée par le biais de temps immortel.

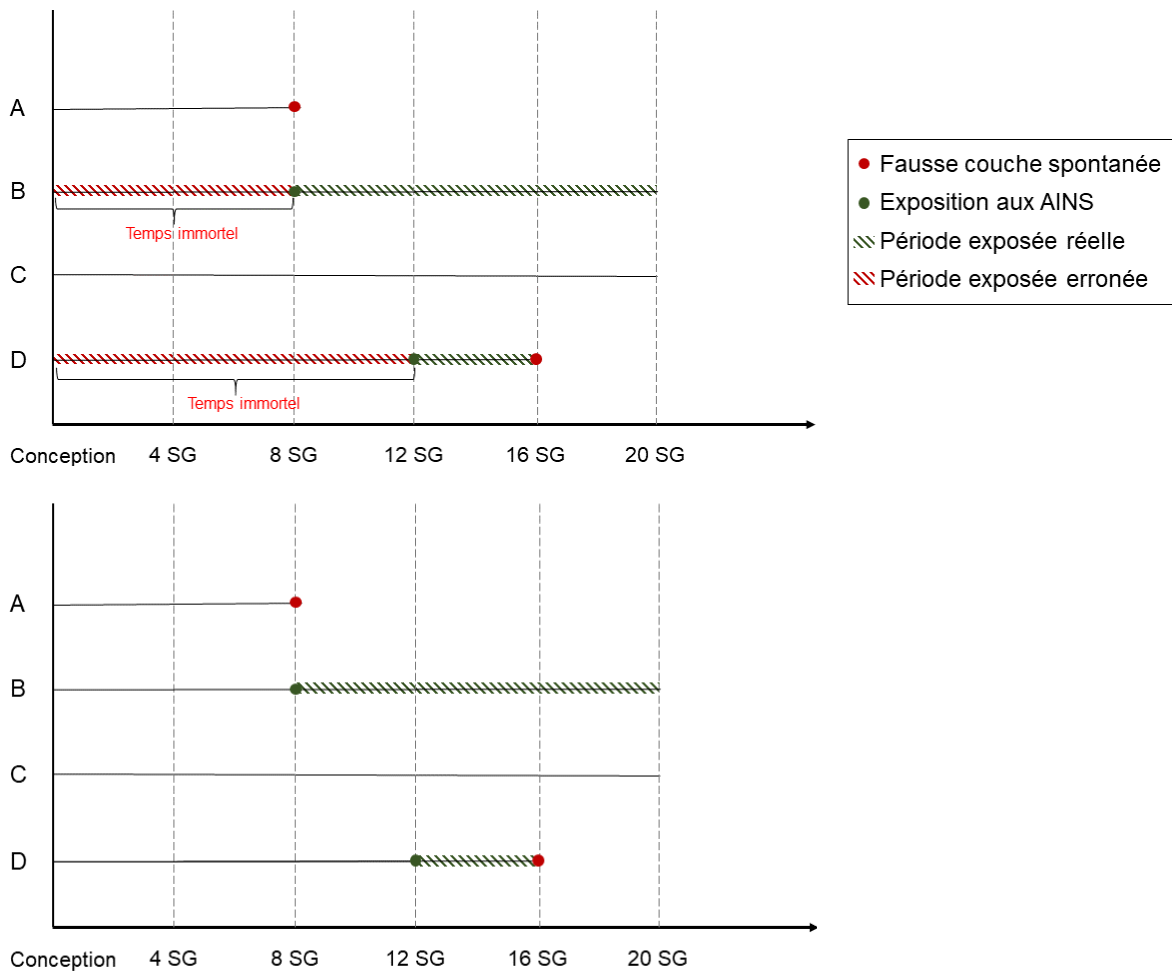


Figure 13. Biais de temps immortel. La figure du haut présente une analyse dans laquelle une femme est considérée comme exposée si elle a au moins une délivrance d’AINS pendant les 20 premières semaines de grossesse (les femmes B et D). Dans cette analyse, la période non exposée des femmes B et D est classée comme « exposée », ce qui entraîne une surestimation du temps de survie des femmes exposées. La figure du bas présente une analyse qui prend en compte le changement de l’exposition et classe correctement la période non exposée des femmes B et D.

Afin de prendre en compte la variation temporelle de l'exposition, nous avons utilisé un modèle de Cox avec variable d'exposition dépendant du temps. Ce modèle est construit à partir d'une table dans laquelle chaque ligne correspond à un intervalle de suivi d'une personne. Une personne peut avoir plusieurs lignes si elle a plusieurs intervalles de suivi. Ce format permet de mettre à jour les valeurs des variables dépendant du temps dans chaque intervalle. Dans les études longitudinales qui disposent de mesures régulières des covariables au cours du suivi, comme les études de cohorte, les intervalles de suivi de chaque individu dans la table de données sont définis par le rythme des visites. Dans notre étude, les femmes enceintes n'avaient pas des visites régulières pour mesurer les covariables, ce qui est fréquent lors de l'utilisation des bases de données médico-administratives, nous avons ainsi divisé le suivi des femmes en intervalles de sept jours correspondants aux semaines de grossesse dans la table de données adaptée à la construction d'un modèle de Cox avec variable dépendant du temps (voir la Figure 14).

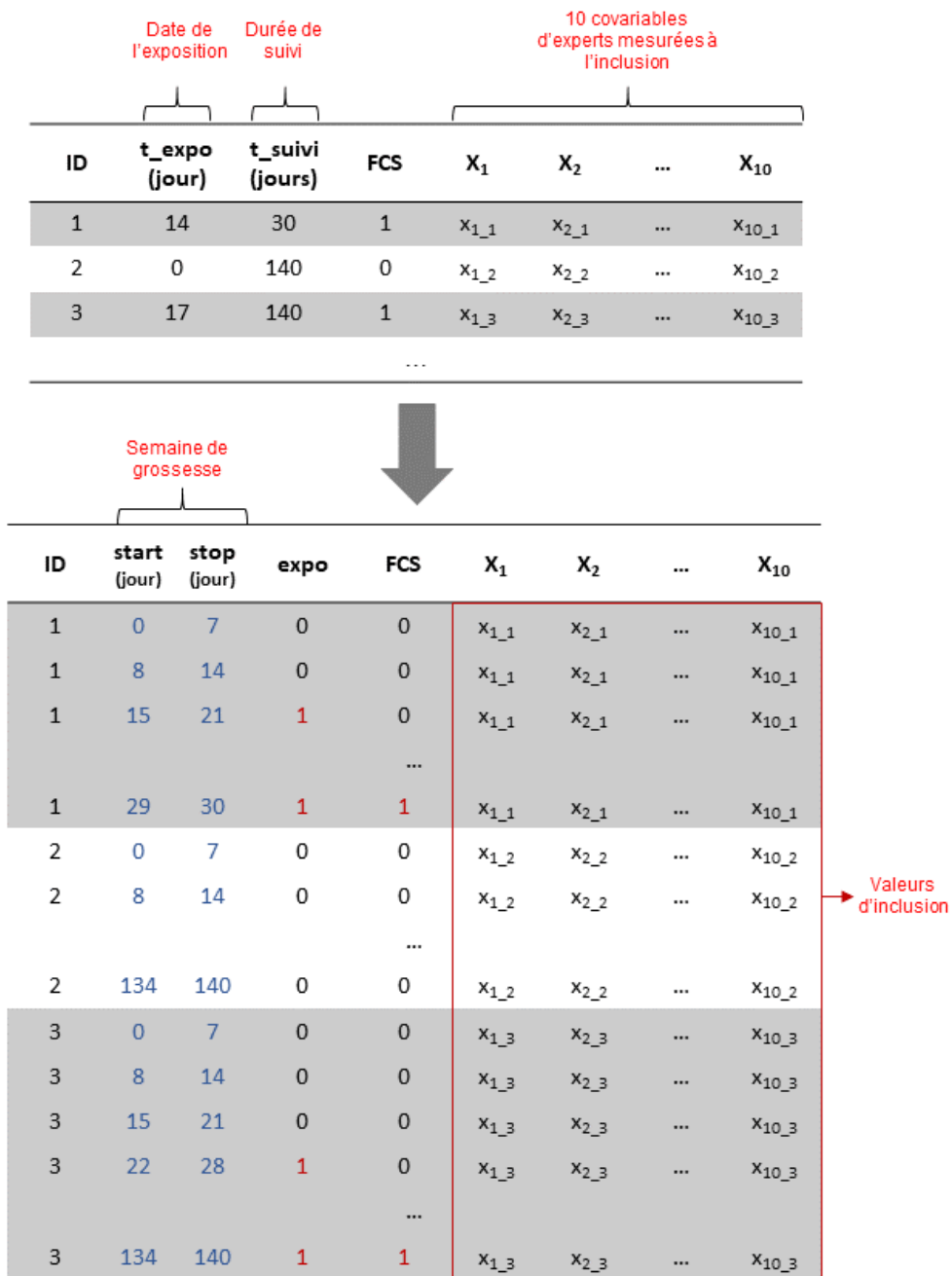


Figure 14. Construction de la table des données dépendant du temps

Dans ce travail, les grossesses ont été suivies depuis la conception jusqu'à la fin de la 20^e semaine de grossesse ou jusqu'à une FCS, selon le premier événement survenu. Le suivi a été censuré à la 20^e semaine de grossesse. Au sein de chaque semaine, nous avons pris en compte le changement du statut d'exposition et d'événement et maintenu les autres covariables à leur valeur d'inclusion. Une femme ayant pris sa première délivrance d'AINS entre la semaine $t - 1$ et t est considérée non exposée au début de la semaine $t - 1$ et exposée au début de la semaine t .

D'après la littérature actuelle, nous avons inclus les variables suivantes dans le modèle de Cox avec variable dépendant du temps en tant que FDC potentiels : l'âge au début de grossesse en tranches d'âges de 5 ans, l'indicateur de CMU-C, l'indice FDep, le nombre de grossesses depuis 2010, l'historique des avortements ou des FCS depuis 2010, et la présence des facteurs suivants pendant une année avant la date de conception : la prise d'AINS, le diabète, les maladies inflammatoires, l'hypothyroïdie, ou l'état d'hypercoagulabilité.

3.2.2.2 Approche « lag-time » et biais protopathique

Le biais protopathique résulte de la prise d'un médicament pour traiter les symptômes précoces de la maladie étudiée, ce qui aboutit à une augmentation des prises de médicament dans le groupe des personnes présentant l'événement même si le médicament n'a aucun effet sur le risque de maladie (Acton, Willis, et Hennessy 2023). Par exemple, dans les études concernant la FCS, les antibiotiques ou les analgésiques sont souvent susceptibles de présenter ce biais, car ils peuvent être prescrits pour traiter les symptômes d'une FCS débutante (van Gelder,

Lupattelli, et Nordeng 2023). Dans notre étude, il est possible que les AINS soient prescrits pour traiter des symptômes précoces d'une FCS débutante au moment de la délivrance (douleur abdominale, fièvre, etc.). Pour limiter le risque de biais protopathique, il est possible de restreindre la définition des personnes exposées aux indications non liées à la maladie. Néanmoins, cette restriction est compliquée en pratique car les motifs de délivrances ne sont pas toujours disponibles dans les bases de données, y compris dans le SNDS. Pour prévenir ce biais, Daniel et al. (2015b) ont exclu les délivrances d'AINS enregistrées dans les trois jours précédant la FCS et ont trouvé une association non significative entre les AINS et la FCS. Cette approche repose sur le choix délicat d'une période, uniformément appliquée à toutes les expositions afin de pallier l'absence d'information sur l'indication chez tous les individus exposés. Basée également sur l'idée d'exclure les délivrances susceptibles d'induire un biais protopathique, l'approche de « lag-time » (période de latence) proposée par Tahim et al. (2007) introduit une période de latence (période « lag ») pendant laquelle toutes les expositions sont ignorées. A la différence de la technique effectuée par Daniel et al., l'approche « lag-time » applique la période de latence à toutes les personnes, qu'elles présentent l'événement ou pas, ce qui permet d'éviter un traitement différencié de l'exposition des individus selon la survenue de l'évènement.

En pratique, la période de latence peut être introduite après l'exposition d'intérêt dans une étude de cohorte ou avant la date d'index (la date où l'évènement est diagnostiqué dans la majorité des études) dans une étude de cas-témoins (Hicks et al. 2023). Tous les événements qui surviennent pendant la période de latence sont considérés comme non exposés en supposant que le lien causal ne

soit pas plausible en raison de courts délais entre l'exposition et la survenue de l'événement. La Figure 15 (A et B) illustre deux exemples d'application de cette méthode dans une étude cas-témoin et une étude de cohorte. La Figure 15A représente une étude cas-témoins où un témoin est apparié à un cas à la date de l'événement du cas (date index). Une période de latence est introduite immédiatement avant la date index, et toutes les expositions qui surviennent pendant cette période sont ignorées. La Figure 15B représente une étude de cohorte qui évalue les deux médicaments X et Y. La date index est celle de la première délivrance des médicaments d'intérêt. La période de latence est introduite après la date index. Tous les individus qui présentent l'événement pendant la période de latence sont exclus.

Il convient de noter que nous n'avons pas appliqué l'approche comme celle que Hicks et al. (2023) ont décrite pour une étude de cohorte dans la mesure où notre étude ne compare pas deux médicaments. Nous avons retardé les dates de délivrance d'un nombre de jours égal à la période de latence choisi pris en compte la période exposée à partir cette nouvelle date. Ainsi, les FCS survenant pendant la période de latence étaient considérées comme non exposées (voir la Figure 15C).

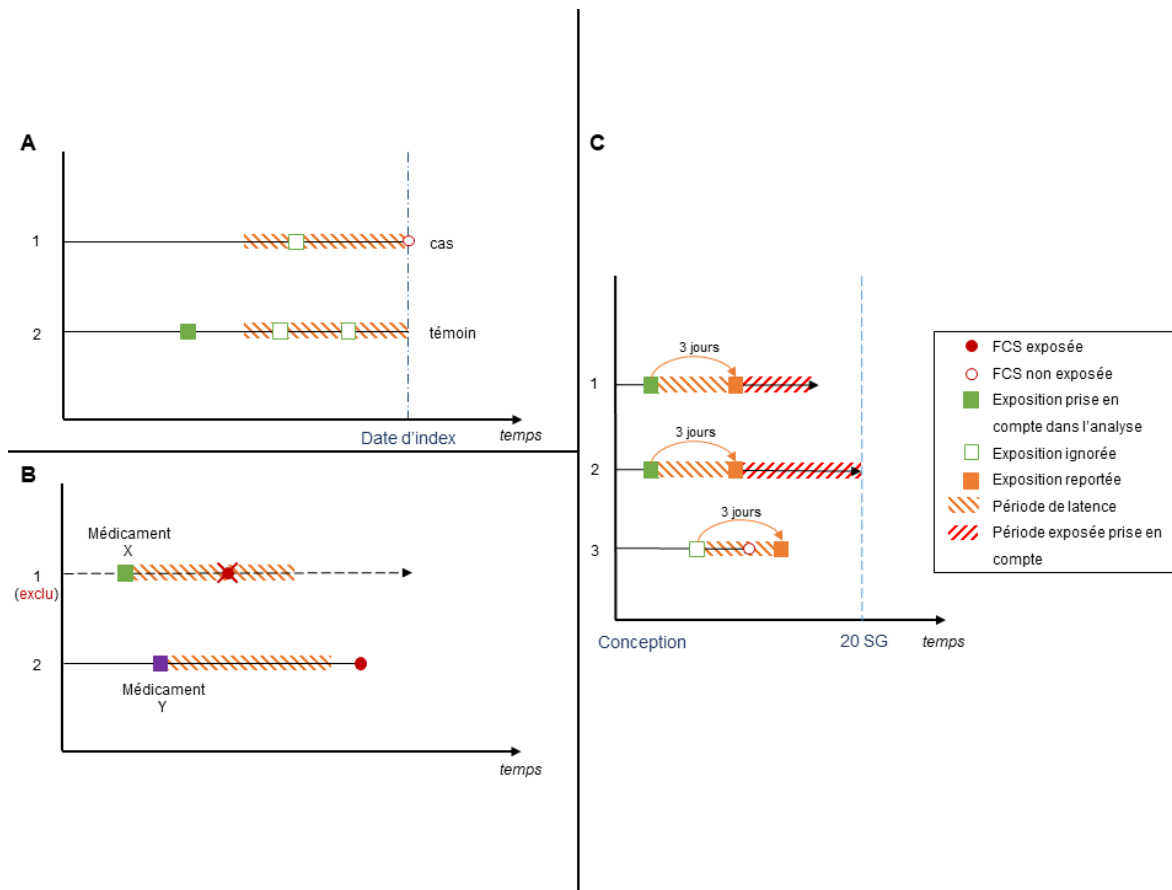


Figure 15. Application de l'approche "lag-time". La figure 15A représente l'exemple d'une étude cas-témoins dans laquelle la personne 1 est considérée non exposée car sa délivrance tombe sur la période de latence, la personne 2 est considéré exposée. La figure 15B représente un exemple d'une étude de cohorte dans laquelle la personne 1 est exclue car elle présente l'événement pendant la période de latence après son exposition. La figure 15C montre le cas de notre étude, dans laquelle une période de trois jours a été ajoutée à toutes les dates d'exposition. Les FCS survenues au cours de cette période ont été considérés non exposés.

En ignorant l'exposition et l'événement d'écart $\leq n$ jours de latence, l'approche « lag-time » postule implicitement qu'il faut un certain temps (n jours) après l'exposition pour que le médicament provoque son effet et affecte la survenue de

l'événement s'il a un lien causal avec ce dernier. Cette hypothèse est notamment plausible dans les études sur les cancers, qui adoptent souvent une période de latence de plusieurs années (Hicks et al. 2023). En pratique, le choix d'une durée pertinente repose sur les connaissances des mécanismes d'action de la molécule et de la physiopathologie de la maladie étudiée. Ce choix peut avoir un impact important sur les résultats d'analyses : une période de latence trop courte ne permet pas d'éliminer efficacement le biais protopathique alors qu'une période trop longue peut diminuer ou annuler l'effet du médicament (voir la Figure 16). Cette difficulté dans le choix de cette période a conduit des auteurs à recommander des études de sensibilité faisant varier cette période autour d'une valeur initiale (Arfè et Corrao 2016).

Dans ce travail, nous avons choisi une période de latence de 3 jours en nous appuyant sur l'étude de Daniel et al. (2015b), qui ont trouvé que le lien entre les AINS et la FCS n'était plus significatif lorsque les femmes exposées aux AINS au cours des trois jours précédant la FCS ont été exclues. Nous avons aussi conduit une analyse de sensibilité faisant varier cette période afin d'évaluer l'impact de ce choix.

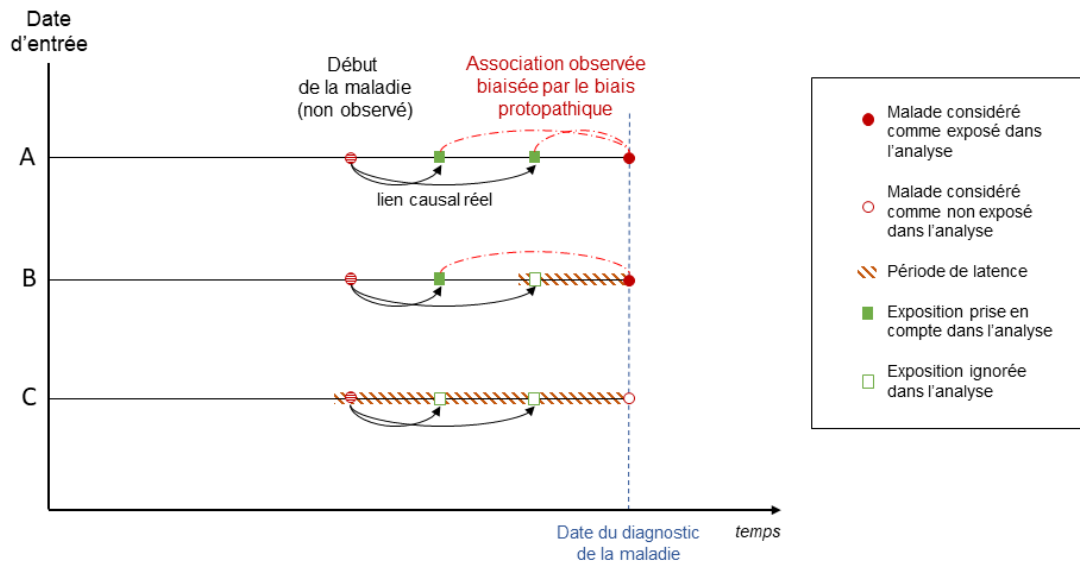


Figure 16. Biais protopathique et approche "lag-time". L'étude B utilise une période de latence trop courte qui ne couvre pas complètement le délai entre l'apparition des symptômes précoces et le diagnostic de la maladie. Par conséquent, le cas est toujours assigné au groupe exposé et le biais protopathique est toujours présent dans l'analyse. La période de latence dans l'analyse C comprend le délai entre les manifestations précoces et le diagnostic de la maladie, et permet donc d'atténuer le risque de biais protopathique.

3.2.2.3 Définition de l'exposition

Dans notre étude, une femme est considérée comme exposée aux AINS si elle a reçu au moins une délivrance d'AINS (code ATC «M01A ») entre les 2 semaines précédant la date de conception et la 20^e semaine de grossesse. Cette femme sera

considérée comme exposée à partir de la date de sa première délivrance + l'éventuelle période de latence jusqu'à la fin de son suivi. La durée de la période de latence est de trois jours dans l'analyse principale et varie entre zéro (sans latence) à sept jours dans les analyses de sensibilité. Le délai de 2 semaines avant la date de conception permet de prendre en compte les effets des expositions aux AINS autour de la conception et de capturer des expositions au début de la grossesse, ce qui est le cas pour des délivrances prises avant la conception dont la durée dépasse la date de conception.

Enfin, nous avons réalisé des analyses pour chaque molécule d'AINS qui ont une prévalence > 1/10000 au cours de la grossesse. Dans ces analyses, nous avons comparé les femmes exposées à chaque molécule d'intérêt avec les femmes qui n'ont utilisé aucun AINS au cours de la grossesse.

3.2.2.4 Résultats

Il y avait 4 857 907 grossesses qui ont été incluses dans l'analyse principale, dont 4 671 366 (96,16 %) grossesses se terminant par une naissance vivante, 22 875 (0,47 %) par un mort-né et 163 666 (3,37 %) par une FCS. Le diagramme de flux est présenté dans la Figure 17.

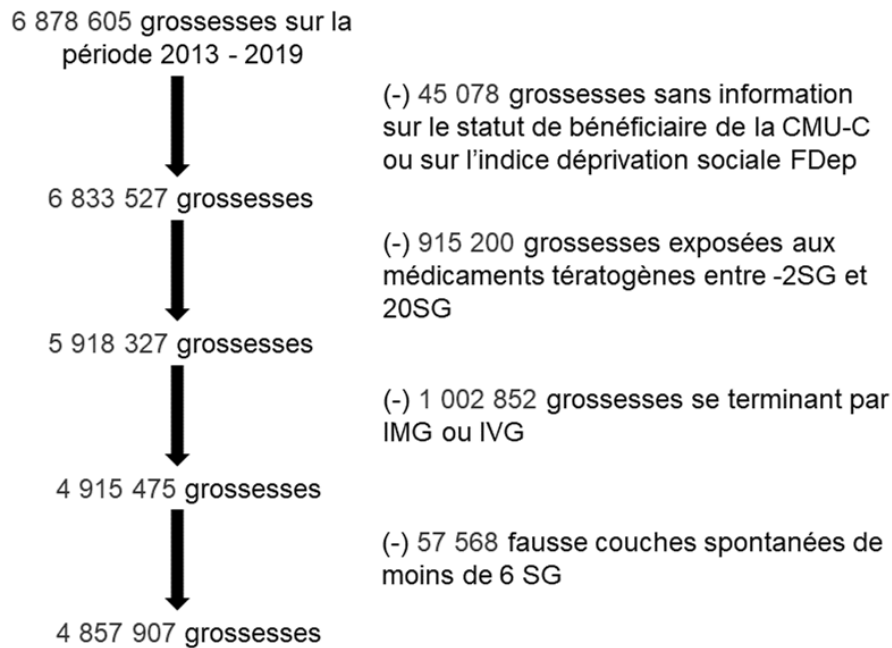


Figure 17. Flow-chart : sélection des grossesses pour les analyses portant sur la FCS

Le Tableau 9 montre la distribution des variables incluses dans le modèle de Cox avec variable dépendant du temps dans les deux groupes. Globalement, la proportion de femmes qui avaient eu une FCS antérieurement était plus élevée chez les femmes exposées que chez les femmes non exposées (5,41 % contre 3,21 %). Le pourcentage de femmes bénéficiant de la CMU-C était plus important chez les femmes exposées aux AINS (18,10 % contre 11,33 %). Les femmes exposées aux AINS au cours de la grossesse étaient exposées plus fréquemment aux AINS l'année précédant la conception (67,12 % contre 42,24 %), avaient plus d'antécédents de FCS et souffraient plus souvent d'autres maladies que les femmes non exposées.

Variable		Femmes exposées N = 349 294 (7,19 %)	Femmes non exposées N = 4 508 613 (92,81 %)
Variables continues*			
Nombre de grossesses depuis 2010		2 (1 ; 2)	2 (1 ; 2)
FDep		3 (2 ; 4)	3 (2 ; 4)
Variables catégorielles			
FCS		5,41 %	3,21 %
Tranche d'âge	<= 20 ans	5,60 %	4,34 %
	21 – 25 ans	18,62 %	17,58 %
	26 – 35 ans	59,90 %	64,37%
	36 – 45 ans	15,79 %	13,65 %
	> 45 ans	0,09 %	0,06 %
CMU-C		18,10 %	11,33%
Antécédents de FCS ou d'avortements		9,90 %	9,08 %
Exposition aux AINS un an avant la date de conception		67,12 %	42,24 %
Diabète		0,81 %	0,53 %
Maladies inflammatoires		0,99 %	0,83 %
Hypothyroïdie		3,05 %	2,53 %
Etat d'hypercoagulabilité		2,49 %	1,82 %

Tableau 9. Caractéristiques des grossesses exposées et non exposées aux AINS

* Médiane (écart d'interquartile)

Le Tableau 10 représente le hazard ratio (HR) de FCS à partir de la 6^e semaine de grossesse de tous les AINS confondus et celui de chaque molécule, classée par ordre décroissant de prévalence.

Médicament	Code d'ATC	3 jours de latence		
		Prévalence	HR	IC95%
AINS	M01A	349 294 (7,19 %)	1.83	1,80-1.86
Analyses par molécules				
Ibuprofène	M01AE01	148919 (3,20 %)	1,89	1,84-1,93
Kétoprofène	M01AE03	57126 (1,25 %)	2,25	2,18-2,32
Flurbiprofène	M01AE09	27021 (0,59 %)	3,27	3,14-3,40
Acide tiaprofénique	M01AE11	22473 (0,49 %)	0,90	0,83-0,98
Diclofénac	M01AB05 ou M01AB55	21538 (0,48 %)	1,05	0,98-1,14
Naproxène	M01AE02	19721 (0,43 %)	1,09	1,01-1,17
Piroxicam	M01AC01	5688 (0,12 %)	0,97	0,83-1,12
Acide niflumique	M01AX02	5598 (0,12 %)	0,92	0,78-1,08
Acide méfénamique	M01AG01	2984 (0,06 %)	2,43	2,11-2,79
Nabumétone	M01AX01	2282 (0,05 %)	1,09	0,87-1,35
Acélofénac	M01AB16	2294 (0,05 %)	0,98	0,77-1,23

Célécoxib	M01AH01	1596 (0,03 %)	1,26	0,99-1,59
Morniflumate	M01AX22	1512 (0,03 %)	1,68	1,25-2,23
Méloxicam	M01AC06	860 (0,01 %)	1,11	0,78-1,58
Etodolac	M01AB08	922 (0,02 %)	1,06	0,73-1,51
Ténoxycam	M01AC02	779 (0,01 %)	0,87	0,56-1,33
Insaponifiable d'huile d'avocat-soja	M01AX26	778 (0,01 %)	1,21	0,81-1,79
Etoricoxib	M01AH05	565 (0,01 %)	0,89	0,56-1,41
Indométacine	M01AB01	608 (0,01 %)	1,93	1,32-2,81

Tableau 10. Résultats de l'analyse intégrant trois jours de latence avec l'ajustement sur 10 covariables choisie selon la littérature : l'âge au début de grossesse en tranches d'âges de 5 ans, l'indicateur de CMU-C, l'indice FDep, le nombre de grossesses depuis 2010, l'historique des avortements ou des FCS depuis 2010, et la présence des facteurs suivants pendant une année avant la date de conception : la prise d'AINS, le diabète, les maladies inflammatoires, l'hypothyroïdie, ou l'état d'hypercoagulabilité.

L'exposition aux AINS était associée à une augmentation du risque de FCS survenant à partir de la 6^e semaine de grossesse (HR = 1,83 ; IC95% : 1,80-1,86). L'effet était hétérogène entre les molécules. Les molécules associées à un risque accru de FCS sont les suivantes : Flurbiprofène (HR = 3,27 ; IC95 % : 3,14-3,40), acide méfénamique (HR = 2,43 ; IC95 % : 2,11-2,79), kétoprofène (HR = 2,25 ; IC95 % : 2,18-2,32), indométacine (HR = 1,93 ; IC95 % : 1,32-2,81), ibuprofène (HR = 1,89 ; IC95 % : 1,84-1,93), morniflumate (HR = 1,68 ; IC95 % : 1,25-2,23), naproxène (HR = 1,09 ; IC95 % : 1,01-1,17).

3.2.2.5 Résultats des analyses de sensibilité

Pour évaluer la robustesse des résultats, différentes analyses de sensibilité ont été effectuées avec : (1) l'inclusion des FCS d'une durée inférieure à 6 semaines ; (2) Différentes durées pour la période de latence. Nous avons aussi calculé l'E-value pour évaluer la robustesse des résultats de l'analyse principale face à la présence de FDC non mesurés. Les résultats sont présentés dans le Tableau 11.

	HR	IC95%
Analyse avec les FCS \geq 6 semaines de grossesse et 3 jours de latence	1,83 (E-value = 3,06)	1,80-1,86 (E-value = 3,00-3,12)
<i>Analyse de sensibilité</i>		
Inclusion de toutes les FCS (3 jours de latence)	1,82	1,79-1,84
	0 jour	2,25 2,21-2,30
	4 jours	1,76 1,73-1,78
Période de latence	5 jours	1,70 1,67-1,72
	6 jours	1,63 1,60-1,66
	7 jours	1,56 1,53-1,59

Tableau 11. Résultats des analyses de sensibilité

Dans cette étude, nous avons exclu les FCS de moins de 6 semaines de grossesse, car ces FCS très précoces sont souvent mal enregistrées dans les bases de données médico-administratives. Ce choix n'a pas souvent été appliqué dans des études antérieures. De ce fait, nous avons mis en œuvre une analyse incluant toutes les FCS afin d'évaluer l'impact de ce choix sur les résultats. Comme indiqué dans le Tableau 11 l'inclusion de toutes les FCS n'a pas changé le HR par rapport à l'analyse principale (HR = 1,82 ; IC95 % : 1,79-1,84 contre HR = 1,83 ; IC95 % : 1,80-1,86).

Concernant la sensibilité à la durée de la période de latence, comme attendu, une augmentation de la durée de la période de latence diminuait le HR, mais le risque restait toujours significatif jusqu'à sept jours de latence (HR = 1,56 ; IC95 % : 1,53-1,59).

L'E-value de l'estimation ponctuelle et la borne inférieure de l'IC95 % étaient de 3,06 et 3,00, respectivement.

3.2.3 Analyse enrichie par une sélection automatisée des covariables d'ajustement

3.2.3.1 Score de propension en grande dimension

Dans notre étude, nous avons estimé le hdPS à partir de codes enregistrés pendant l'année précédant la date de conception concernant diverses dimensions : 18 codes liés aux actes biologiques, 56 codes liés aux consultations chez les généralistes et les spécialistes, 55 codes liés aux pathologies survenant avant la conception et 518 codes liés aux consommations médicamenteuses. Nous avons exclu 210 codes possédant une prévalence inférieure à 1/10 000. Dans chaque dimension, les codes ont été classés par ordre décroissant de prévalence, et les 250 premiers ont été retenus pour le calcul des fréquences et leur recodage en trois variables (une fois, sporadique, et fréquent, voir la Figure 4, section 1.1.6, page 42). Pour les dimensions contenant moins de 250 codes, tous les codes ont été conservés. Au total, 557 variables ont été générées et leur potentiel de confusion a été évalué à l'aide de la formule de Bross (voir Équation 3, section 1.1.6, page 41). Les 200 variables présentant le plus grand potentiel de confusion ainsi que les variables prédéfinies (voir la section 3.2.2.1) ont été incluses dans un modèle de régression logistique classique pour estimer le hdPS. La Figure 18 présente un diagramme de flux de la sélection de variables pour le hdPS.

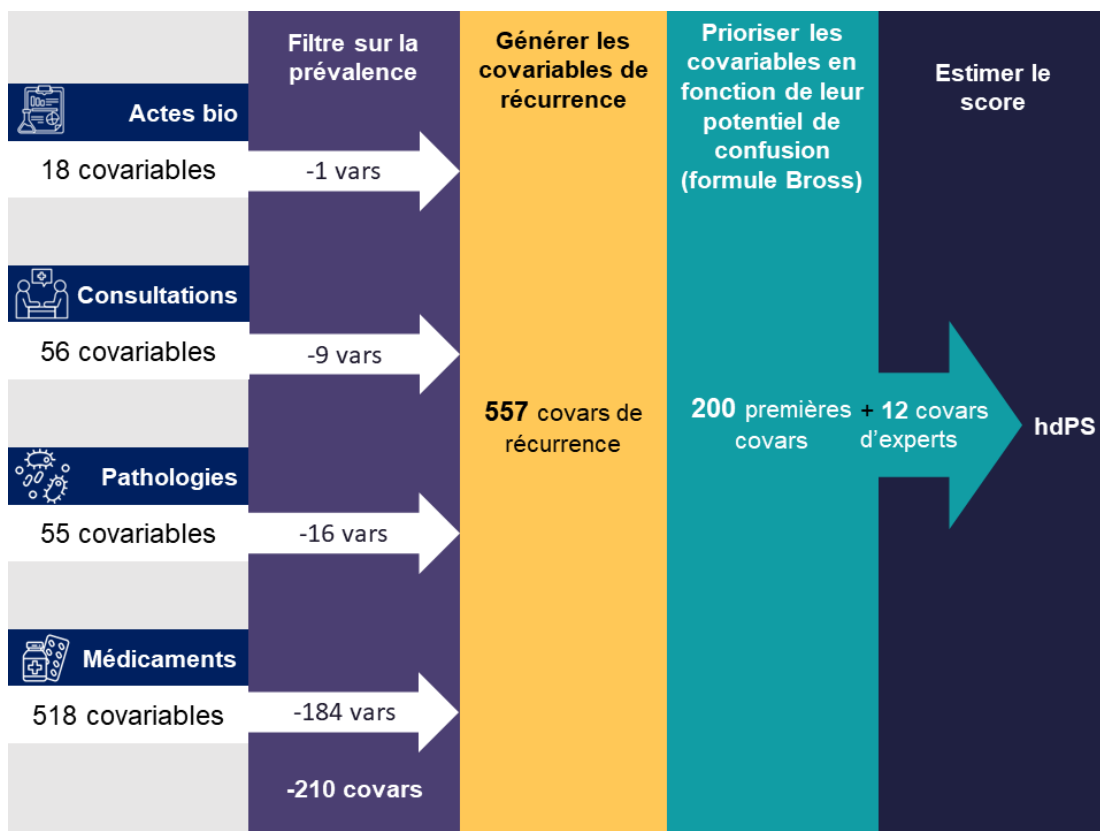


Figure 18. Étapes pour la sélection des variables incluses dans le hdPS

Nous avons estimé plusieurs hdPS au cours du suivi de la cohorte pour prendre en compte les variations temporelles de l'exposition. Au début de chaque semaine de grossesse, le hdPS a été réestimé chez les femmes n'ayant pas observé de FCS (ensemble à risque). Afin de réduire le coût computationnel, nous avons sélectionné une seule fois les 200 covariables pour les modèles du hdPS de toutes les semaines de grossesse. Cette sélection initiale s'appuyait sur la formule de Bross prenant en compte l'état d'exposition et d'événement à la fin du suivi. La Figure 19 présente l'intégration des 200 covariables dans la table des données et l'estimation du hdPS dépendant du temps.

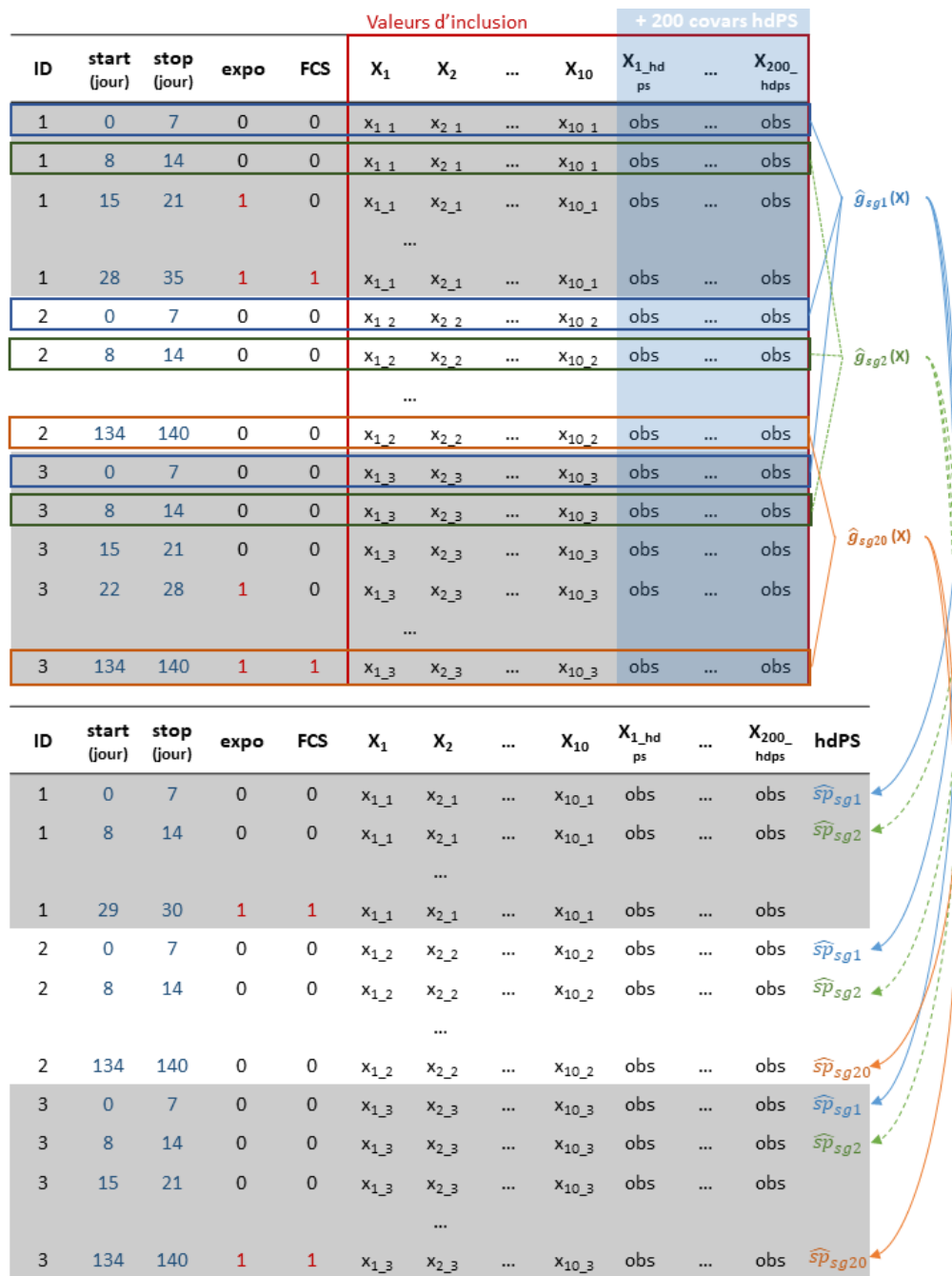


Figure 19. Estimation des hdPS dépendants du temps. Pour chaque semaine de grossesse i , un modèle logistique classique, $\widehat{g}_{sg_i}(X)$, a été utilisé pour prédire les SP des individus encore à risque à partir des 10 covariables d'expert et des 200 covariables hdPS, qui étaient constantes au fil du temps. Dans cet exemple, la femme n°1 participera aux constructions des hdPS pour les semaines 1 à 5 (FCS à la 5^e semaine) tandis que la femme n°2 participera à l'ensemble des hdPS (pas fin de suivi à la 20^e semaine)

Le score de propension des individus au début de chaque semaine de grossesse a été estimé à partir de des valeurs d'inclusion des 200 covariables hdPS et 10 covariables d'expert chez les femmes restant à risque. Une femme ayant eu sa première délivrance d'AINS entre la semaine $t - 1$ et t , après avoir ajouté trois jours de latence, était considérée comme non exposée de la conception jusqu'au début de la semaine t , puis comme exposée à partir de la semaine t jusqu'à la fin de son suivi. Bien que la réestimation du score à chaque semaine n'implique pas de variables dépendant du temps, elle demeure utile pour maintenir l'équilibre entre les groupes et lutter contre la déplétion différentielle des susceptibles dans les deux groupes (Wyss et al. 2020). Cette dernière peut survenir lorsque, en plus du traitement, il existe d'autres facteurs qui augmentent également le risque d'événement. Les individus présentant ces facteurs dans le groupe traité pourraient présenter l'événement et quitter l'étude plus rapidement que ceux dans le groupe non traité, entraînant un déséquilibre entre les deux groupes au fil du temps.

Finalement, le score a été intégré dans l'analyse par ajustement ou pondération IPTW dans un modèle de Cox avec variable dépendant du temps en prenant une période de latence de 3 jours pour la mesure de l'exposition. Dans les sections suivantes, afin de distinguer les analyses intégrant le hdPS de celle reposant uniquement sur les covariables choisies *a priori* selon la littérature (section 3.2.2), nous désignerons cette dernière par le terme « analyse classique », qui fait référence aux analyses classiquement mises en œuvre en pharmacoépidémiologie, sans intégrer des méthodes exploitant les données en grande dimension.

3.2.3.2 Résultats

Dans notre analyse, un chevauchement adéquat des SP a été observé entre les femmes exposées et non exposées, à chaque semaine de grossesse. La distribution des scores des femmes au début de chaque semaine de grossesse est présentée dans l'Annexe B.3 : *Distribution des hdPS au début de chaque semaine de grossesse (SG)*.

Le Tableau 12 représente les résultats des analyses.

Méthode	HR	IC95%
Analyse classique	1,83	1,80-1,86
<i>Analyse avec le hdPS</i>		
Ajustement	1,86	1,83-1,89
IPTW	2,24	2,20-2,28

Tableau 12. Résultats des analyses intégrant le hdPS au modèle de Cox avec variable dépendant du temps et une période de latence de trois jours. Les résultats de l'analyse classique (voir la section 3.2.2, page 89).

L'HR obtenu par l'ajustement sur le hdPS était similaire à celui de l'analyse sans hdPS (HR = 1,86 ; IC95% : 1,82-1,89 contre HR = 1,83 ; IC95% : 1,80-1,86). En revanche, l'IPTW a généré un HR plus élevé que celui de l'analyse sans hdPS (HR = 2,24 ; IC95% : 2,20-2,28 contre HR = 1,83 ; IC95% : 1,80-1,86).

3.3 Discussion

Les AINS sont des médicaments utilisés chez les femmes enceintes, même si les recommandations restreignent largement leur usage. Généralement, leur

utilisation est déconseillée à partir de la 20^e semaine de grossesse ou le 4^e mois de grossesse en France et est contre-indiqué au troisième trimestre ou le 6^e mois de grossesse en France. Au premier trimestre, ils sont utilisés avec précaution en raison de l'association débattue de ces médicaments à certaines conséquences sur le fœtus, dont la FCS. En effet, certaines études ont trouvé un surrisque de FCS chez les femmes exposées aux AINS au cours de la grossesse. Cependant, les conclusions sur ce lien restent inconsistantes à ce jour (Ying et al. 2022). L'objectif de ce travail était d'évaluer l'association entre les AINS et la FCS en utilisant une grande base de données médico-administrative concernant les grossesses sur la période de 2013-2019 en France, et en prenant en compte par modélisation et par plan d'analyse différents biais affectant l'estimation du risque causal. Nous avons estimé le HR à l'aide d'un modèle de régression de Cox avec variable dépendant du temps en intégrant la période de latence dans la définition des personnes exposées afin d'éviter les biais de temps immortel et protopathique. Nous avons aussi implémenté un hdPS dépendant du temps dans l'analyse pour évaluer les mérites de l'utilisation d'une méthode exploitant les données en grande dimension dans la prise en compte des facteurs de confusion non mesurés.

Dans notre étude, nous avons constaté que l'exposition aux AINS entre les deux semaines précédant la conception et la 20^e semaine de grossesse augmentait le risque de FCS à partir de la 6^e semaine de grossesse (HR = 1,83 ; IC95 % : 1,80-1,86). Bien que notre conclusion sur l'existence d'un risque soit en accord avec certaines études publiées. (D.-K. Li et al. 2018; Nakhai-Pour et al. 2011; Nielsen et al. 2001), elle diffère de celles d'autres études. (Daniel et al. 2015b; Dathe et al. 2018; Edwards et al. 2012).

Une première explication réside dans l'ensemble des médicaments AINS étudié dans chaque étude. A l'instar d'autres études, notre travail a mis en évidence une hétérogénéité dans l'effet de chaque molécule sur le risque de FCS. Ainsi, l'effet de l'exposition générale aux AINS reflète l'effet moyen de toutes les molécules composant cette classe, et peut donc varier si les médicaments concernés sont différents entre les études. Ainsi, tandis que l'étude de Daniel et al. (2015b), a évalué sept AINS : l'ibuprofène, le diclofénac, le naproxène, l'étodolac, l'indométacine, le lornoxicam, et la nabumétone, Dathe et al. s'intéressaient à l'exposition à l'ibuprofène. Notre étude a porté sur 25 médicaments. Au niveau des molécules, nos résultats étaient cohérents en terme de significativité avec l'étude de Daniel et al.(2015b) pour l'étodolac (HR = 1,06 ; IC95 % = 0,73-1,51 contre HR = 1,28 ; IC95 % : 0,91-1,79, respectivement), le diclofénac (HR = 1,05 ; IC95 % : 0,98-1,14 contre HR = 1,16 ; IC95 % : 0,94-1,43, respectivement), et l'indométacine (HR = 1,93 ; IC95 % : 1,32-2,81 contre HR = 2,64 ; IC95 % : 1,56-4,47, respectivement). C'était aussi le cas pour le célécoxib et l'étoricoxib qui font partie de la famille des inhibiteurs de la cyclo-oxygénase 2 (COX-2), un sous-groupe d'AINS contre-indiqués pendant toute la grossesse en France et pour lesquels nous n'avons pas observé un surrisque significatif de FCS. En ce qui concerne l'ibuprofène, nous avons trouvé un risque de FCS significatif (HR = 1,89 ; IC95 % : 1,84-1,93), risque qui n'était pas mis en évidence dans l'étude de Dathe et al. (2018).

La différence dans la définition de la période d'exposition peut être une autre raison expliquant des différences dans les résultats observés. Dans notre étude, une femme était considérée exposée à partir de la date de sa première délivrance d'AINS, avec un décalage de trois jours pour prendre en compte la période de latence, jusqu'à la fin de son suivi. Dathe et al. (2018) se sont

appuyés, quant à eux, sur la durée des délivrances pour déterminer les périodes exposées qui ont majoritairement duré moins de 7 jours. Par rapport à notre définition, celle de Dathe et al. tend donc à réduire le nombre de FCS exposées et à les assigner au groupe non exposé, car les périodes exposées sont plus courtes.

L'inconsistance des résultats peut également provenir de différences dans la fenêtre d'exposition aux AINS. Par exemple, tandis que la fenêtre d'exposition de notre étude s'étend des 2 semaines avant la conception jusqu'à la 20^e semaine de grossesse, celle de l'étude de Dathe et al. va de la 2^e semaine de grossesse jusqu'à la 12^e semaine de grossesse + 6 jours, et celle de l'étude d'Edward du premier jour des dernières règles à la 6^e semaine de grossesse (Dathe et al. 2018; Edwards et al. 2012).

Comme constaté par Li et al., le moment de l'exposition peut également influencer le risque de FCS avec un effet plus fort pour les expositions autour de la conception (D.-K. Li, Liu, et Odouli 2003a; D.-K. Li et al. 2018). Ainsi, la différence dans la fenêtre d'exposition peut être l'une des raisons qui expliquent la disparité entre les études.

Nous avons effectué des analyses de sensibilité pour évaluer la robustesse de nos résultats à la durée de la période de latence ou lorsqu'on inclut toutes les FCS dans l'analyse. Tandis que l'inclusion de toutes les FCS n'a pas changé les résultats, l'augmentation de la durée de la période de latence a réduit le risque de FCS. Cependant, le risque demeurerait significatif dans les analyses avec une période de latence de 7 jours. Les E-values calculées pour l'estimation ponctuelle et la borne inférieure de l'IC95% ont indiqué une forte robustesse des résultats face à la présence de FDC non mesurés. En effet, l'E-value de

l'estimation ponctuelle et la borne inférieure de l'IC95 % de l'analyse classique étaient respectivement de 3,06 et 3,00. Cela signifie qu'afin de faire disparaître l'association observée entre les AINS et la FCS, il faut qu'un FDC non mesuré ait simultanément une force d'association d'environ 3 avec les AINS et la FCS conditionnellement aux covariables ajustées dans l'analyse. Bien que cela dépasse le cadre de ce travail, d'autres analyses de sensibilité, telles que des analyses quantitatives des biais (Lash et al. 2014), pourraient être envisagées pour évaluer plus en profondeur la robustesse des résultats face à la présence de FDC non mesurés, comme l'ont suggéré VanderWeele et Mathur (2020).

Dans le cas d'une étude où la durée de la période de latence est laissée au choix de l'investigateur, notamment lorsque la durée pertinente n'est pas connue, un allongement trop important de la période de latence peut conduire à conclure à une association non significative. Tamim et al. (2007) ont proposé deux méthodes pilotées par les données visant à choisir la durée de la période de latence. Cependant, ces méthodes reposent sur l'hypothèse que les mesures d'effet présentent une tendance monotone vers le plateau lorsque la période de latence continue d'être prolongée, ce qui n'est pas nécessairement valide en réalité (Arfè et Corrao 2016). Une autre possibilité est d'évaluer différentes périodes de latence et de rapporter de manière transparente leur impact sur les résultats. Dans certains cas, les informations disponibles sur le mécanisme de l'effet de l'exposition, ou sur le délai depuis les manifestations précoces de la maladie et son diagnostic, etc. peuvent aider à choisir la période de latence. Dans notre étude, reporter toutes les expositions à 3 jours et assigner les FCS qui surviennent pendant la période de latence au groupe non exposé semble supposer que tous les effets potentiels des AINS sur la FCS ne peuvent pas se déclencher pendant les 3 jours suivant l'exposition, et que les prises d'AINS

pendant cette période sont susceptibles d'être liées aux symptômes précoces d'une FCS débutante. Cette hypothèse semble plausible au regard du mécanisme d'action des AINS, qui interfèrent avec l'implantation du fœtus liée via l'inhibition de la synthèse des prostaglandines dans l'utérus. Le choix d'une période de trois jours s'est également aligné sur notre stratégie d'étude, consistant à postuler une valeur initiale en suivant l'article de Daniel et al. (2015b), puis prolonger cette période dans une analyse de sensibilité afin d'évaluer la robustesse des résultats pour différentes durées de latence. Ainsi, le risque demeurait significatif même avec une semaine de latence (HR = 1,56 ; IC95% : 1,53-1,59).

L'intégration du hdPS par l'ajustement a donné un HR similaire à celui de l'analyse sans score (HR = 1,86 ; IC95 % : 1,83-1,89 contre HR = 1,83 ; IC95% : 1,80-1,86, respectivement). Cette similarité peut être expliquée par le fait que les covariables souvent proposées dans la littérature captent la plupart de la confusion mesurée ou indirectement mesurée, et l'ajout des covariables identifiées par l'algorithme d'hdPS n'apporte pas des changements significatifs dans les estimations par rapport au modèle avec une sélection « experte » voire ne les améliorent pas toujours comme suggéré par certains auteurs (Schneeweiss et Rassen 2011; Toh, García Rodríguez, et Hernán 2011). Par ailleurs, une étude de simulation a montré que l'ajustement par le SP pourrait induire un biais vers la valeur nulle pour l'estimation de l'odds ratio ou le hazard ratio. Ce biais s'expliquerait par la différence dans l'objectif d'estimation de la régression et le SP. Alors que le modèle de régression cible un effet conditionnel, le SP est utilisé plutôt pour estimer l'effet marginal (Austin et al. 2007). La présence potentielle de la sous-estimation du HR notamment quand l'effet était assez important a également été démontré par une autre étude de

simulation (Austin 2013). Cependant, même si ce biais est présent dans notre analyse, il ne remet pas en question notre conclusion sur la présence d'un risque accru de FCS lié aux AINS car le vrai risque serait potentiellement plus élevé.

L'utilisation de l'IPTW a conduit à un HR plus élevé (HR = 2,24 ; IC95 % : 2,20-2,28). Cette différence observée peut s'expliquer par des estimands théoriques ciblés différents entre les approches. Alors que l'IPTW cible l'effet marginal des AINS, les autres méthodes ciblent un effet conditionnel. L'ajustement mesure l'effet des AINS au sein des groupes de femmes dont le SP est similaire. De même, l'analyse classique mesure l'effet des AINS au sein de chaque groupe dont les valeurs des covariables incluses dans le modèle de Cox sont identiques. Il faudrait mettre en œuvre d'autres méthodes qui ciblent le même estimant théorique que l'IPTW pour pouvoir les comparer par exemple, la GC. En raison du temps limité, et du coût computationnel, nous ne pouvions pas effectuer la GC dans le cadre de ce travail. L'apport de cette méthode sera étudié dans un travail à venir. Par ailleurs, la discordance des résultats entre les deux approches basées sur le SP pourrait s'expliquer par l'hypothèse supplémentaire sur la relation linéaire entre le SP et l'événement en utilisant le score comme variable dans le modèle de régression. La violation de cette hypothèse peut biaiser l'estimation obtenue par l'ajustement.

A la différence de l'étude de Li et al., nous n'avons pas utilisé le paracétamol en tant que témoin d'exposition négatif (D.-K. Li, Liu, et Odouli 2003a; D.-K. Li et al. 2018). Cet emploi se base sur la similarité supposée entre ses indications et celles des AINS. Un risque de FCS non significatif pour le paracétamol

obtenu par les analyses appliquées aux AINS peut indiquer une absence de biais résiduel dans les résultats concernant les AINS et la FCS. Cependant, comme ce qui est indiqué dans la section 1.1.5.1 (page 34), pour qu'un témoin négatif soit valide, il doit être le plus « U-comparable » possible avec l'exposition d'intérêt (pour un témoin d'exposition) ou l'événement d'intérêt (pour un témoin d'événement). Autrement dit, pour être utilisé comme témoin négatif, il faut que l'ensemble des FDC de l'association paracétamol-FCS soit le plus similaire possible à celui de l'association AINS-FCS. Néanmoins, malgré des indications semblables, les caractéristiques des individus qui reçoivent ces médicaments ne sont pas forcément similaires. Une étude a montré que les individus qui ont reçu le paracétamol avaient tendance à souffrir plus de comorbidités, telles que l'infarctus du myocarde, l'hémorragie gastro-intestinale, des maladies rénales et l'accident vasculaire cérébral, par rapport aux individus qui ont reçu l'ibuprofène (Weinstein et al. 2017). Dans le cas des FCS, si les femmes ayant une grossesse compliquée présentent une probabilité plus élevée de recevoir le paracétamol, il sera possible que l'ensemble de FDC du lien paracétamol-FCS soit assez différent de celui du lien AINS-FCS. Ainsi, une association significative entre le paracétamol et la FCS ne signifie pas forcément l'existence de biais résiduels dans l'association AINS-FCS ; inversement, une association nulle ne signifie pas nécessairement l'absence de biais dans les résultats de l'AINS. Dans nos données, nous avons remarqué que la manière de prescrire les deux médicaments n'était pas similaire. En effet, 75 % des femmes exposées aux AINS ont reçu leur première délivrance avant le premier mois de grossesse, tandis que plus des trois quarts des femmes exposées au paracétamol ont reçu leur première délivrance après ce moment. Cela remet donc en cause la validité de l'utilisation du paracétamol comme témoin négatif. La distribution des temps d'exposition aux deux médicaments

est présentée dans l'*Annexe B.4 : Distribution des délais de première prescription de paracétamol et d'AINS après le début de grossesse*.

Comme toute étude utilisant les données du SNDS, notre étude est susceptible d'erreurs de classification des expositions, dans le cas de délivrances non prises, ou de médicaments délivrés sans ordonnance, même si pour cette dernière situation, la sous-estimation de l'exposition semble limitée chez les femmes enceintes (Bérard et al. 2019). Il existe aussi des limites spécifiques à notre étude. En intégrant la période de latence, nous avons risqué de sous-estimer l'effet des AINS en raison de l'omission d'un éventuel effet instantané. Néanmoins, cela signifie qu'en présence d'un tel effet, le risque de FCS serait potentiellement encore plus élevé que celui trouvé dans notre étude. Une autre limite est le manque de variables dépendant du temps concernant les caractéristiques des femmes. En effet, à la différence des études de cohorte dans lesquelles les informations des participants sont mises à jour au cours du suivi, les variables dans les bases de données médico-administratives ne sont pas mesurées régulièrement. Ainsi, il est possible que la vraie valeur des facteurs qui changent la probabilité d'être exposé ne soit pas prise en compte dans le modèle du score de chaque semaine de grossesse, ce qui peut aboutir à de moindres performances du hdPS dans notre analyse.

3.4 Conclusion

Dans notre étude, l'exposition aux AINS a été associée à un surrisque de FCS à partir de la 6^e semaine de grossesse. Cette association est restée significative à travers différentes analyses de sensibilité ou avec l'incorporation du hdPS. Nos résultats sont cohérents avec ceux d'autres études et confirment la nécessité de prudence lors de l'utilisation des AINS au premier trimestre de grossesse.

CHAPITRE 4

Conclusion générale et perspectives

4.1 Conclusion générale

Dans le cadre de cette thèse, nous avons évalué et mis en œuvre des méthodes utilisées dans le domaine de la causalité sur les données du SNDS, une très large base de données médico-administratives, dans le cadre d'études en pharmacoépidémiologie. Ces méthodes ont été étudiées dans un contexte lié à l'effet des consommations médicamenteuses au cours de la grossesse sur le fœtus et le nouveau-né. Dans le volet méthodologique présenté au chapitre 2, nous avons évalué les performances de trois méthodes pour atténuer le biais causé par les FDC : le score de propension en grande dimension, la G-computation, et l'estimation du maximum de vraisemblance ciblée (TMLE), appliquées aux données en grande dimension. Le chapitre 3 s'est intéressé à la mise en œuvre de l'une de ces approches dans le cadre d'une association d'intérêt, impliquant un plan d'étude plus spécifique et une modélisation plus complexe pour pallier d'autres biais. Il intègre des informations issues des

données en grande dimension grâce à l'utilisation du score de propension en grande dimension.

Dans le chapitre 2, il convient de noter que, parmi les trois méthodes étudiées, le hdPS est la seule spécifiquement conçue pour l'exploitation des données en grande dimension. Les autres méthodes ont été adaptées au contexte de la grande dimension grâce à l'intégration de la régression LASSO, une méthode issue de l'apprentissage statistique, utilisée pour ajuster des modèles à partir d'un grand nombre de prédicteurs. Au travers d'une étude empirique à partir d'un ensemble de référence de grande taille, notre travail complète les résultats issus des études de simulation (un autre type d'étude empirique) en fournissant un aperçu des performances de ces méthodes dans des situations fréquemment rencontrées dans des études en pharmacoépidémiologie utilisant les grandes bases de données de santé, telles que les expositions rares ou le volume énorme des données. Toutes les méthodes étudiées ont contribué à améliorer le contrôle de la confusion causée par les FDC mesurés ou indirectement mesurés. Parmi elles, la G-computation était associée à de meilleures performances en termes de proportion de vrais positifs et de faux négatifs générés. Lorsque l'exposition était rare, l'utilisation du SP, ainsi que les méthodes qui en dépendent, comme les estimateurs doublement robustes, pouvait nuire à la qualité des estimations en raison de la présence de scores extrêmes ou de la violation de l'hypothèse de positivité.

Dans le chapitre 3, nous avons implémenté l'une des méthodes étudiées dans le chapitre 2, le hdPS, dans la cadre d'une étude de pharmacoépidémiologie portant sur le lien entre les expositions aux anti-inflammatoires non stéroïdiens et le risque de fausse couche spontanée. Dans cette étude, outre la question

concernant le biais de confusion, il existe potentiellement d'autres biais, tels que le biais de temps immortel et le biais protopathique qui nécessitent des approches plus complexes pour les atténuer comme le modèle de Cox avec variable dépendant du temps et l'intégration d'une période de latence à la définition de l'exposition. Nous avons comparé les analyses avec et sans hdPS pour évaluer l'impact de l'intégration du score sur les résultats. Dans ce travail, l'exposition aux AINS au cours de la grossesse était associée à un risque accru de FCS. Cette association demeurait significative dans les différentes analyses de sensibilité. L'intégration du hdPS n'a apporté qu'un changement mineur avec l'ajustement, mais a augmenté l'estimation du risque avec l'IPTW. Cette différence peut s'expliquer de plusieurs manières, telles que l'hypothèse supplémentaire de l'ajustement sur la relation linéaire entre le risque de FCS et le SP, ou les estimands théoriques ciblés par des méthodes (marginal ou conditionnel). Malgré la différence entre les résultats de l'IPTW et ceux des autres analyses, cela ne réfute pas notre conclusion sur la présence d'un lien entre l'exposition aux AINS et la FCS. Cette conclusion est en accord avec plusieurs études (Nielsen et al. 2001; Nakhai-Pour et al. 2011; D.-K. Li et al. 2018).

Bien que le hdPS ait permis d'atténuer le biais lié aux FDC dans le chapitre 2, sa contribution était moindre par rapport aux analyses s'appuyant uniquement sur des variables choisies *a priori* dans le chapitre 3. Cette différence peut provenir des différents contextes auxquelles la méthode a été appliquée. Le chapitre 2 impliquait un plan d'étude simple avec une exposition binaire, non dépendante du temps et un événement binaire, tandis que le chapitre 3 s'intéressait à une situation plus complexe avec une exposition dépendante du temps et une analyse de survie. La différence peut également s'expliquer par

le fait que, dans l'étude présentée au chapitre 3, les covariables sélectionnées par des experts capturent déjà une grande partie de la confusion qui biaise l'association d'intérêt, y compris celle liée aux facteurs non mesurés, comme l'ont suggéré par Schneeweiss et al. (2011).

Au travers de cette thèse, nous avons été confrontés à la faisabilité de l'application des méthodes statistiques étudiées en pratique. Ces méthodes ont été souvent développées et évaluées sur des données simulées ou des simulations plasmodes, qui ciblent un type de biais précis ou une question particulière. En réalité, les investigateurs doivent traiter simultanément différents types de biais, ce qui soulève deux points majeurs à considérer lors de l'application des méthodes : (1) le coût computationnel et (2) la complexité de la modélisation et du plan d'étude. Bien que ces défis concernent toutes les trois méthodes étudiées, nous avons notamment rencontré des difficultés dans la mise en œuvre de la GC, la méthode la plus performante dans le chapitre 2, dans le cadre de l'étude réelle menée dans le chapitre 3.

En effet, concernant le coût computationnel, et selon l'algorithme présenté par Keil et al. (2014), l'utilisation de la GC pour estimer un HR consiste à construire un modèle de prédiction pour chaque variable dépendante du temps et à prédire leur valeur pour chaque jour de suivi et pour chaque individu. Appliqué à notre contexte, cela nous aurait donc imposé de prédire le statut d'événement pour chaque jour de suivi des 4 millions de femmes, générant ainsi un coût computationnel considérable. Ce coût aurait été encore amplifié par l'utilisation du bootstrap pour estimer la variance des estimations. Pour finir, l'extension de la GC aux données en grande dimension aurait impliqué d'intégrer plus de 200 variables dans les modèles de prédictions.

Concernant la complexité de la modélisation et du plan d'étude, il était, par exemple, difficile d'incorporer les périodes de latence dans la GC, qui consiste à fixer les expositions en fonction des interventions d'intérêt, par exemple, une population contrefactuelle toujours exposée versus jamais exposée.

Toutes ces difficultés soulignent l'intérêt de développer de nouvelles méthodes ou d'adapter celles existantes aux analyses conduites sur de grandes bases de données, notamment dans un contexte où les études pharmacoépidémiologiques s'appuyant sur de grandes bases de données médico-administratives sont de plus en plus courantes.

4.2 Perspectives

A partir des résultats de ces travaux, plusieurs pistes sont envisageables dans de futurs travaux. Au chapitre 2, nous avons relevé certains problèmes émergeant souvent dans les études en pharmacoépidémiologie, tels que l'amélioration de la sélection de variables et de l'estimation de l'effet de l'exposition, la présence d'expositions rares, et le coût computationnel élevé lors de l'application des méthodes aux données en grande dimension. Ces problèmes peuvent faire l'objet d'études à venir qui impliquent certaines approches proposées ces dernières années.

Lors du choix du modèle d'estimation, le compromis biais-variance est une question importante. Alors qu'un modèle complexe peut offrir une bonne qualité de prédiction au sein d'une base de données donnée et présenter un faible biais, il risque de posséder une qualité de prédiction médiocre lorsqu'il est appliqué à une autre base de données à cause de sa dépendance excessive à la base de données d'origine. L'estimateur « Super Learner (SL)» (van der Laan, Polley, et Hubbard 2007) permet de combiner différents modèles « simples » pour produire une estimation moyenne, optimisant ainsi le compromis biais-variance. Cette méthode peut être utilisée pour estimer le SP, les issues potentielles, et ainsi les estimations de la TMLE. Chaque estimation composant le « Super learner » est pondérée en fonction de sa qualité de prédiction, évaluée à l'aide de la validation croisée. Ces estimateurs peuvent être basés sur différents algorithmes et/ou sur un même algorithme avec différents ensembles de prédicteurs. On peut ainsi construire, par exemple, un hdPS qui s'appuie sur plusieurs modèles de régression classiques intégrant un nombre différent de covariables sélectionnées à partir de

différents seuils pour la formule de Bross. Le SL peut donc aider à répondre à la question sur le choix du nombre de variables et du type de modèle d'estimation pour la construction du hdPS (Wyss et al. 2018).

Concernant la sélection de variables pour le hdPS, certaines méthodes alternatives peuvent être envisagées. Parmi celles-ci, deux méthodes retiennent notre attention : l'« outcome-adaptive LASSO » (Shortreed et Ertefaie 2017) et la « Scalable collaborative TMLE (SC-TMLE) » (Ju et al. 2019).

L'OAL modélise le score de propension à l'aide d'une extension du LASSO, l'« adaptive LASSO » (Zou 2006), qui permet d'intégrer des pénalités spécifiques à chaque covariable du modèle. L'originalité du travail de Shortreed et al. est de proposer que ces pénalités tiennent compte de la force d'association entre chaque covariable et l'événement. En pratique ces poids sont déterminés au moyen d'un modèle de régression multiple préliminaire incluant l'événement comme variable à expliquer, et l'exposition ainsi que toutes les covariables comme prédicteurs. Les poids proposés sont définis comme $|\tilde{\beta}_j|^{-\gamma}$ avec $\gamma > 1$.

En prenant en compte la relation des covariables avec l'événement conditionnellement au traitement, l'OAL impose une contrainte plus forte sur les coefficients des covariables ayant une association faible ou nulle avec l'événement. L'application de cette méthode, seule ou combinée avec une présélection de variables à l'aide de l'algorithme du hdPS, peut faire l'objet d'études futures.

La SC-TMLE a été développée pour réduire le coût computationnel de la méthode TMLE collaborative gloutonne (« greedy C-TMLE »). Cette dernière

repose sur un processus itératif de sélection de modèles de SP parmi une séquence de modèles candidats. A chaque itération, une séquence de modèles SP est construite en ajoutant une nouvelle covariable, qui n'a pas encore été incluse dans les itérations précédentes. Ainsi, k itérations sont réalisées pour un ensemble de k variables. Chaque modèle SP produit un estimateur $TMLE_j$, et à chaque itération, celui qui minimise la fonction de perte par rapport aux prédictions de l'itération précédente est sélectionné comme candidat de cette itération. En fin de compte, la validation croisée est utilisée pour choisir le meilleur estimateur parmi les k estimateurs TMLE. Grâce à la capacité de considérer différents modèles SP basés sur divers ensembles de covariables, la C-TMLE gloutonne favorise le compromis biais-variance en choisissant un ensemble de variables suffisant pour minimiser le biais sans augmenter de manière excessive la variance. Cependant, la nécessité d'effectuer de nombreuses itérations et la validation croisée augmente énormément le coût computationnel et entrave l'application de la méthode aux grandes bases de données. La SC-TMLE diminue le coût computationnel en intégrant un ordre prédéfini des covariables à inclure dans chaque itération. Ainsi, un seul modèle SP est construit à chaque itération. Ju et al. (2019) ont proposé deux algorithmes guidés par les données pour déterminer l'ordre des covariables. Le premier algorithme ordonne les variables en fonction de l'amélioration de la qualité de prédiction des issues potentielles par rapport aux prédictions initiales de la GC, à l'aide d'un modèle logistique préliminaire. La qualité de prédiction est évaluée à l'aide d'une fonction de perte définie par les investigateurs. Le second algorithme ordonne les variables en fonction de leur corrélation avec la déviance entre les prédictions initiales de la GC et les réponses observées Y au sein de chaque strate du traitement A .

Dans un autre travail, Ju et al.(2019) ont développé la « Positivity-C-TMLE » qui consiste à construire une troncature des SP pilotée par les données, pour traiter la violation de l'hypothèse de la positivité et la présence de SP extrêmes. Cette méthode vise à déterminer le meilleur seuil de troncature parmi une séquence de seuils candidats (les quartiles des SP estimé) à l'aide de la validation croisée. Les auteurs ont démontré que cette méthode permet d'avoir un meilleur compromis entre le biais et la variance.

A partir des résultats du chapitre 3, plusieurs perspectives peuvent être envisagées visant à améliorer la pertinence des hypothèses concernant l'exposition aux AINS, telles que la détermination des épisodes d'exposition, le déclin de l'effet au fils du temps ou la relation dose-effet.

En ce qui concerne la construction des épisodes d'exposition, bien que nous ayons appliqué une définition couramment utilisée dans des études antérieures, qui considère qu'une femme est exposée dès sa première délivrance d'AINS jusqu'à la fin de son suivi, cette définition semble reposer sur une hypothèse relativement forte concernant la persistance de l'effet des AINS. Par exemple, l'hypothèse d'un lien éventuel entre une FCS à la vingtième semaine de grossesse et l'exposition aux AINS à la conception semble relativement forte. A notre connaissance, il n'existe actuellement aucune preuve permettant de valider ou de réfuter cette hypothèse. Nous suggérons donc qu'il serait utile de restreindre la durée de la période d'exposition et d'évaluer l'impact de cette modification sur les résultats. Les épisodes d'exposition peuvent être construits en tenant compte des informations sur la dose et la durée de la prescription, ou en appliquant une même période d'exposition après chaque délivrance, avec une variation de sa durée pour

examiner le changement des résultats. Il serait également possible d'introduire une période supplémentaire à la fin de l'épisode d'exposition pour tenir compte de la persistance de l'effet du médicament.

Une nouvelle piste de recherche s'ouvre avec la prise en compte de la relation dose-effet, de l'effet cumulatif et de l'évolution temporelle du risque de FCS lié à l'exposition aux AINS. L'une des approches permettant de répondre à ces questions est la méthode de l'exposition cumulative pondérée (WCE pour « weighted cumulative exposure ») (Abrahamowicz et al. 2006). La WCE pondère les périodes d'expositions en fonction de leur proximité temporelle avec l'événement. Cette méthode permet donc de prendre en compte le fait que le risque d'événement ne reste pas constant au cours du suivi après l'exposition, mais est généralement plus important immédiatement après celle-ci, et diminue progressivement au fil du temps. Les poids des périodes d'exposition sont déterminés en utilisant une fonction estimée à partir des données. Lors de l'analyse, la dose d'exposition à un moment u durant le suivi est calculée comme la somme cumulative des doses précédentes pondérées $w(u-t)X(t)$, ce qui peut être représenté par l'équation $X(u) = \sum_t^u w(u-t)X(t)$, où $w(u-t)$ est une fonction qui attribue un poids à la dose réelle au temps t , $X(t)$, en fonction du délai entre le temps t et le moment considéré u . En pratique, cette fonction est souvent estimée à l'aide de splines cubiques flexibles.

La conclusion sur la relation dose-effet reste à ce jour inconsistante (D.-K. Li, Liu, et Odouli 2003b; Nakhai-Pour et al. 2011). Cependant, la compréhension de cette relation pourrait contribuer à expliquer les divergences entre les études concernant le lien entre les AINS et la FCS. En effet, Edward et al. ont

suggéré que la différence de dosage pourrait expliquer la discordance entre des études antérieures et la leur, qui n'a pas trouvé d'association significative entre les AINS et la FCS (Edwards et al. 2012). Ils ont proposé que des doses d'AINS plus élevées que celles vendues sans ordonnance étudiées dans leur étude, pourraient induire un effet significatif sur le risque de FCS. Une analyse approfondie de la relation dose-effet pourrait donc apporter une compréhension supplémentaire sur la disparité entre les résultats des études publiées. Par ailleurs, étant donné qu'il est généralement recommandé d'utiliser les AINS à la dose efficace la plus faible possible et pendant la durée la plus courte possible au premier trimestre (ANSM 2024), les études portant sur la relation dose-effet spécifiquement chez les femmes enceintes peuvent aider les professionnels de santé à choisir une dose pertinente dans un contexte actuel où il manque encore d'informations concernant le changement des caractéristiques de pharmacocinétiques des AINS chez les femmes enceintes (Coppola et al. 2022).

REFERENCES

- Abrahamowicz, Michal, Gillian Bartlett, Robyn Tamblyn, et Roxane du Berger. 2006. « Modeling Cumulative Dose and Exposure Duration Provided Insights Regarding the Associations between Benzodiazepines and Injuries ». *Journal of Clinical Epidemiology* 59 (4): 393-403. <https://doi.org/10.1016/j.jclinepi.2005.01.021>.
- Acton, Emily K., Allison W. Willis, et Sean Hennessy. 2023. « Core concepts in pharmacoepidemiology: Key biases arising in pharmacoepidemiologic studies ». *Pharmacoepidemiology and drug safety* 32 (1): 9-18. <https://doi.org/10.1002/pds.5547>.
- ANSM. 2024. « Dossier thématique - Traitement de la douleur durant la grossesse ». ANSM. 10 octobre 2024. <https://ansm.sante.fr/dossiers-thematiques/medicaments-et-grossesse/traitement-de-la-douleur-durant-la-grossesse>.
- Arfè, Andrea, et Giovanni Corrao. 2016. « The Lag-Time Approach Improved Drug–Outcome Association Estimates in Presence of Protopathic Bias ». *Journal of Clinical Epidemiology* 78 (octobre):101-7. <https://doi.org/10.1016/j.jclinepi.2016.03.003>.
- Austin, Peter C. 2010. « Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score ». *American Journal of Epidemiology* 172 (9): 1092. <https://doi.org/10.1093/aje/kwq224>.

- . 2013. « The Performance of Different Propensity Score Methods for Estimating Marginal Hazard Ratios ». *Statistics in Medicine* 32 (16): 2837-49. <https://doi.org/10.1002/sim.5705>.
- . 2022. « Bootstrap vs Asymptotic Variance Estimation When Using Propensity Score Weighting with Continuous and Binary Outcomes ». *Statistics in Medicine* 41 (22): 4426-43. <https://doi.org/10.1002/sim.9519>.
- Austin, Peter C., Paul Grootendorst, Sharon-Lise T. Normand, et Geoffrey M. Anderson. 2007. « Conditioning on the Propensity Score Can Result in Biased Estimation of Common Measures of Treatment Effect: A Monte Carlo Study ». *Statistics in Medicine* 26 (4): 754-68. <https://doi.org/10.1002/sim.2618>.
- Benevent, Justine, Mélanie Araujo, Caroline Hurault-Delarue, Jean-Louis Montastruc, Agnès Sommet, Isabelle Lacroix, et Christine Damase-Michel. 2019. « Pharmacoepidemiology in pregnancy ». *Therapies, PHARMACOEPIDEMIOLGY*, 74 (2): 289-300. <https://doi.org/10.1016/j.therap.2018.11.014>.
- Bérard, Anick, Fatima Abbas-Chorfa, Behrouz Kassai, Thierry Vial, Kim An Nguyen, Odile Sheehy, et Anne-Marie Schott. 2019. « The French Pregnancy Cohort: Medication Use during Pregnancy in the French Population ». *PLOS ONE* 14 (7): e0219095. <https://doi.org/10.1371/journal.pone.0219095>.
- Bérard, Anick, et Shashidhar Kori. 2012. « Dihydroergotamine (DHE) Use During Gestation and the Risk of Adverse Pregnancy Outcomes ». *Headache: The Journal of Head and Face Pain* 52 (7): 1085-93. <https://doi.org/10.1111/j.1526-4610.2012.02172.x>.
- Blum, Manuel R, Yuan Jin Tan, et John P A Ioannidis. 2020. « Use of E-values for addressing confounding in observational studies—an empirical assessment of the literature ». *International Journal of Epidemiology* 49 (5): 1482-94. <https://doi.org/10.1093/ije/dyz261>.
- Brookhart, M. Alan, Sebastian Schneeweiss, Kenneth J. Rothman, Robert J. Glynn, Jerry Avorn, et Til Stürmer. 2006. « Variable Selection for Propensity Score Models ». *American Journal of Epidemiology* 163 (12): 1149-56. <https://doi.org/10.1093/aje/kwj149>.

- Cannas, Massimo, et Bruno Arpino. 2019. « A Comparison of Machine Learning Algorithms and Covariate Balance Measures for Propensity Score Matching and Weighting ». *Biometrical Journal* 61 (4): 1049-72. <https://doi.org/10.1002/bimj.201800132>.
- Chatton, Arthur, Florent Le Borgne, Clémence Leyrat, Florence Gillaizeau, Chloé Rousseau, Laetitia Barbin, David Laplaud, Maxime Léger, Bruno Giraudeau, et Yohann Foucher. 2020. « G-Computation, Propensity Score-Based Methods, and Targeted Maximum Likelihood Estimator for Causal Inference with Different Covariates Sets: A Comparative Simulation Study ». *Scientific Reports* 10 (1): 9219. <https://doi.org/10.1038/s41598-020-65917-x>.
- Choi, Eun-Young, Han Eol Jeong, Yunha Noh, Ahhyung Choi, Dong Keon Yon, Jung Yeol Han, Ji-Hee Sung, Seung-Ah Choe, et Ju-Young Shin. 2023. « Neonatal and Maternal Adverse Outcomes and Exposure to Nonsteroidal Anti-Inflammatory Drugs during Early Pregnancy in South Korea: A Nationwide Cohort Study ». *PLoS Medicine* 20 (2): e1004183. <https://doi.org/10.1371/journal.pmed.1004183>.
- Cochran, W. G. 1968. « The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies ». *Biometrics* 24 (2): 295-313.
- Coppola, Paola, Essam Kerwash, Janet Nooney, Amro Omran, et Susan Cole. 2022. « Pharmacokinetic Data in Pregnancy: A Review of Available Literature Data and Important Considerations in Collecting Clinical Data ». *Frontiers in Medicine* 9 (octobre). <https://doi.org/10.3389/fmed.2022.940644>.
- Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, et Oscar A. Mitnik. 2009. « Dealing with limited overlap in estimation of average treatment effects ». *Biometrika* 96 (1): 187-99. <https://doi.org/10.1093/biomet/asn055>.
- Daniel, Sharon, Gideon Koren, Eitan Lunenfeld, Natalya Bilenko, Ronit Ratzon, et Amalia Levy. 2014. « Fetal Exposure to Nonsteroidal Anti-Inflammatory Drugs and Spontaneous Abortions ». *CMAJ* 186 (5): E177-82. <https://doi.org/10.1503/cmaj.130605>.

- Daniel, Sharon, Gideon Koren, Eitan Lunenfeld, et Amalia Levy. 2015a. « Immortal Time Bias in Drug Safety Cohort Studies: Spontaneous Abortion Following Nonsteroidal Antiinflammatory Drug Exposure ». *American Journal of Obstetrics and Gynecology* 212 (3): 307.e1-6. <https://doi.org/10.1016/j.ajog.2014.09.028>.
- . 2015b. « NSAIDs and spontaneous abortions – true effect or an indication bias? » *British Journal of Clinical Pharmacology* 80 (4): 750-54. <https://doi.org/10.1111/bcp.12653>.
- Dathe, Katarina, Anne-Katrin Fietz, Lucas William Pritchard, Stephanie Padberg, Stefanie Hultsch, Katja Meixner, Reinhard Meister, et Christof Schaefer. 2018. « No evidence of adverse pregnancy outcome after exposure to ibuprofen in the first trimester – Evaluation of the national Embryotox cohort ». *Reproductive Toxicology* 79 (août):32-38. <https://doi.org/10.1016/j.reprotox.2018.05.003>.
- Demilly, Romain, Sylvie Escolano, Françoise Haramburu, Pascale Tubert-Bitter, et Ismaïl Ahmed. 2020. « Identifying Drugs Inducing Prematurity by Mining Claims Data with High-Dimensional Confounder Score Strategies ». *Drug Safety* 43 (6): 549-59. <https://doi.org/10.1007/s40264-020-00916-5>.
- Demilly, Romain, Sylvie Escolano, Catherine Quantin, Pascale Tubert-Bitter, et Ismaïl Ahmed. 2017. « Prescription Drug Use during Pregnancy in France: A Study from the National Health Insurance Permanent Sample ». *Pharmacoepidemiology and Drug Safety* 26 (9): 1126-34. <https://doi.org/10.1002/pds.4265>.
- Edwards, Digna R. Velez, Tiara Aldridge, Donna D. Baird, Michele Jonsson Funk, David A. Savitz, et Katherine E. Hartmann. 2012. « Periconceptional Over-the-Counter Nonsteroidal Anti-Inflammatory Drug Exposure and Risk for Spontaneous Abortion ». *Obstetrics & Gynecology* 120 (1): 113. <https://doi.org/10.1097/AOG.0b013e3182595671>.
- Ericson, Anders, et Bengt A. J. Källén. 2001. « Nonsteroidal anti-inflammatory drugs in early pregnancy ». *Reproductive Toxicology* 15 (4): 371-75. [https://doi.org/10.1016/S0890-6238\(01\)00137-X](https://doi.org/10.1016/S0890-6238(01)00137-X).

- FDA. 2024. « Nonsteroidal Anti-Inflammatory Drugs (NSAIDs): Drug Safety Communication - Avoid Use of NSAIDs in Pregnancy at 20 Weeks or Later ». *FDA*, septembre. <https://www.fda.gov/safety/medical-product-safety-information/nonsteroidal-anti-inflammatory-drugs-nsaids-drug-safety-communication-avoid-use-nsaids-pregnancy-20>.
- Flanders, W. Dana, Mitchel Klein, Lyndsey A. Darrow, Matthew J. Strickland, Stefanie E. Sarnat, Jeremy A. Sarnat, Lance A. Waller, Andrea Winquist, et Paige E. Tolbert. 2011. « A Method for Detection of Residual Confounding in Time-Series and Other Observational Studies ». *Epidemiology* 22 (1): 59. <https://doi.org/10.1097/EDE.0b013e3181fdcab>.
- Fox, Matthew P, Onyebuchi A Arah, et Elizabeth A Stuart. 2020. « Commentary: The value of E-values and why they are not enough ». *International Journal of Epidemiology* 49 (5): 1505-6. <https://doi.org/10.1093/ije/dyaa093>.
- Gelder, Marleen M. H. J. van, Angela Lupattelli, et Hedvig M. E. Nordeng. 2023. « Risk of Spontaneous Abortion after Periconceptional Medication Use: Time to Tackle the Methodological Challenges ». *Paediatric and Perinatal Epidemiology* 37 (3): 188-90. <https://doi.org/10.1111/ppe.12967>.
- Gelder, Marleen M. H. J. van, Nel Roeleveld, et Hedvig Nordeng. 2011. « Exposure to Non-Steroidal Anti-Inflammatory Drugs during Pregnancy and the Risk of Selected Birth Defects: A Prospective Cohort Study ». *PLoS ONE* 6 (7): e22174. <https://doi.org/10.1371/journal.pone.0022174>.
- Germay, Sibylle de, Cécile Conte, Joëlle Micallef, Emilie Bouquet, Laurent Chouchana, Margaux Lafaurie, et Antoine Pariente. 2023. « Performing pharmacoepidemiological studies using the French health insurance data warehouse (SNDS): How to translate guidelines into practice ». *Therapies* 78 (6): 679-89. <https://doi.org/10.1016/j.therap.2023.01.009>.
- Greenland, Sander. 2020. « Commentary: An argument against E-values for assessing the plausibility that an association could be explained away by residual confounding ». *International Journal of Epidemiology* 49 (5): 1501-3. <https://doi.org/10.1093/ije/dyaa095>.

- Groenwold, Rolf H. H. 2013. « Falsification End Points for Observational Studies ». *JAMA* 309 (17): 1769-71. <https://doi.org/10.1001/jama.2013.3089>.
- Guertin, Jason R., Elham Rahme, et Jacques LeLorier. 2016. « Performance of the High-Dimensional Propensity Score in Adjusting for Unmeasured Confounders ». *European Journal of Clinical Pharmacology* 72 (12): 1497-1505. <https://doi.org/10.1007/s00228-016-2118-x>.
- Hajage, David, Florence Tubach, Philippe Gabriel Steg, Deepak L. Bhatt, et Yann De Rycke. 2016. « On the use of propensity scores in case of rare exposure ». *BMC Medical Research Methodology* 16 (1): 38. <https://doi.org/10.1186/s12874-016-0135-1>.
- Hennessy, Sean. 2006. « Use of Health Care Databases in Pharmacoepidemiology ». *Basic & Clinical Pharmacology & Toxicology* 98 (3): 311-13. https://doi.org/10.1111/j.1742-7843.2006.pto_368.x.
- Hicks, Blánaid, James A. Kaye, Laurent Azoulay, Kasper Bruun Kristensen, Laurel A. Habel, et Anton Pottegård. 2023. « The Application of Lag Times in Cancer Pharmacoepidemiology: A Narrative Review ». *Annals of Epidemiology* 84 (août):25-32. <https://doi.org/10.1016/j.annepidem.2023.05.004>.
- Höfler, M. 2005. « Causal Inference Based on Counterfactuals ». *BMC Medical Research Methodology* 5 (1): 28. <https://doi.org/10.1186/1471-2288-5-28>.
- Imbens, Guido W. 2004. « Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review ». *The Review of Economics and Statistics* 86 (1): 4-29. <https://doi.org/10.1162/003465304323023651>.
- Ioannidis, John P. A., Yuan Jin Tan, et Manuel R. Blum. 2019. « Limitations and Misinterpretations of E-Values for Sensitivity Analyses of Observational Studies ». *Annals of Internal Medicine* 170 (2): 108-11. <https://doi.org/10.7326/M18-2159>.
- Jackson, Lisa A., Michael L. Jackson, Jennifer C. Nelson, Kathleen M. Neuzil, et Noel S. Weiss. 2006. « Evidence of Bias in Estimates of Influenza Vaccine

- Effectiveness in Seniors ». *International Journal of Epidemiology* 35 (2): 337-44. <https://doi.org/10.1093/ije/dyi274>.
- Jain, Venu, Sarah D. McDonald, William R. Mundle, et Dan Farine. 2020. « Guideline No. 398: Progesterone for Prevention of Spontaneous Preterm Birth ». *Journal of Obstetrics and Gynaecology Canada* 42 (6): 806-12. <https://doi.org/10.1016/j.jogc.2019.04.012>.
- Jones, Mark, et Robert Fowler. 2016. « Immortal time bias in observational studies of time-to-event outcomes ». *Journal of Critical Care* 36 (décembre):195-99. <https://doi.org/10.1016/j.jcrc.2016.07.017>.
- Ju, Cheng, Susan Gruber, Samuel D Lendle, Antoine Chambaz, Jessica M Franklin, Richard Wyss, Sebastian Schneeweiss, et Mark J van der Laan. 2019. « Scalable collaborative targeted learning for high-dimensional data ». *Statistical methods in medical research* 28 (2): 532-54. <https://doi.org/10.1177/0962280217729845>.
- Ju, Cheng, Joshua Schwab, et Mark J. van der Laan. 2019. « On Adaptive Propensity Score Truncation in Causal Inference ». *Statistical Methods in Medical Research* 28 (6): 1741-60. <https://doi.org/10.1177/0962280218774817>.
- Kahlert, Johnny, Sigrid Bjerger Gribsholt, Henrik Gammelager, Olaf M Dekkers, et George Luta. 2017. « Control of confounding in the analysis phase – an overview for clinicians ». *Clinical Epidemiology* 9 (mars):195-204. <https://doi.org/10.2147/CLEP.S129886>.
- Kang, Joseph D. Y., et Joseph L. Schafer. 2007. « Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data ». *Statistical Science* 22 (4): 523-39. <https://doi.org/10.1214/07-STS227>.
- Karim, Mohammad Ehsanul, Menglan Pang, et Robert W. Platt. 2018. « Can We Train Machine Learning Methods to Outperform the High-Dimensional Propensity Score Algorithm? ». *Epidemiology* 29 (2): 191. <https://doi.org/10.1097/EDE.0000000000000787>.
- Keil, Alexander P., Jessie K. Edwards, David R. Richardson, Ashley I. Naimi, et Stephen R. Cole. 2014. « The parametric G-formula for time-to-event

data: towards intuition with a worked example ». *Epidemiology (Cambridge, Mass.)* 25 (6): 889-97. <https://doi.org/10.1097/EDE.000000000000160>.

Kreif, Noémi, Susan Gruber, Rosalba Radice, Richard Grieve, et Jasjeet S Sekhon. 2016. « Evaluating Treatment Effectiveness under Model Misspecification: A Comparison of Targeted Maximum Likelihood Estimation with Bias-Corrected Matching ». *Statistical Methods in Medical Research* 25 (5): 2315-36. <https://doi.org/10.1177/0962280214521341>.

Laan, Mark J. van der, et Susan Gruber. 2010. « Collaborative Double Robust Targeted Maximum Likelihood Estimation ». *The International Journal of Biostatistics* 6 (1). <https://doi.org/10.2202/1557-4679.1181>.

Laan, Mark J. van der, Eric C. Polley, et Alan E. Hubbard. 2007. « Super Learner ». *Statistical Applications in Genetics and Molecular Biology* 6:Article25. <https://doi.org/10.2202/1544-6115.1309>.

Laan, Mark J. van der, et Daniel Rubin. 2006. « Targeted Maximum Likelihood Learning ». *The International Journal of Biostatistics* 2 (1). <https://doi.org/10.2202/1557-4679.1043>.

Lash, Timothy L., Matthew P. Fox, Richard F. MacLehose, George Maldonado, Lawrence C. McCandless, et Sander Greenland. 2014. « Good Practices for Quantitative Bias Analysis ». *International Journal of Epidemiology* 43 (6): 1969-85. <https://doi.org/10.1093/ije/dyu149>.

Le Borgne, Florent, Arthur Chatton, Maxime Léger, Rémi Lenain, et Yohann Foucher. 2021. « G-Computation and Machine Learning for Estimating the Causal Effects of Binary Exposure Statuses on Binary Outcomes ». *Scientific Reports* 11 (1): 1435. <https://doi.org/10.1038/s41598-021-81110-0>.

Leathers, Tess A., et Crystal D. Rogers. 2023. « Nonsteroidal Anti-Inflammatory Drugs and Implications for the Cyclooxygenase Pathway in Embryonic Development ». *American Journal of Physiology - Cell Physiology* 324 (2): C532. <https://doi.org/10.1152/ajpcell.00430.2022>.

- Lee, Brian K., Justin Lessler, et Elizabeth A. Stuart. 2010. « Improving propensity score weighting using machine learning ». *Statistics in medicine* 29 (3): 337-46. <https://doi.org/10.1002/sim.3782>.
- Lendle, Samuel D., Bruce Fireman, et Mark J. van der Laan. 2013. « Targeted Maximum Likelihood Estimation in Safety Analysis ». *Journal of Clinical Epidemiology* 66 (8): S91-98. <https://doi.org/10.1016/j.jclinepi.2013.02.017>.
- Li, De-Kun, Jeannette R. Ferber, Roxana Odouli, et Charles Quesenberry. 2018. « Use of Nonsteroidal Antiinflammatory Drugs during Pregnancy and the Risk of Miscarriage ». *American Journal of Obstetrics and Gynecology* 219 (3): 275.e1-275.e8. <https://doi.org/10.1016/j.ajog.2018.06.002>.
- Li, De-Kun, Liyan Liu, et Roxana Odouli. 2003a. « Exposure to non-steroidal anti-inflammatory drugs during pregnancy and risk of miscarriage: population based cohort study ». *BMJ: British Medical Journal* 327 (7411): 368.
- . 2003b. « Exposure to non-steroidal anti-inflammatory drugs during pregnancy and risk of miscarriage: population based cohort study ». *BMJ: British Medical Journal* 327 (7411): 368.
- Li, Fan, Laine E Thomas, et Fan Li. 2019. « Addressing Extreme Propensity Scores via the Overlap Weights ». *American Journal of Epidemiology* 188 (1): 250-57. <https://doi.org/10.1093/aje/kwy201>.
- Li, Liang, et Tom Greene. 2013. « A Weighting Analogue to Pair Matching in Propensity Score Analysis ». *The International Journal of Biostatistics* 9 (2): 215-34. <https://doi.org/10.1515/ijb-2012-0030>.
- Lipsitch, Marc, Eric Tchetgen Tchetgen, et Ted Cohen. 2010. « Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies ». *Epidemiology* 21 (3): 383. <https://doi.org/10.1097/EDE.0b013e3181d61eeb>.
- Little, R. J., et D. B. Rubin. 2000. « Causal Effects in Clinical and Epidemiological Studies via Potential Outcomes: Concepts and Analytical Approaches ». *Annual Review of Public Health* 21:121-45. <https://doi.org/10.1146/annurev.publhealth.21.1.121>.

- Louchet, Margaux, Mathis Collier, Nathanaël Beeker, Laurent Mandelbrot, Jeanne Sibiude, Laurent Chouchana, et Jean Marc Treluyer. 2024. « Trends in Harmful Drug Exposure during Pregnancy in France between 2013 and 2019: A Nationwide Cohort Study ». *PLOS ONE* 19 (1): e0295897. <https://doi.org/10.1371/journal.pone.0295897>.
- Lunceford, Jared K., et Marie Davidian. 2004. « Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study ». *Statistics in Medicine* 23 (19): 2937-60. <https://doi.org/10.1002/sim.1903>.
- Lupattelli, A., O. Spigset, M. J. Twigg, K. Zagorodnikova, A. C. Mårdby, M. E. Moretti, M. Drozd, et al. 2014. « Medication Use in Pregnancy: A Cross-Sectional, Multinational Web-Based Study ». *BMJ Open* 4 (2): e004365. <https://doi.org/10.1136/bmjopen-2013-004365>.
- Luque-Fernandez, Miguel Angel, Michael Schomaker, Bernard Rachet, et Mireille E. Schnitzer. 2018. « Targeted Maximum Likelihood Estimation for a Binary Treatment: A Tutorial ». *Statistics in Medicine* 37 (16): 2530-46. <https://doi.org/10.1002/sim.7628>.
- Maillard, Olivier, René Bun, Moussa Laanani, Amandine Verga-Gérard, Taylor Leroy, Nathalie Gault, Candice Estellat, et al. 2024. « Use of the French National Health Data System (SNDS) in pharmacoepidemiology: A systematic review in its maturation phase ». *Therapies*, mai. <https://doi.org/10.1016/j.therap.2024.05.003>.
- Medicines and Healthcare products Regulatory Agency. 2024. « Non-Steroidal Anti-Inflammatory Drugs (NSAIDs): Potential Risks Following Prolonged Use after 20 Weeks of Pregnancy ». GOV.UK. 3 juillet 2024. <https://www.gov.uk/drug-safety-update/non-steroidal-anti-inflammatory-drugs-nsaids-potential-risks-following-prolonged-use-after-20-weeks-of-pregnancy>.
- Miranda, S., J. Drouin, J. Botton, T. Le Tri, A. Meyer, M. Zureik, A. Weill, et R. Dray-Spira. 2023. « Registre national de données Mères-Enfants à partir du Système national des données de santé pour des études pharmaco-épidémiologiques ». *Revue d'Épidémiologie et de Santé Publique*, Congrès national Emois 2023, 71 (mars):101539. <https://doi.org/10.1016/j.respe.2023.101539>.

- Naimi, Ashley I, Alan E Mishler, et Edward H Kennedy. 2021. « Challenges in Obtaining Valid Causal Effect Estimates with Machine Learning Algorithms ». *American Journal of Epidemiology*, juillet, kwab201. <https://doi.org/10.1093/aje/kwab201>.
- Nakhai-Pour, Hamid Reza, Perrine Broy, Odile Sheehy, et Anick Bérard. 2011. « Use of Nonaspirin Nonsteroidal Anti-Inflammatory Drugs during Pregnancy and the Risk of Spontaneous Abortion ». *CMAJ* 183 (15): 1713-20. <https://doi.org/10.1503/cmaj.110454>.
- New Zealand Medicines and Medical Devices Safety Authority. 2024. « Non-steroidal anti-inflammatory drugs (NSAIDs): avoid use in pregnancy ». 8 octobre 2024. <https://medsafe.govt.nz/profs/PUArticles/September2022/Non-steroidal-anti-inflammatory%20drugs-NSAIDs-avoid-use%20in-pregnancy.html>.
- Nielsen, Gunnar Lauge, Henrik Toft Sørensen, Helle Larsen, et Lars Pedersen. 2001. « Risk of adverse birth outcome and miscarriage in pregnant users of non-steroidal anti-inflammatory drugs: population based observational study and case-control study ». *BMJ: British Medical Journal* 322 (7281): 266-70.
- Norman, Jane E. 2020. « Progesterone and preterm birth ». *International Journal of Gynaecology and Obstetrics* 150 (1): 24-30. <https://doi.org/10.1002/ijgo.13187>.
- Pang, Menglan, Tibor Schuster, Kristian B. Fillion, Maria Eberg, et Robert W. Platt. 2016. « Targeted Maximum Likelihood Estimation for Pharmacoepidemiologic Research ». *Epidemiology* 27 (4): 570. <https://doi.org/10.1097/EDE.0000000000000487>.
- Petersen, Maya L, Kristin E Porter, Susan Gruber, Yue Wang, et Mark J van der Laan. 2012. « Diagnosing and Responding to Violations in the Positivity Assumption ». *Statistical Methods in Medical Research* 21 (1): 31-54. <https://doi.org/10.1177/0962280210386207>.
- Pirracchio, Romain, Maya L. Petersen, et Mark van der Laan. 2015. « Improving Propensity Score Estimators' Robustness to Model Misspecification

- Using Super Learner ». *American Journal of Epidemiology* 181 (2): 108-19. <https://doi.org/10.1093/aje/kwu253>.
- Prada-Ramallal, Guillermo, Bahi Takkouche, et Adolfo Figueiras. 2019. « Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review ». *BMC Medical Research Methodology* 19 (1): 53. <https://doi.org/10.1186/s12874-019-0695-y>.
- Prasad, Vinay, et Anupam B. Jena. 2013. « Prespecified Falsification End Points: Can They Validate True Observational Associations? » *JAMA* 309 (3): 241-42. <https://doi.org/10.1001/jama.2012.96867>.
- Quantin, C, C Yamdjieu Ngadeu, J Cottenet, S Escolano, S Bechraoui-Quantin, P Rozenberg, P Tubert-Bitter, et J-B Gouyon. 2021. « Early Exposure of Pregnant Women to Non-Steroidal Anti-Inflammatory Drugs Delivered Outside Hospitals and Preterm Birth Risk: Nationwide Cohort Study ». *BJOG: An International Journal of Obstetrics & Gynaecology* 128 (10): 1575-84. <https://doi.org/10.1111/1471-0528.16670>.
- Rachas, Antoine, Christelle Gastaldi-Ménager, Pierre Denis, Pauline Barthélémy, Panayotis Constantinou, Jérôme Drouin, Dimitri Lastier, et al. 2022. « The Economic Burden of Disease in France From the National Health Insurance Perspective: The Healthcare Expenditures and Conditions Mapping Used to Prepare the French Social Security Funding Act and the Public Health Act ». *Medical Care* 60 (9): 655. <https://doi.org/10.1097/MLR.0000000000001745>.
- Rassen, Jeremy A., Patrick Blin, Sebastian Kloss, Romain S. Neugebauer, Robert W. Platt, Anton Pottegård, Sebastian Schneeweiss, et Sengwee Toh. 2023. « High-Dimensional Propensity Scores for Empirical Covariate Selection in Secondary Database Studies: Planning, Implementation, and Reporting ». *Pharmacoepidemiology and Drug Safety* 32 (2): 93-106. <https://doi.org/10.1002/pds.5566>.
- Rassen, Jeremy A., Robert J. Glynn, M. Alan Brookhart, et Sebastian Schneeweiss. 2011. « Covariate Selection in High-Dimensional Propensity Score Analyses of Treatment Effects in Small Samples ». *American Journal of Epidemiology* 173 (12): 1404-13. <https://doi.org/10.1093/aje/kwr001>.

- Rey, Grégoire, Eric Jouglu, Anne Fouillet, et Denis Hémon. 2009. « Ecological association between a deprivation index and mortality in France over the period 1997 – 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death ». *BMC Public Health* 9 (janvier):33. <https://doi.org/10.1186/1471-2458-9-33>.
- Ridder, Maria A J de, Marcel de Wilde, Christina de Ben, Armando R Leyba, Bartholomeus M T Mosseveld, Katia M C Verhamme, Johan van der Lei, et Peter R Rijnbeek. 2022. « Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands ». *International Journal of Epidemiology* 51 (6): e314-23. <https://doi.org/10.1093/ije/dyac026>.
- Robins, James. 1986. « A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect ». *Mathematical Modelling* 7 (9): 1393-1512. [https://doi.org/10.1016/0270-0255\(86\)90088-6](https://doi.org/10.1016/0270-0255(86)90088-6).
- Romero, Roberto, Lami Yeo, Piya Chaemsaitong, Tinnakorn Chaiworapongsa, et Sonia Hassan. 2014. « Progesterone to prevent spontaneous preterm birth ». *Seminars in fetal & neonatal medicine* 19 (1): 15-26. <https://doi.org/10.1016/j.siny.2013.10.004>.
- Rosenbaum, Paul R., et Donald B. Rubin. 1983. « The central role of the propensity score in observational studies for causal effects ». *Biometrika* 70 (1): 41-55. <https://doi.org/10.1093/biomet/70.1.41>.
- . 1984. « Reducing Bias in Observational Studies Using Subclassification on the Propensity Score ». *Journal of the American Statistical Association* 79 (387): 516-24. <https://doi.org/10.2307/2288398>.
- . 1985. « Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score ». *The American Statistician* 39 (1): 33-38. <https://doi.org/10.1080/00031305.1985.10479383>.
- Sabourin, Jeremy A., William Valdar, et Andrew B. Nobel. 2015. « A Permutation Approach for Selecting the Penalty Parameter in Penalized Model Selection ». *Biometrics* 71 (4): 1185-94. <https://doi.org/10.1111/biom.12359>.

- San Lazaro Campillo, Indra, Sarah Meaney, Maria Harrington, Karen McNamara, Anna Maria Verling, Paul Corcoran, et Keelin O'Donoghue. 2021. « Assessing the Concordance and Accuracy between Hospital Discharge Data, Electronic Health Records, and Register Books for Diagnosis of Inpatient Admissions of Miscarriage: A Retrospective Linked Data Study ». *The Journal of Obstetrics and Gynaecology Research* 47 (6): 1987-96. <https://doi.org/10.1111/jog.14785>.
- Sanson-Fisher, Robert William, Billie Bonevski, Lawrence W. Green, et Cate D'Este. 2007. « Limitations of the Randomized Controlled Trial in Evaluating Population-Based Health Interventions ». *American Journal of Preventive Medicine* 33 (2): 155-61. <https://doi.org/10.1016/j.amepre.2007.04.007>.
- Schafer, Joseph L., et Joseph Kang. 2008. « Average Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example ». *Psychological Methods* 13 (4): 279-313. <https://doi.org/10.1037/a0014268>.
- Schneeweiss, Sebastian. 2018. « Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects ». *Clinical Epidemiology* 10 (juillet):771-88. <https://doi.org/10.2147/CLEP.S166545>.
- Schneeweiss, Sebastian, et Jerry Avorn. 2005. « A review of uses of health care utilization databases for epidemiologic research on therapeutics ». *Journal of Clinical Epidemiology* 58 (4): 323-37. <https://doi.org/10.1016/j.jclinepi.2004.10.012>.
- Schneeweiss, Sebastian, Wesley Eddings, Robert J. Glynn, Elisabetta Patorno, Jeremy Rassen, et Jessica M. Franklin. 2017. « Variable Selection for Confounding Adjustment in High-Dimensional Covariate Spaces When Analyzing Healthcare Databases ». *Epidemiology (Cambridge, Mass.)* 28 (2): 237-48. <https://doi.org/10.1097/EDE.0000000000000581>.
- Schneeweiss, Sebastian, et Jeremy Rassen. 2011. « Letter to the Editor ». *Pharmacoepidemiology and Drug Safety* 20 (10): 1110-11. <https://doi.org/10.1002/pds.2238>.
- Schneeweiss, Sebastian, Jeremy A. Rassen, Robert J. Glynn, Jerry Avorn, Helen Mogun, et M. Alan Brookhart. 2009. « High-Dimensional Propensity

Score Adjustment in Studies of Treatment Effects Using Health Care Claims Data ». *Epidemiology* 20 (4): 512-22. <https://doi.org/10.1097/EDE.0b013e3181a663cc>.

Schuler, Megan S., et Sherri Rose. 2017. « Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies ». *American Journal of Epidemiology* 185 (1): 65-73. <https://doi.org/10.1093/aje/kww165>.

Schuster, Tibor, Wilfrid Kouokam Lowe, et Robert W. Platt. 2016. « Propensity score model overfitting led to inflated variance of estimated odds ratios ». *Journal of clinical epidemiology* 80 (décembre):97-106. <https://doi.org/10.1016/j.jclinepi.2016.05.017>.

Shmueli, Galit. 2010. « To Explain or to Predict? » *Statistical Science* 25 (3): 289-310. <https://doi.org/10.1214/10-STS330>.

Shortreed, Susan M., et Ashkan Ertefaie. 2017. « Outcome-Adaptive Lasso: Variable Selection for Causal Inference ». *Biometrics* 73 (4): 1111-22. <https://doi.org/10.1111/biom.12679>.

Smith, Matthew J., Rachael V. Phillips, Miguel Angel Luque-Fernandez, et Camille Maringe. 2023. « Application of Targeted Maximum Likelihood Estimation in Public Health and Epidemiological Studies: A Systematic Review ». *Annals of Epidemiology* 86 (octobre):34-48.e28. <https://doi.org/10.1016/j.annepidem.2023.06.004>.

Snowden, Jonathan M., Sherri Rose, et Kathleen M. Mortimer. 2011. « Implementation of G-Computation on a Simulated Data Set: Demonstration of a Causal Inference Technique ». *American Journal of Epidemiology* 173 (7): 731-38. <https://doi.org/10.1093/aje/kwq472>.

Sofrygin, Oleg, Zheng Zhu, Julie A. Schmittdiel, Alyce S. Adams, Richard W. Grant, Mark J. van der Laan, et Romain Neugebauer. 2019. « Targeted Learning with Daily EHR Data ». *Statistics in Medicine* 38 (16): 3073-90. <https://doi.org/10.1002/sim.8164>.

Stürmer, Til, Kenneth J. Rothman, Jerry Avorn, et Robert J. Glynn. 2010. « Treatment Effects in the Presence of Unmeasured Confounding: Dealing with Observations in the Tails of the Propensity Score

- Distribution--a Simulation Study ». *American Journal of Epidemiology* 172 (7): 843-54. <https://doi.org/10.1093/aje/kwq198>.
- Suissa, Samy. 2008. « Immortal Time Bias in Pharmacoepidemiology ». *American Journal of Epidemiology* 167 (4): 492-99. <https://doi.org/10.1093/aje/kwm324>.
- Surén, Pål, Christine Roth, Michaeline Bresnahan, Margaretha Haugen, Mady Hornig, Deborah Hirtz, Kari Kveim Lie, et al. 2013. « Association between Maternal Use of Folic Acid Supplements and Risk of Autism Spectrum Disorders in Children ». *JAMA* 309 (6): 570-77. <https://doi.org/10.1001/jama.2012.155925>.
- Tamim, H., A. A. Tahami Monfared, et J. LeLorier. 2007. « Application of Lag-Time into Exposure Definitions to Control for Protopathic Bias ». *Pharmacoepidemiology and Drug Safety* 16 (3): 250-58. <https://doi.org/10.1002/pds.1360>.
- Tazare, John, Richard Wyss, Jessica M. Franklin, Liam Smeeth, Stephen J. W. Evans, Shirley V. Wang, Sebastian Schneeweiss, Ian J. Douglas, Joshua J. Gagne, et Elizabeth J. Williamson. 2022. « Transparency of High-Dimensional Propensity Score Analyses: Guidance for Diagnostics and Reporting ». *Pharmacoepidemiology and Drug Safety* 31 (4): 411-23. <https://doi.org/10.1002/pds.5412>.
- Thake, Miriam, et Andrew Lowry. 2017. « A systematic review of trends in the selective exclusion of older participant from randomised clinical trials ». *Archives of Gerontology and Geriatrics* 72 (septembre):99-102. <https://doi.org/10.1016/j.archger.2017.05.017>.
- Thangaraju, Pugazhenthana, Shoban Babu Varthya, Prakash Srinivasan Timiri Shanmugam, T. Y. Sree Sudha, et Chaitanya Mittal. 2021. « Hydroxyprogesterone caproate for preterm birth: A situational analysis from the Indian perspective ». *Gynecology and Obstetrics Clinical Medicine* 1 (3): 144-47. <https://doi.org/10.1016/j.gocm.2021.08.002>.
- Tian, Yuxi, Martijn J Schuemie, et Marc A Suchard. 2018. « Evaluating large-scale propensity score performance through real-world and synthetic

- data experiments ». *International Journal of Epidemiology* 47 (6): 2005-14. <https://doi.org/10.1093/ije/dyy120>.
- Tibshirani, Robert. 1996. « Regression Shrinkage and Selection Via the Lasso ». *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267-88. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
- Toh, Sengwee, Luis A. García Rodríguez, et Miguel A. Hernán. 2011. « Confounding Adjustment via a Semi-Automated High-Dimensional Propensity Score Algorithm: An Application to Electronic Medical Records ». *Pharmacoepidemiology and Drug Safety* 20 (8): 849-57. <https://doi.org/10.1002/pds.2152>.
- Trinquart, Ludovic, Adrienne L Erlinger, Julie M Petersen, Matthew Fox, et Sandro Galea. 2019. « Applying the E Value to Assess the Robustness of Epidemiologic Fields of Inquiry to Unmeasured Confounding ». *American Journal of Epidemiology* 188 (6): 1174-80. <https://doi.org/10.1093/aje/kwz063>.
- Tuppin, P., J. Rudant, P. Constantinou, C. Gastaldi-Ménager, A. Rachas, L. de Roquefeuil, G. Maura, et al. 2017. « Value of a national administrative database to guide public decisions: From the *système national d'information interrégimes de l'Assurance Maladie* (SNIIRAM) to the *système national des données de santé* (SNDS) in France ». *Revue d'Épidémiologie et de Santé Publique, Réseau REDSIAM*, 65 (octobre):S149-67. <https://doi.org/10.1016/j.respe.2017.05.004>.
- Uddin, Md Jamal, Rolf H. H. Groenwold, Mohammed Sanni Ali, Anthonius de Boer, Kit C. B. Roes, Muhammad A. B. Chowdhury, et Olaf H. Klungel. 2016. « Methods to Control for Unmeasured Confounding in Pharmacoepidemiology: An Overview ». *International Journal of Clinical Pharmacy* 38 (3): 714-23. <https://doi.org/10.1007/s11096-016-0299-0>.
- VanderWeele, Tyler J. 2009. « Concerning the Consistency Assumption in Causal Inference ». *Epidemiology (Cambridge, Mass.)* 20 (6): 880-83. <https://doi.org/10.1097/EDE.0b013e3181bd5638>.
- VanderWeele, Tyler J., et Peng Ding. 2017. « Sensitivity Analysis in Observational Research: Introducing the E-Value ». *Annals of Internal Medicine* 167 (4): 268-74. <https://doi.org/10.7326/M16-2607>.

- VanderWeele, Tyler J., Peng Ding, et Maya Mathur. 2019. « Technical Considerations in the Use of the E-Value ». *Journal of Causal Inference* 7 (2). <https://doi.org/10.1515/jci-2018-0007>.
- VanderWeele, Tyler J, et Maya B Mathur. 2020. « Commentary: Developing best-practice guidelines for the reporting of E-values ». *International Journal of Epidemiology* 49 (5): 1495-97. <https://doi.org/10.1093/ije/dyaa094>.
- Vansteelandt, Stijn, et Niels Keiding. 2011. « Invited Commentary: G-Computation—Lost in Translation? ». *American Journal of Epidemiology* 173 (7): 739-42. <https://doi.org/10.1093/aje/kwq474>.
- Weinstein, Rachel B., Patrick Ryan, Jesse A. Berlin, Amy Matcho, Martijn Schuemie, Joel Swerdel, Kayur Patel, et Daniel Fife. 2017. « Channeling in the Use of Nonprescription Paracetamol and Ibuprofen in an Electronic Medical Records Database: Evidence and Implications ». *Drug Safety* 40 (12): 1279-92. <https://doi.org/10.1007/s40264-017-0581-7>.
- Wolf, Achim, Daniel Dedman, Jennifer Campbell, Helen Booth, Darren Lunn, Jennifer Chapman, et Puja Myles. 2019. « Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum ». *International Journal of Epidemiology* 48 (6): 1740-1740g. <https://doi.org/10.1093/ije/dyz034>.
- Wyss, Richard, Joshua J. Gagne, Yueqin Zhao, Esther H. Zhou, Jacqueline M. Major, Shirley V. Wang, Rishi J. Desai, et al. 2020. « Use of Time-Dependent Propensity Scores to Adjust Hazard Ratio Estimates in Cohort Studies with Differential Depletion of Susceptibles ». *Epidemiology (Cambridge, Mass.)* 31 (1): 82-89. <https://doi.org/10.1097/EDE.0000000000001107>.
- Wyss, Richard, Sebastian Schneeweiss, Mark van der Laan, Samuel D. Lendle, Cheng Ju, et Jessica M. Franklin. 2018. « Using Super Learner Prediction Modeling to Improve High-Dimensional Propensity Score Estimation ». *Epidemiology* 29 (1): 96-106. <https://doi.org/10.1097/EDE.0000000000000762>.
- Xu, Stanley, Colleen Ross, Marsha A. Raebel, Susan Shetterly, Christopher Blanchette, et David Smith. 2010. « Use of Stabilized Inverse Propensity Scores as Weights to Directly Estimate Relative Risk and Its Confidence

- Intervals ». *Value in Health* 13 (2): 273-77. <https://doi.org/10.1111/j.1524-4733.2009.00671.x>.
- Ying, Xian-hua, Dan-ni Bao, Hai-yin Jiang, et Yu-dan Shi. 2022. « Maternal Non-Steroidal Anti-Inflammatory Drug Exposure during Pregnancy and Risk of Miscarriage: A Systematic Review and Meta-Analysis ». *European Journal of Clinical Pharmacology* 78 (2): 171-80. <https://doi.org/10.1007/s00228-021-03222-w>.
- Z, Ren, Bremer Aa, et Pawlyk Ac. 2021. « Drug Development Research in Pregnant and Lactating Women ». *American Journal of Obstetrics and Gynecology* 225 (1). <https://doi.org/10.1016/j.ajog.2021.04.227>.
- Zhang, Linying, Yixin Wang, Martijn J. Schuemie, David M. Blei, et George Hripcsak. 2022. « Adjusting for indirectly measured confounding using large-scale propensity score ». *Journal of Biomedical Informatics* 134 (octobre):104204. <https://doi.org/10.1016/j.jbi.2022.104204>.
- Zhang, Xiang, Douglas E. Faries, Hu Li, James D. Stamey, et Guido W. Imbens. 2018. « Addressing Unmeasured Confounding in Comparative Observational Research ». *Pharmacoepidemiology and Drug Safety* 27 (4): 373-82. <https://doi.org/10.1002/pds.4394>.
- Zhu, Yaqian, Rebecca A. Hubbard, Jessica Chubak, Jason Roy, et Nandita Mitra. 2021. « Core Concepts in Pharmacoepidemiology: Violations of the Positivity Assumption in the Causal Analysis of Observational Data: Consequences and Statistical Approaches ». *Pharmacoepidemiology and drug safety* 30 (11): 1471-85. <https://doi.org/10.1002/pds.5338>.
- Zivich, Paul N., et Alexander Breskin. 2021. « Machine Learning for Causal Inference: On the Use of Cross-Fit Estimators ». *Epidemiology (Cambridge, Mass.)* 32 (3): 393-401. <https://doi.org/10.1097/EDE.0000000000001332>.
- Zou, Hui. 2006. « The Adaptive Lasso and Its Oracle Properties ». *Journal of the American Statistical Association* 101 (476): 1418-29. <https://doi.org/10.1198/016214506000000735>.

**ANNEXES A : DOCUMENTS
SUPPLEMENTAIRES DES ANALYSES DU
CHAPITRE 2**

Annexe A.1 : Algorithme de sélection des grossesses dans le PMSI

Accouchements	
<u>Diagnostiques d'hospitalisation*</u>	
Z37	Résultat de l'accouchement
Z3900	Soins et examens immédiatement après un accouchement hors d'un établissement de santé
O80	Accouchement unique et spontané
O81	Accouchement unique par forceps et ventouse
O82	Accouchement unique par césarienne
O83	Autres accouchements uniques avec assistance
O84	Accouchements multiples
<u>Actes CCAM</u>	
JQGD010	Accouchement céphalique unique par voie naturelle, chez une primipare
JQGD012	Accouchement céphalique unique par voie naturelle, chez une multipare
JQGD004	Accouchement unique par le siège par voie naturelle, chez une primipare
JQGD001	Accouchement unique par le siège par voie naturelle, chez une multipare
JQGD003	Accouchement unique par le siège par voie naturelle avec petite extraction, chez une primipare
JQGD008	Accouchement unique par le siège par voie naturelle avec petite extraction, chez une multipare
JQGD013	Accouchement unique par le siège par voie naturelle avec grande extraction, chez une primipare
JQGD005	Accouchement unique par le siège par voie naturelle avec grande extraction, chez une multipare
JQGD002	Accouchement multiple par voie naturelle, chez une primipare
JQGD007	Accouchement multiple par voie naturelle, chez une multipare
JQGA002	Accouchement par césarienne programmée, par laparotomie
JQGA004	Accouchement par césarienne en urgence en dehors du travail, par laparotomie
JQGA003	Accouchement par césarienne au cours du travail, par laparotomie
JQGA005	Accouchement par césarienne, par abord vaginal
Interruptions de grossesse	
<u>Diagnostiques d'hospitalisation*</u>	
O04 et	Avortement médical
O35	Soins maternels pour anomalie et lésion fœtales, connues ou présumées
O28	Résultats anormaux constatés au cours de l'examen prénatal systématique de la mère
<u>Diagnostiques d'hospitalisation*</u>	
O03	Avortement spontané
Grossesses extra-utérines	
<u>Diagnostiques d'hospitalisation*</u>	
O00	Grossesse extra-utérine
<u>Actes CCAM</u>	

JFA001	Salpingectomie partielle ou totale pour grossesse extra-utérine, par laparotomie
JFC001	Salpingectomie partielle ou totale pour grossesse extra-utérine, par coelioscopie
JJA002	Expression tubaire pour évacuation tubo abdominale de grossesse extra-utérine, par laparotomie
JJC002	Expression tubaire pour évacuation tubo abdominale de grossesse extra-utérine, par coelioscopie
JLJ001	Injection intra ovulaire d'agent pharmacologique pour grossesse extra-utérine, par voie transvaginale avec guidage échographique
JPA001	Salpingotomie avec aspiration de grossesse extra-utérine, par laparotomie
JPC001	Salpingotomie avec aspiration de grossesse extra-utérine, par coelioscopie
JQA001	Extraction de grossesse extra-utérine abdominale au-delà de 13 semaines d'aménorrhée, par laparotomie
Autres types de grossesses	
<u>Diagnostics d'hospitalisation*</u>	
O01	Môle hydatiforme
O02	Autres produits anormaux de la conception

*Les codes CIM-10 ont été sélectionnés à partir des diagnostics principaux, associés ou reliés.

Annexe A.2 : Liste des codes CIM10 utilisés pour le repérage des pathologies observées au cours de la grossesse

Pathologie	Code	Libellé
Grossesse multiple	Z372	Naissance gémellaire, jumeaux nés vivants
	Z373	Naissance gémellaire, l'un des jumeaux né vivant, l'autre mort-né
	Z374	Naissance gémellaire, jumeaux mort-nés
	Z375	Autres naissances multiples, tous nés vivants
	Z376	Autres naissances multiples, certains enfants nés vivants
	Z377	Autres naissances multiples, tous mort-nés
	O84	Accouchements multiples
	O30	Grossesse multiple
Diabète gestationnel	O244	Diabète sucré survenant au cours de la grossesse
Oligoamnios	O410	Oligoamnios
	P012	Fœtus et nouveau-né affectés par un oligoamnios
Grossesse extra-utérine	O00	Grossesse extra-utérine
	O367	Soins maternels pour grossesse abdominale avec fœtus viable
	P014	Fœtus et nouveau-né affectés par une grossesse extra-utérine
Hydramnios	O40	Hydramnios
	P013	Fœtus et nouveau-né affectés par un hydramnios
RPM	O42	Rupture prématurée des membranes
Menace d'accouchement prématuré	O600	Travail prématuré sans accouchement
	O601	Travail prématuré spontané avec accouchement prématuré
	O602	Travail prématuré spontané avec accouchement à terme
	O609	Travail prématuré avec accouchement à terme
	O470	Faux travail avant 37 semaines entières de gestation [MAP]
	O200	Menace d'avortement
Restriction de croissance intra-utérine	P05	Retard de croissance et malnutrition du fœtus
	Z364	Dépistage prénatal de retard de croissance du foetus par échographie et autres méthodes physiques
	O365	Soins maternels pour croissance insuffisante du foetus
Macrosomie	O366	Soins maternels pour croissance excessive du fœtus
	P081	Autres enfants gros pour l'âge gestationnel [Macrosomie foetale]
	O654	Dystocie due à une disproportion foeto-pelvienne, sans précision
Pathologies hypertensives	O11	Prééclampsie surajoutée à une hypertension chronique
	O13	Hypertension gestationnelle [liée à la grossesse]
	O14	Prééclampsie
	O15	Éclampsie

	O16	Hypertension de la mère, sans précision
Pathologies hémorragiques	O20	Hémorragie du début de la grossesse
	H113	Hémorragie conjonctivale
	H313	Hémorragie et rupture de la choroïde
	H356	Hémorragie rétinienne
	H431	Hémorragie du corps vitré
	H450	Hémorragie du corps vitré au cours de maladies classées ailleurs
	I60	Hémorragie sous-arachnoïdienne
	I61	Hémorragie intracérébrale
	I62	Autres hémorragies intracrâniennes non traumatiques
	K920	Hématémèse
	K921	Mélaena
	K922	Hémorragie gastro-intestinale, sans précision
	N92	Ménorragie, polyménorrhée et métrorragie
	N93	Autres saignements anormaux de l'utérus et du vagin
	O071	Échec d'une tentative d'avortement médical, compliqué d'une hémorragie retardée ou sévère
	O076	Échec d'une tentative d'avortement, autres et sans précision, compliqués d'une hémorragie retardée ou sévère
	O081	Hémorragie retardée ou sévère consécutive à un avortement, une grossesse extra-utérine et molaire
	O450	Décollement prématuré du placenta avec anomalie de la coagulation
	O46	Hémorragie précédant l'accouchement, non classée ailleurs
	O67	Travail et accouchement compliqués d'une hémorragie, non classée ailleurs
	O72	Hémorragie du post-partum
	P021	Foetus et nouveau-né affectés par d'autres formes de décollement et d'hémorragie placentaires
	R04	Hémorragie des voies respiratoires
R58	Hémorragie, non classée ailleurs	
D65	Coagulation intravasculaire disséminée [syndrome de défibrination]	
Hyperemesis gravidarum	O21	Vomissements incoercibles au cours de la grossesse
	R11	Nausées et vomissements
Cholestase gravidique	K710	Maladie toxique du foie avec cholestase
	O266	Affections hépatiques au cours de la grossesse, de l'accouchement et de la puerpéralité
Insuffisance rénale aiguë	N17	Insuffisance rénale aiguë
	O084	Insuffisance rénale consécutive à un avortement, une grossesse extra-utérine et molaire
	O904	Insuffisance rénale aiguë du post-partum
	I120	Néphropathie hypertensive, avec insuffisance rénale
	I131	Cardionéphropathie hypertensive, avec insuffisance rénale
	K711	Maladie toxique du foie avec nécrose hépatique

Insuffisance hépatique aiguë	K712	Maladie toxique du foie avec hépatite aiguë
	K720	Insuffisance hépatique aiguë et subaiguë
Pancréatite aiguë	K85	Pancréatite aiguë
Maladies thrombotiques	O22	Complications veineuses et hémorroïdes au cours de la grossesse
	I63	Infarctus cérébral
	I26	Embolie pulmonaire
	I24	Autres cardiopathies ischémiques aiguës
	I74	Embolie et thrombose artérielles
	I80	Phlébite et thrombophlébite
	I81	Thrombose de la veine porte
	I82	Autres embolies et thromboses veineuses
	O222	Thrombophlébite au cours de la grossesse
	O223	Phlébothrombose profonde au cours de la grossesse
	O225	Thrombose veineuse cérébrale au cours de la grossesse
	O870	Thrombophlébite superficielle au cours de la puerpéralité
	O871	Phlébothrombose profonde au cours de la puerpéralité
	O873	Thrombose veineuse cérébrale au cours de la puerpéralité
	O882	Embolie obstétricale
	G08	Phlébite et thrombophlébite intracrâniennes et intrarachidiennes
	K763	Infarctus hépatique
	N280	Ischémie et infarctus du rein
	D735	Infarctus de la rate
I21	Infarctus aigu du myocarde	
Mort fœtale in-utero	O364	Soins maternels pour mort intra-utérine du fœtus
	O021	Rétention d'un fœtus mort in utero
	P95	Mort fœtale
	O312	Poursuite de la grossesse après mort intra-utérine d'un ou plusieurs fœtus
Colique néphrétique	N20	Calcul du rein et de l'uretère
	N21	Calcul des voies urinaires inférieures
	N22	Calcul des voies urinaires au cours de maladies classées ailleurs
	N23	Colique néphrétique, sans précision
Fausse couche spontanée	O03	Avortement spontané
	O311	Poursuite de la grossesse après avortement d'un ou plusieurs fœtus
Infection	F024	Démence de la maladie due au virus de l'immunodéficience humaine [VIH] (B22.0)
	D762	Syndrome hémophagocytaire associé à une infection
	D823	Déficit immunitaire avec réponse héréditaire anormale au virus d'Epstein-Barr
	E033	Hypothyroïdie post-infectieuse

I301	Péricardite infectieuse
I521	Autres cardiopathies au cours d'autres maladies infectieuses et parasitaires classées ailleurs
I321	Péricardite au cours d'autres maladies infectieuses et parasitaires classées ailleurs
I330	Endocardite infectieuse (aiguë et subaiguë)
I400	Myocardite infectieuse
I412	Myocardite au cours d'autres maladies infectieuses et parasitaires classées ailleurs
I430	Myocardiopathie au cours de maladies infectieuses et parasitaires classées ailleurs
H320	Affections chorio-rétiniennes au cours de maladies infectieuses et parasitaires classées ailleurs
H3200	Rétinite à cytomégalovirus (B25.8)
H3208	Affections chorio-rétiniennes au cours d'autres maladies infectieuses et parasitaires classées ailleurs
H603	Autres otites externes infectieuses
H605	Otite externe aiguë, (non infectieuse)
H611	Affections non infectieuses du pavillon de l'oreille
G940	Hydrocéphalie au cours de maladies infectieuses et parasitaires classées ailleurs (A00-B99)
H011	Dermatoses non infectieuses de la paupière
H031	Atteinte de la paupière au cours d'autres maladies infectieuses classées ailleurs
H131	Conjonctivite au cours de maladies infectieuses et parasitaires classées ailleurs
H191	Kératite et kératoconjonctivite dues au virus de l'herpès (B00.5)
H192	Kératite et kératoconjonctivite au cours d'autres maladies infectieuses et parasitaires classées ailleurs
H220	Iridocyclite au cours de maladies infectieuses et parasitaires classées ailleurs
G02	Méningite au cours d'autres maladies infectieuses et parasitaires classées ailleurs
G020	Méningite au cours d'infections virales classées ailleurs
G028	Méningite au cours d'autres maladies infectieuses et parasitaires précisées classées ailleurs
G050	Encéphalite, myélite et encéphalomyélite au cours d'infections bactériennes classées ailleurs
G051	Encéphalite, myélite et encéphalomyélite au cours d'infections virales classées ailleurs
G052	Encéphalite, myélite et encéphalomyélite au cours d'autres maladies infectieuses et parasitaires classées ailleurs
G531	Paralysie de plusieurs nerfs crâniens au cours de maladies infectieuses et parasitaires classées ailleurs (A00-B99)
H623	Otite externe au cours d'autres maladies infectieuses et parasitaires classées ailleurs

H750	Mastoïdite au cours de maladies infectieuses et parasitaires classées ailleurs
H940	Névrite acoustique au cours de maladies infectieuses et parasitaires classées ailleurs
B97	Virus, cause de maladies classées dans d'autres chapitres
B970	Adénovirus, cause de maladies classées dans d'autres chapitres
B971	Entérovirus, cause de maladies classées dans d'autres chapitres
B972	Coronavirus, cause de maladies classées dans d'autres chapitres
B973	Rétrovirus, cause de maladies classées dans d'autres chapitres
B974	Virus respiratoire syncytial, cause de maladies classées dans d'autres chapitres
B975	Réovirus, cause de maladies classées dans d'autres chapitres
B976	Parvovirus, cause de maladies classées dans d'autres chapitres
B977	Papillovirus, cause de maladies classées dans d'autres chapitres
B978	Autres virus, cause de maladies classées dans d'autres chapitres
B98	Autres agents infectieux précisés, cause de maladies classées dans d'autres chapitres
B99	Maladies infectieuses, autres et non précisées
G630	Polynévrite au cours de maladies infectieuses et parasitaires classées ailleurs
G734	Myopathie au cours de maladies infectieuses et parasitaires classées ailleurs
B71	Autres infections à cestodes
B718	Autres infections par cestodes, précisées
B94	Séquelles de maladies infectieuses et parasitaires, autres et non précisées
B948	Séquelles d'autres maladies infectieuses et parasitaires précisées
B949	Séquelles de maladie infectieuse ou parasitaire, sans précision
B963	Haemophilus influenzae, cause de maladies classées dans d'autres chapitres
A492	Infection à Haemophilus influenzae, siège non précisé
A498	Autres infections bactériennes, siège non précisé
A542	Pelvipéritonite gonococcique et autres infections génito-urinaires gonococciques
A548	Autres infections gonococciques
A56	Autres infections à Chlamydia transmises par voie sexuelle
A60	Infection ano-génitale par le virus de l'herpès [herpes simplex]
A600	Infection des organes génitaux et de l'appareil génito-urinaire par le virus de l'herpès
A601	Infection de la marge cutanée de l'anus et du rectum, par le virus de l'herpès
A609	Infection ano-génitale par le virus de l'herpès, sans précision
A69	Autres infections à spirochètes
A691	Autres infections dites de Vincent
A698	Autres infections à spirochètes précisées
A74	Autres infections à Chlamydia

A748	Autres infections à Chlamydia précisées
A800	Poliomyélite paralytique aiguë, associée au virus vaccinal
A801	Poliomyélite paralytique aiguë, virus sauvage importé
A802	Poliomyélite paralytique aiguë, virus sauvage indigène
A81	Infections atypiques à virus, du système nerveux central
A818	Autres infections atypiques à virus, du système nerveux central
A819	Infection atypique à virus, du système nerveux central, sans précision
A836	Maladie à virus de Rocio
A850	Encéphalite à entérovirus (G05.1*)
A851	Encéphalite à adénovirus (G05.1*)
A870	Méningite à entérovirus (G02.0*)
A871	Méningite à adénovirus (G02.0)
A88	Autres infections virales du système nerveux central, non classées ailleurs
A880	Fièvre exanthématique à entérovirus [exanthème de Boston]
A888	Autres infections virales précisées du système nerveux central
A91	Fièvre hémorragique due au virus de la dengue
A921	Fièvre due au virus O'nyong-nyong
A923	Fièvre due au virus West-Nile
A96	Fièvre hémorragique à arénavirus
A968	Autres fièvres hémorragiques à arénavirus
A969	Fièvre hémorragique à arénavirus, sans précision
A984	Maladie à virus Ebola
B00	Infections par le virus de l'herpès [herpes simplex]
B001	Dermite vésiculaire due au virus de l'herpès
B002	Gingivo-stomatite et pharyngo-amygdalite due au virus de l'herpès
B003	Méningite due au virus de l'herpès (G02.0*)
B004	Encéphalite due au virus de l'herpès (G05.1*)
B005	Affections oculaires dues au virus de l'herpès
B007	Maladie disséminée due au virus de l'herpès
B008	Autres formes d'infection due au virus de l'herpès
B009	Infection due au virus de l'herpès, sans précision
B08	Autres infections virales caractérisées par des lésions cutanéomuqueuses, non classées ailleurs
B080	Autres infections à orthopoxvirus
B083	Érythème infectieux [cinquième maladie] [mégalythème épiémique]
B084	Stomatite vésiculaire avec exanthème, due à un entérovirus
B085	Pharyngite vésiculaire due à un entérovirus
B088	Autres infections virales précisées, caractérisées par des lésions cutanéomuqueuses
B170	(Sur)infection aiguë par agent delta d'un sujet porteur de l'hépatite B

B20	Immunodéficience humaine virale [VIH], à l'origine de maladies infectieuses et parasitaires
B200	Maladie par VIH à l'origine d'une infection mycobactérienne
B201	Maladie par VIH à l'origine d'autres infections bactériennes
B202	Maladie par VIH à l'origine d'infections à cytomégalovirus
B203	Maladie par VIH à l'origine d'autres infections virales
B207	Maladie par VIH à l'origine d'infections multiples
B208	Maladie par VIH à l'origine d'autres maladies infectieuses et parasitaires
B209	Maladie par VIH à l'origine d'une maladie infectieuse ou parasitaire non précisée
B230	Syndrome d'infection aiguë par VIH
B25	Maladie à cytomégalovirus
B250	Pneumopathie à cytomégalovirus (J17.1*)
B251	Hépatite à cytomégalovirus (K77.0*)
B252	Pancréatite à cytomégalovirus (K87.1*)
B258	Autres maladies à cytomégalovirus
B259	Maladie à cytomégalovirus, sans précision
B27	Mononucléose infectieuse
B270	Mononucléose due à herpèsvirus gamma
B271	Mononucléose à cytomégalovirus
B278	Autres mononucléoses infectieuses
B279	Mononucléose infectieuse, sans précision
B300	Kératoconjonctivite due à un adénovirus (H19.2*)
B301	Conjonctivite due à un adénovirus (H13.1*)
B33	Autres maladies à virus, non classées ailleurs
B333	Infections à rétrovirus, non classées ailleurs
B334	Syndrome (cardio-)pulmonaire à virus Hanta [hantavirus] [SPH] [SCPH] (J17.1*)
B338	Autres maladies à virus précisées
B348	Autres infections virales, siège non précisé
A028	Autres infections précisées à Salmonella
A04	Autres infections intestinales bactériennes
A044	Autres infections intestinales à Escherichia coli
A048	Autres infections intestinales bactériennes précisées
A068	Autres localisations d'une infection amibienne
A08	Infections virales intestinales et autres infections intestinales précisées
A080	Entérite à rotavirus
A082	Entérite à adénovirus
A085	Autres infections intestinales précisées
A09	Autres gastroentérites et colites d'origine infectieuse et non précisée
A090	Gastroentérites et colites d'origine infectieuse, autres et non précisées

A157	Primo-infection tuberculeuse de l'appareil respiratoire, avec confirmation bactériologique et histologique
A167	Primo-infection tuberculeuse de l'appareil respiratoire, (sans mention de confirmation bactériologique ou histologique)
A318	Autres infections à mycobactéries atypiques
A398	Autres infections à méningocoques
A413	Sepsis à Haemophilus influenzae
I681	Artérite cérébrale au cours de maladies infectieuses et parasitaires
I89	Autres atteintes non infectieuses des vaisseaux et des ganglions lymphatiques
I898	Autres atteintes non infectieuses précisées des vaisseaux et des ganglions lymphatiques
I899	Atteinte non infectieuse des vaisseaux et des ganglions lymphatiques, sans précision
I981	Troubles cardio-vasculaires au cours d'autres maladies infectieuses et parasitaires classées ailleurs
J068	Autres infections aiguës des voies respiratoires supérieures, à localisations multiples
J09	Grippe, à virus aviaire identifié
J10	Grippe, à autre virus grippal identifié
J100	Grippe avec pneumopathie, autre virus grippal identifié
J101	Grippe avec d'autres manifestations respiratoires, autre virus grippal identifié
J108	Grippe avec d'autres manifestations, autre virus grippal identifié
J11	Grippe, virus non identifié
J110	Grippe avec pneumopathie, virus non identifié
J111	Grippe avec d'autres manifestations respiratoires, virus non identifié
J118	Grippe avec d'autres manifestations, virus non identifié
J121	Pneumopathie due au virus respiratoire syncytial [VRS]
J122	Pneumopathie due au virus paragrippaux
J123	Pneumonie due à métapneumovirus humain
J13	Pneumonie due à Streptococcus pneumoniae
J14	Pneumopathie due à Haemophilus influenzae
B66	Autres infections par douves [distomatoses]
B668	Autres infections par douves précisées
J680	Bronchite et pneumopathie dues à des agents chimiques, des émanations, des fumées et des gaz
M018	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs
M0180	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Sièges multiples
M0181	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Articulations acromio-claviculaire, scapulo-humérale, et sterno-claviculaire

M0182	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Articulation du coude
M0183	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Articulation du poignet
M0184	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Articulations de la main
M0185	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Articulations de la hanche et sacro-iliaque
M0186	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Articulation du genou
M0187	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Articulations de la cheville et du pied
M0188	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Autres articulations
M0189	Arthrite au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Siège non précisé
M01	Arthrites infectieuses directes au cours de maladies infectieuses et parasitaires classées ailleurs
K52	Autres gastro-entérites et colites non infectieuses
K528	Autres gastro-entérites et colites non infectieuses précisées
K529	Gastro-entérite et colite non infectieuses, sans précision
K67	Atteintes péritonéales au cours de maladies infectieuses classées ailleurs
K678	Autres atteintes péritonéales au cours de maladies infectieuses classées ailleurs
J16	Pneumopathie due à d'autres micro-organismes infectieux, non classée ailleurs
J168	Pneumopathie due à d'autres micro-organismes infectieux
J20	Bronchite aiguë
J200	Bronchite aiguë due à Mycoplasma pneumoniae
J201	Bronchite aiguë due à Haemophilus influenzae
J202	Bronchite aiguë due à des streptocoques
J203	Bronchite aiguë due au virus Coxsackie
J204	Bronchite aiguë due aux virus paragrrippaux
J205	Bronchite aiguë due au virus respiratoire syncytial [VRS]
J206	Bronchite aiguë due à des rhinovirus
J207	Bronchite aiguë due à des virus ECHO
J208	Bronchite aiguë due à d'autres micro-organismes précisés
J209	Bronchite aiguë, sans précision
J210	Bronchiolite aiguë due au virus respiratoire syncytial [VRS]
J211	Bronchiolite aiguë due à métapneumovirus humain
J40	Bronchite, (non précisée comme aiguë ou chronique)
J41	Bronchite chronique simple et mucopurulente
J410	Bronchite chronique simple
J411	Bronchite chronique mucopurulente
J418	Bronchite chronique simple et mucopurulente
J42	Bronchite chronique, sans précision

J440	Maladie pulmonaire obstructive chronique avec infection aiguë des voies respiratoires inférieures
K93820	Colite à cytomégalovirus (B25.8)
K93830	Infection du rectum par le virus de l'herpès (A60.1)
K93840	Infection de l'anus par le virus de l'herpès (A60.1)
L08	Autres infections localisées de la peau et du tissu cellulaire sous-cutané
L088	Autres infections localisées précisées de la peau et du tissu cellulaire sous-cutané
L303	Dermite infectée
M03	Arthropathies post-infectieuses et réactionnelles au cours de maladies classées ailleurs
M031	Arthropathie syphilitique post-infectieuse
M0310	Arthropathie syphilitique post-infectieuse - Sièges multiples
M0311	Arthropathie syphilitique post-infectieuse - Articulations acromio-claviculaire, scapulo-humérale, et sterno-claviculaire
M0312	Arthropathie syphilitique post-infectieuse - Articulation du coude
M0313	Arthropathie syphilitique post-infectieuse - Articulation du poignet
M0314	Arthropathie syphilitique post-infectieuse - Articulations de la main
M0315	Arthropathie syphilitique post-infectieuse - Articulations de la hanche et sacro-iliaque
M0316	Arthropathie syphilitique post-infectieuse - Articulation du genou
M0317	Arthropathie syphilitique post-infectieuse - Articulations de la cheville et du pied
M0318	Arthropathie syphilitique post-infectieuse - Autres articulations
M0319	Arthropathie syphilitique post-infectieuse - Siège non précisé
M032	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs
M0320	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Sièges multiples
M0321	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Articulations acromio-claviculaire, scapulo-humérale, et sterno-claviculaire
M0322	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Articulation du coude
M0323	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Articulation du poignet
M0324	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Articulations de la main
M0325	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Articulations de la hanche et sacro-iliaque
M0326	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Articulation du genou

M0327	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Articulations de la cheville et du pied
M0328	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Autres articulations
M0329	Autres arthropathies post-infectieuses au cours de maladies classées ailleurs - Siège non précisé
K770	Atteintes hépatiques au cours de maladies infectieuses et parasitaires classées ailleurs
K8700	Cholangite à cytomégalovirus (B25.8)
M4939	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Localisation vertébrale non précisée
M600	Myosite infectieuse
M6000	Myosite infectieuse - Sièges multiples
M6001	Myosite infectieuse - Région scapulaire
M6002	Myosite infectieuse - Bras
M6003	Myosite infectieuse - Avant-bras
M6004	Myosite infectieuse - Main
M6005	Myosite infectieuse - Région pelvienne et cuisse
M6006	Myosite infectieuse - Jambe
M6007	Myosite infectieuse - Cheville et pied
M6008	Myosite infectieuse - Autres localisations
M6009	Myosite infectieuse - Siège non précisé
M493	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs
M4930	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Localisations vertébrales multiples
M4931	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Région occipito-atloïdo-axoïdienne
M4932	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Région cervicale
M4933	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Région cervico-dorsale
M4934	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Région dorsale
M4935	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Région dorso-lombaire
M4936	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Région lombaire
M4937	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Région lombo-sacrée
M4938	Spondylopathie au cours d'autres maladies infectieuses et parasitaires classées ailleurs - Région sacrée et sacro-coccygienne
M711	Autres bursites infectieuses
M7110	Autres bursites infectieuses - Sièges multiples

M7111	Autres bursites infectieuses - Région scapulaire
M7112	Autres bursites infectieuses - Bras
M7113	Autres bursites infectieuses - Avant-bras
M7114	Autres bursites infectieuses - Main
M7115	Autres bursites infectieuses - Région pelvienne et cuisse
M7116	Autres bursites infectieuses - Jambe
M7117	Autres bursites infectieuses - Cheville et pied
M7118	Autres bursites infectieuses - Autres localisations
M7119	Autres bursites infectieuses - Siège non précisé
M465	Autres spondylopathies infectieuses
M4650	Autres spondylopathies infectieuses - Localisations vertébrales multiples
M4651	Autres spondylopathies infectieuses - Région occipito-atloïdo-axoïdienne
M4652	Autres spondylopathies infectieuses - Région cervicale
M4653	Autres spondylopathies infectieuses - Région cervico-dorsale
M4654	Autres spondylopathies infectieuses - Région dorsale
M4655	Autres spondylopathies infectieuses - Région dorso-lombaire
M4656	Autres spondylopathies infectieuses - Région lombaire
M4657	Autres spondylopathies infectieuses - Région lombo-sacrée
M4658	Autres spondylopathies infectieuses - Région sacrée et sacro-coccygienne
M4659	Autres spondylopathies infectieuses - Localisation vertébrale non précisée
M631	Myosite au cours d'infections parasitaires et à protozoaires classées ailleurs
M632	Myosite au cours d'autres maladies infectieuses classées ailleurs
M651	Autres (téno)synovites infectieuses
M6510	Autres (téno)synovites infectieuses - Sièges multiples
M6511	Autres (téno)synovites infectieuses - Région scapulaire
M6512	Autres (téno)synovites infectieuses - Bras
M6513	Autres (téno)synovites infectieuses - Avant-bras
M6514	Autres (téno)synovites infectieuses - Main
M6515	Autres (téno)synovites infectieuses - Région pelvienne et cuisse
M6516	Autres (téno)synovites infectieuses - Jambe
M6517	Autres (téno)synovites infectieuses - Cheville et pied
M6518	Autres (téno)synovites infectieuses - Autres localisations
M6519	Autres (téno)synovites infectieuses - Siège non précisé
N080	Glomérulopathie au cours de maladies infectieuses et parasitaires classées ailleurs
N110	Pyélonéphrite (non obstructive chronique) associée à un reflux
N111	Pyélonéphrite obstructive (chronique)
N431	Hydrocèle infectée
N160	Maladie rénale tubulo-interstitielle au cours de maladies infectieuses et parasitaires classées ailleurs

N291	Autres affections du rein et de l'uretère au cours de maladies infectieuses et parasitaires classées ailleurs
N351	Rétrécissement urétral post-infectieux, non classé ailleurs
O753	Autres infections au cours du travail
O86	Autres infections puerpérales
O861	Autres infections des voies génitales, après accouchement
O863	Autres infections des voies génito-urinaires, après accouchement
O868	Autres infections puerpérales précisées
N770	Ulcération de la vulve au cours de maladies infectieuses et parasitaires classées ailleurs
N771	Vaginite, vulvite et vulvo-vaginite au cours de maladies infectieuses et parasitaires classées ailleurs
M901	Périostite au cours d'autres maladies infectieuses classées ailleurs
M9010	Périostite au cours d'autres maladies infectieuses classées ailleurs - Sièges multiples
M9011	Périostite au cours d'autres maladies infectieuses classées ailleurs - Région scapulaire
M9012	Périostite au cours d'autres maladies infectieuses classées ailleurs - Bras
M9013	Périostite au cours d'autres maladies infectieuses classées ailleurs - Avant-bras
M9014	Périostite au cours d'autres maladies infectieuses classées ailleurs - Main
M9015	Périostite au cours d'autres maladies infectieuses classées ailleurs - Région pelvienne et cuisse
M9016	Périostite au cours d'autres maladies infectieuses classées ailleurs - Jambe
M9017	Périostite au cours d'autres maladies infectieuses classées ailleurs - Cheville et pied
M9018	Périostite au cours d'autres maladies infectieuses classées ailleurs - Autres localisations
M9019	Périostite au cours d'autres maladies infectieuses classées ailleurs - Siège non précisé
M902	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs
M9020	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Sièges multiples
M9021	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Région scapulaire
M9022	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Bras
M9023	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Avant-bras
M9024	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Main

M9025	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Région pelvienne et cuisse
M9026	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Jambe
M9027	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Cheville et pied
M9028	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Autres localisations
M9029	Ostéopathie au cours d'autres maladies infectieuses classées ailleurs - Siège non précisé
P783	Diarrhée (non infectieuse) néonatale
O030	Avortement spontané incomplet, compliqué d'une infection de l'appareil génital et des organes pelviens
O035	Avortement spontané complet ou sans précision, compliqué d'une infection de l'appareil génital et des organes pelviens
O040	Avortement médical incomplet, compliqué d'une infection de l'appareil génital et des organes pelviens
O045	Avortement médical complet ou sans précision, compliqué d'une infection de l'appareil génital et des organes pelviens
O050	Autres formes d'avortement incomplet, compliqué d'une infection de l'appareil génital et des organes pelviens
O055	Autres formes d'avortement complet ou sans précision, compliqué d'une infection de l'appareil génital et des organes pelviens
O060	Avortement sans précision, incomplet, compliqué d'une infection de l'appareil génital et des organes pelviens
O065	Avortement sans précision, complet ou sans précision, compliqué d'une infection de l'appareil génital et des organes pelviens
O070	Échec d'une tentative d'avortement médical, compliqué d'une infection de l'appareil génital et des organes pelviens
O075	Échec d'une tentative d'avortement, autres et sans précision, compliqués d'une infection de l'appareil génital et des organes pelviens
P351	Infection congénitale à cytomégalovirus
P368	Autres infections bactériennes du nouveau-né
P37	Autres maladies infectieuses et parasitaires congénitales
P378	Autres maladies infectieuses et parasitaires congénitales précisées
P379	Maladie infectieuse ou parasitaire congénitale, sans précision
P39	Autres infections spécifiques de la période périnatale
P390	Mastite infectieuse néonatale
P398	Autres infections spécifiques précisées de la période périnatale
P582	Ictère néonatal dû à une infection
R650	Syndrome de réponse inflammatoire systémique d'origine infectieuse sans défaillance d'organe

R651	Syndrome de réponse inflammatoire systémique d'origine infectieuse avec défaillance d'organe
R652	Syndrome de réponse inflammatoire systémique d'origine non infectieuse sans défaillance d'organe
R653	Syndrome de réponse inflammatoire systémique d'origine non infectieuse avec défaillance d'organe
R75	Mise en évidence par des examens de laboratoire du virus de l'immunodéficience humaine [VIH]
T37	Intoxication par d'autres anti-infectieux et antiparasitaires systémiques
T378	Intoxication par autres anti-infectieux et antiparasitaires systémiques précisés
T379	Intoxication par anti-infectieux et anti-parasitaire systémiques, sans précision
O987	Maladie due au virus de l'immunodéficience humaine [VIH], compliquant la grossesse, l'accouchement et la puerpéralité
O988	Autres maladies infectieuses et parasitaires de la mère compliquant la grossesse, l'accouchement et la puerpéralité
O989	Maladie infectieuse ou parasitaire de la mère compliquant la grossesse, l'accouchement et la puerpéralité, sans précision
P002	Foetus et nouveau-né affectés par des maladies infectieuses et parasitaires de la mère
P027	Foetus et nouveau-né affectés par une chorio-amniotite
Z717	Conseil à propos du virus de l'immunodéficience humaine [VIH]
Z830	Antécédents familiaux de maladie due au virus de l'immunodéficience humaine [VIH]
Z831	Antécédents familiaux d'autres maladies infectieuses et parasitaires
Z861	Antécédents personnels de maladies infectieuses et parasitaires
Z883	Antécédents personnels d'allergie à d'autres agents anti-infectieux
Z11	Examen spécial de dépistage de maladies infectieuses et parasitaires
Z110	Examen spécial de dépistage de maladies infectieuses intestinales
Z113	Examen spécial de dépistage des infections dont le mode de transmission est essentiellement sexuel
Z114	Examen spécial de dépistage du virus de l'immunodéficience humaine [VIH]
Z115	Examen spécial de dépistage d'autres maladies à virus
Z118	Examen spécial de dépistage d'autres maladies infectieuses et parasitaires
Z119	Examen spécial de dépistage de maladies infectieuses et parasitaires, sans précision
Z200	Sujets en contact avec et exposés à des maladies infectieuses intestinales

Z202	Sujets en contact avec et exposés à des infections dont le mode de transmission est essentiellement sexuel
Z206	Sujets en contact avec et exposés au virus de l'immunodéficience humaine [VIH]
Z21	Infection asymptomatique par le virus de l'immunodéficience humaine [VIH]
Z22	Sujet porteur de germes responsables d'une maladie infectieuse
Z221	Sujet porteur d'autres maladies intestinales infectieuses
Z226	Sujet porteur d'une infection à virus lymphotrope des lymphocytes T humains de type 1 [HTLV-1]
Z228	Sujet porteur d'autres maladies infectieuses
Z229	Sujet porteur d'une maladie infectieuse, sans précision
Y41	Effets indésirables des autres anti-infectieux et antiparasitaires systémiques
Y418	Effets indésirables d'autres anti-infectieux et antiparasitaires systémiques précisés au cours de leur usage thérapeutique
Y419	Effets indésirables d'un anti-infectieux et antiparasitaire systémiques, sans précision au cours de son usage thérapeutique
Z26	Nécessité d'une vaccination contre d'autres maladies infectieuses uniques
Z268	Nécessité d'une vaccination contre d'autres maladies infectieuses uniques précisées
Z269	Nécessité d'une vaccination (contre une maladie infectieuse, sans précision)
Z27	Nécessité d'une vaccination associée contre plusieurs maladies infectieuses
Z278	Nécessité d'une vaccination contre d'autres associations de maladies infectieuses
Z279	Nécessité d'une vaccination associée contre plusieurs maladies infectieuses, sans précision
O98	Maladies infectieuses et parasitaires de la mère classées ailleurs mais compliquant la grossesse, l'accouchement et la puerpéralité
Y560	Effets indésirables des médicaments antifongiques, anti-infectieux et anti-inflammatoires pour application locale, non classés ailleurs au cours de leur usage thérapeutique
T490	Intoxication par médicaments antifongiques, anti-infectieux et anti-inflammatoires à usage topique, non classés ailleurs

**ANNEXES B : DOCUMENTS
SUPPLEMENTAIRES DES ANALYSES DU
CHAPITRE 3**

Annexe B.1 : Liste des médicaments

foétotoxiques

Groupe de médicament	Nom de médicament	Code_ATC
Inhibiteurs de l'enzyme de conversion de l'angiotensine (ECA)	captopril	C09AA01, C09BA01
	bénazetril	C09AA07, C09BA07
	énalapril	C09AA02, C09BA02, C09BB02
	fosinopril	C09AA09, C09BA09
	lisinopril	C09AA03, C09BA03
	périndopril	C09AA04, C09BA04, C09BB04, C09BX01, C09BX02
	quinapril	C09AA06, C09BA06
	ramipril	C09AA05, C09BA05
	trandolapril	C09AA10, C09AB10
	zofenopril	C09AA15, C09BA15
Bloqueurs des récepteurs de l'angiotensine II	candésartan	C09CA06, C09DA06
	éprosartan	C09CA02
	irbesartan	C09CA04, C09DA04
	losartan	C09CA01, C09DA01
	olmésartan	C09CA08, C09DA08, C09DB02
	telmisartan	C09CA07, C09DA07, C09DB04
	valsartan	C09CA03, C09DA03, C09DB01, C09DX01, C09DX04
Inhibiteurs de la rénine	aliskiren	C09XA02, C09XA52
Contraceptifs hormonaux à usage systémique	progestatifs et œstrogènes, combinaisons fixes	G03AA01, G03AA02, G03AA03, G03AA04, G03AA05, G03AA06, G03AA07, G03AA08, G03AA09, G03AA10, G03AA11, G03AA12, G03AA13, G03AA14, G03AA15, G03AA16, G03AA17
	progestatifs et œstrogènes, préparation séquentielle	G03AB01, G03AB02, G03AB03, G03AB04, G03AB05, G03AB06, G03AB07, G03AB08, G03AB09
	désogestrel	G03AC09
	étonogestrel	G03AC08
	lévonorgestrel	G03AC03
	médroxyprogestérone	G03AC06
Contraceptifs d'urgence	lévonorgestrel, urgence	G03D01
	ulipristal	3400939450377, 3400930124086

Contraceptifs à usage topique	dispositif intra-utérin en plastique avec progestatif	G02BA03, 3400930094754
	anneau vaginal avec progestatif et œstrogène	G02BB01
Androgènes	androstanolone	G03BB02
	mestérolone	G03BB01
	testostérone	G03BA03
Œstrogènes (usage systémique)	estradiol	G03CA03
	éthinyloestradiol	G03CA01
	tibolone	G03CX01
Œstrogènes (usage topique)	estriol	G03CA04
	promestriène	G03CA09
Progestatifs	chlormadinone	G03DB06
	diénogest	G03DB08
	lynestrénol	G03DC03
	médrogestone	G03DB03
	médroxyprogestérone	G03DA02
	nomégestrol	G03DB04
	noréthistérone	G03DC02
	promégestone	G03DB07
Anti-androgène	cyprotérone	G03HA01
	cyprotérone et œstrogènes	G03HB01
Traitement hormonal substitutif	œstrogène et progestérone avec combinaison fixe	G03FA01, G03FA02, G03FA03, G03FA04, G03FA05, G03FA06, G03FA07, G03FA08, G03FA09, G03FA10, G03FA11, G03FA12, G03FA13, G03FA14, G03FA15, G03FA16, G03FA17
	œstrogène et progestérone avec préparation séquentielle	G03FB01, G03FB02, G03FB03, G03FB04, G03FB05, G03FB06, G03FB07, G03FB08, G03FB09, G03FB10, G03FB11, G03FB12
Gonadotrophines et autres stimulants de l'ovulation	follitropine alfa	G03GA05
	Lutropine alfa	G03GA07
	follitropine bêta	G03GA06
	gonadotrophine chorionique	G03GA01
	gonadotrophine ménopausique humaine	G03GA02
	urofollitropine	G03GA04
	clomifène	G03GB02
Antigonadotrophines et agents similaires	danazol	G03XA01

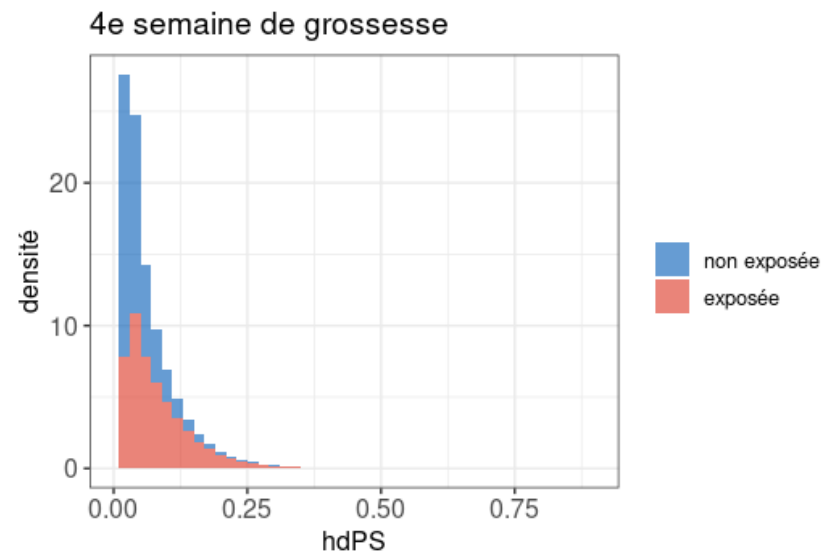
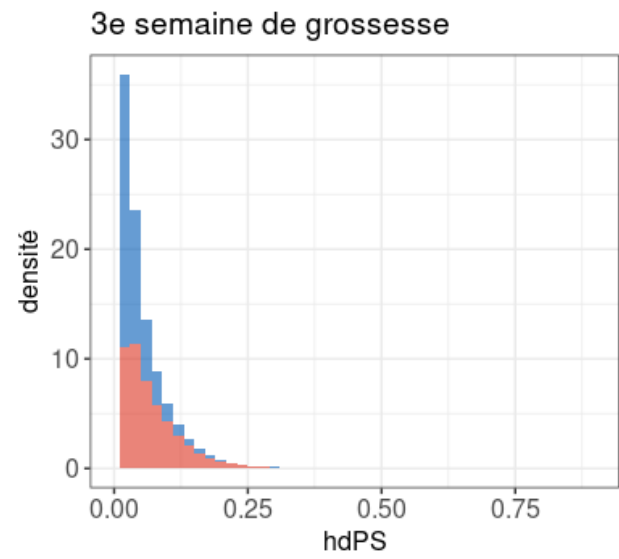
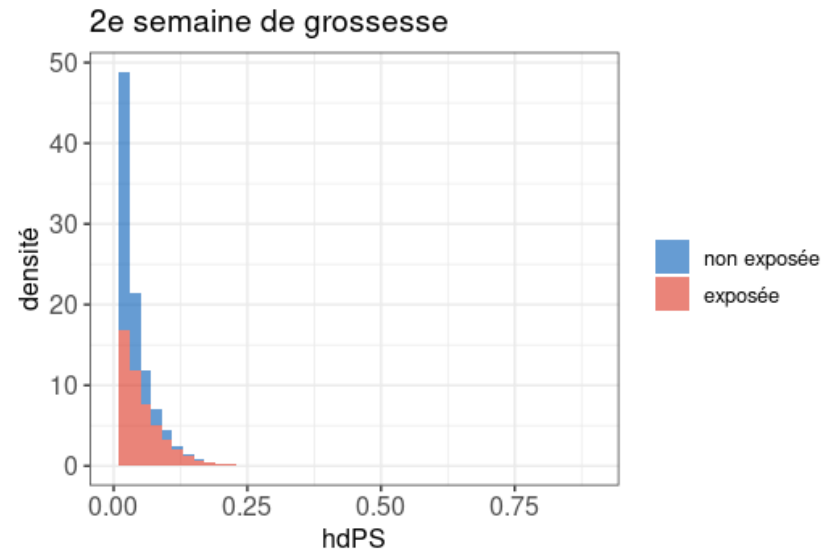
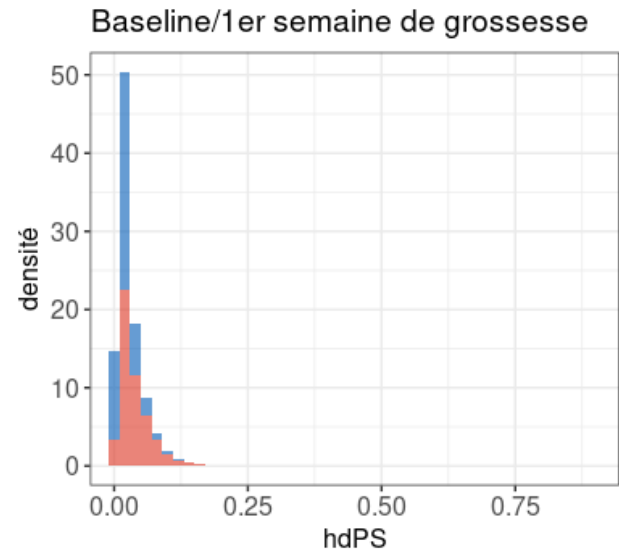
Hormones sexuels pour les maladies systémiques	bazedoxifene	G03XC02
	raloxifène	G03XC01
	ulipristal	3400922257891, 3400927875304

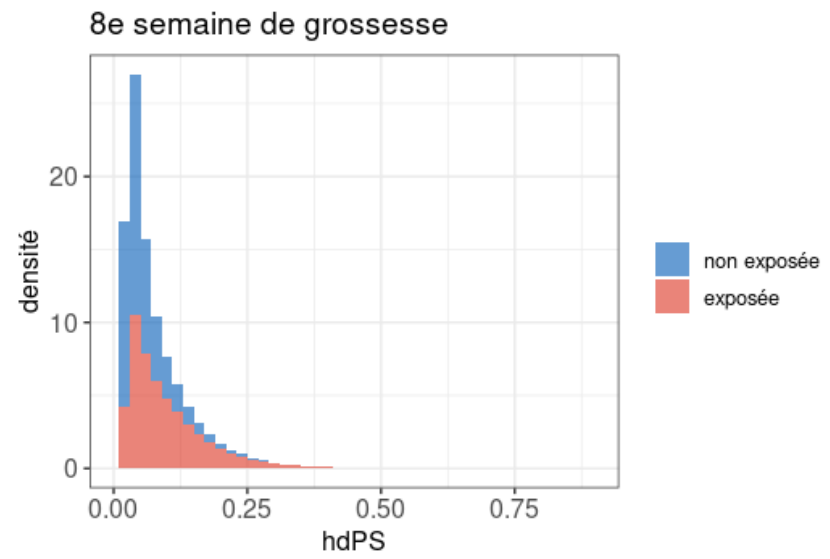
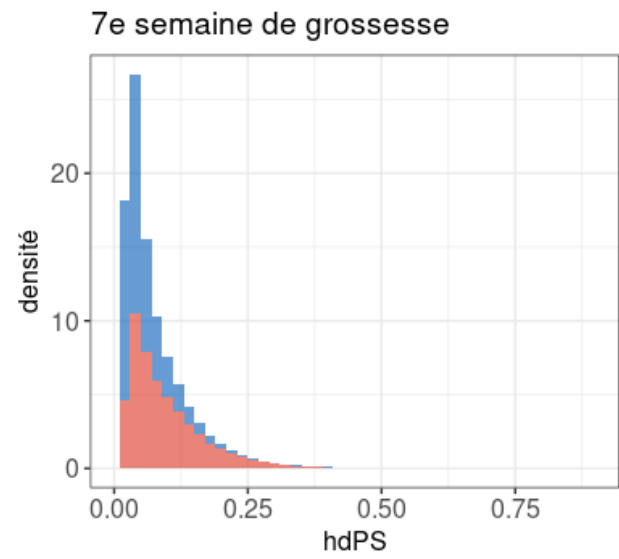
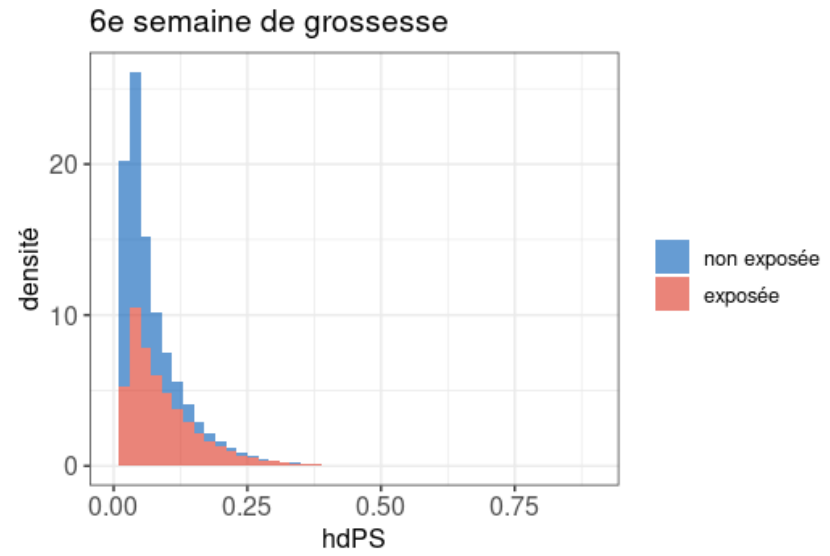
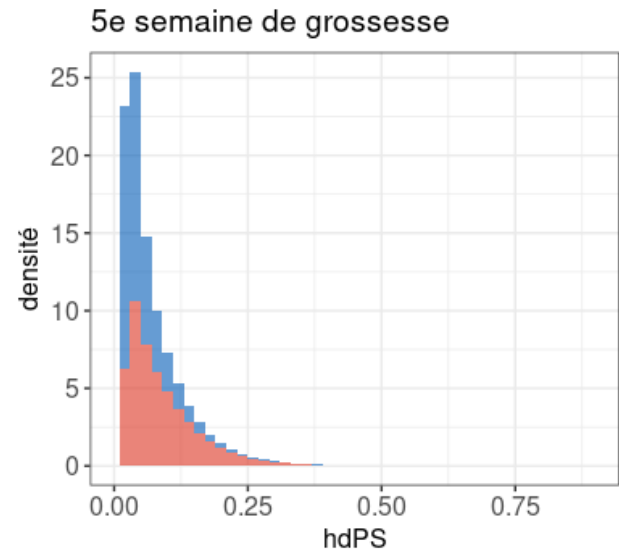
Annexe B.2 : Liste de médicaments tératogènes

Groupe de médicament	Drug_name	Code_ATC
Médicaments antinéoplasiques et immunomodulateurs	fingolimod	L04AA27
	léflunomide	L04AA13
	lénalidomide	L04AX04
	méthotrexate	L04AX03, L01BA01
	acide mycophénolique	L04AA06
	thalidomide	L04AX02
	tériflunomide	L04AA31
Rétinoïdes à usage systémique	acitrétine (pour le traitement du psoriasis)	D05BB02
	agents à base d'alitrétinoïne pour la dermatite	D11AH04
	étrétinate	D05BB01
	isotrétinoïne pour usage systémique	D10BA01
	trétinoïne	L01XX14
Rétinoïdes à usage topique	adapalène	D10AD03, D10AD53
	alitrétinoïne agent antinéoplasique	L01XF02
	isotrétinoïne	D10AD04
	tifarotène	D10AD06
	trétinoïne	D10AD01, D10AD51
Médicaments antiépileptiques	carbamazépine	N03AF01
	fosphénytoïne	N03AB05
	oxcarbazépine	N03AF02
	phénytoïne	N03AB52
	topiramate	N03AX11
	acide valproïque	N03AG01
Médicaments pour les troubles affectifs	valpromide	N03AG02
	lithium	N05AN01
	divalproate	3400934876233, 3400935444271, 3400934876691
Préparations antithyroïdiennes	carbimazole	H03BB01
	thiamazole	H03BB02
Antagonistes de la vitamine K	acénocoumarol	B01AA07
	fluindione	3400933484132, 3400931183280, 3400931183112

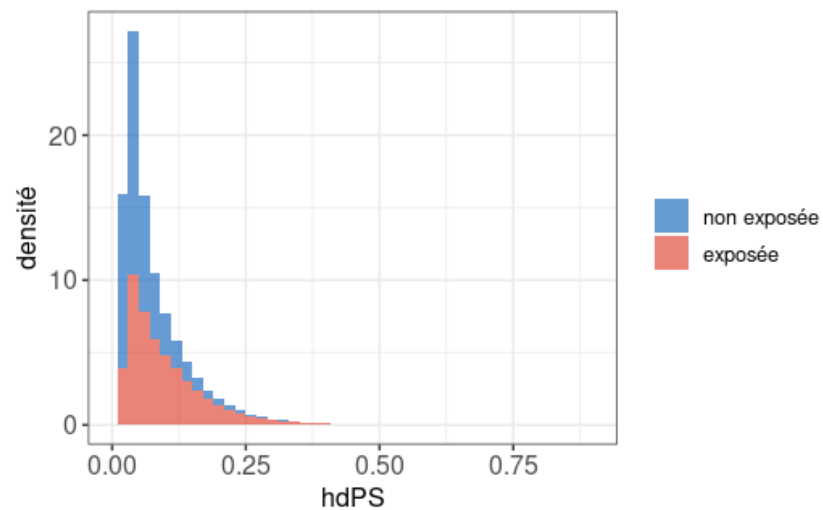
	warfarine	B01AA03
Inhibiteurs de la HMG CoA réductase	atorvastatine	C10AA05, C10BA06, C10BX03, C10BX06, C10BX11, C10BX15
	fluvastatine	C10AA04
	lovastatine	C10AA02
	pitavastatine	C10AA08
	pravastatine	C10AA03, C10BX02
	rosuvastatine	C10AA07, C10BA06
	simvastatine	C10AA01, C10BA02, C10BA04
Autres médicaments agissant comme tératogènes	acetazolamide, antiglaucoma preparation	S01EC01

Annexe B.3 : Distribution des hdPS au début de chaque semaine de grossesse (SG)

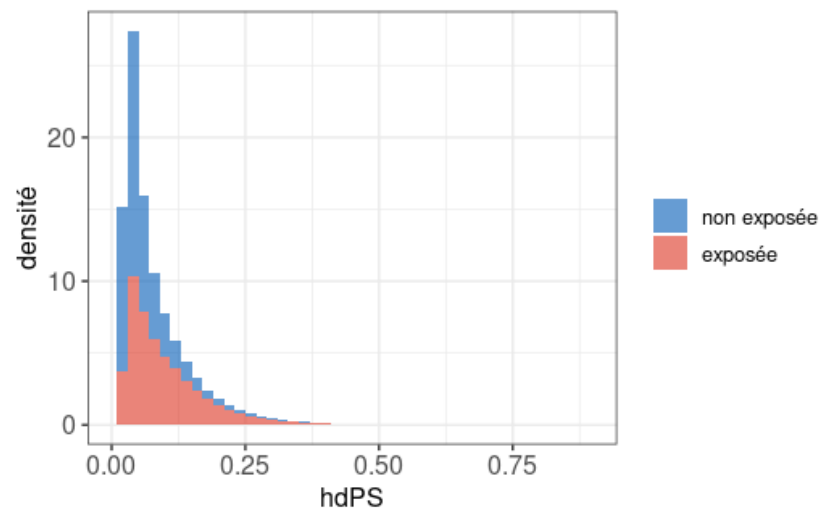




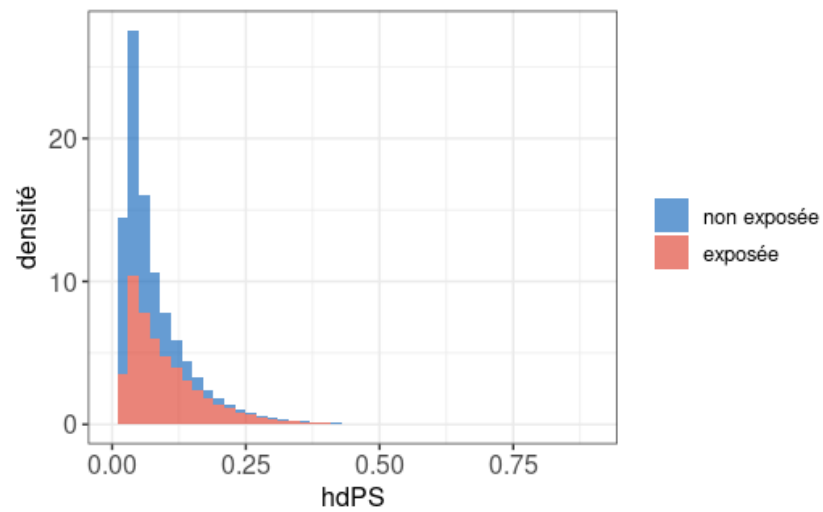
9e semaine de grossesse



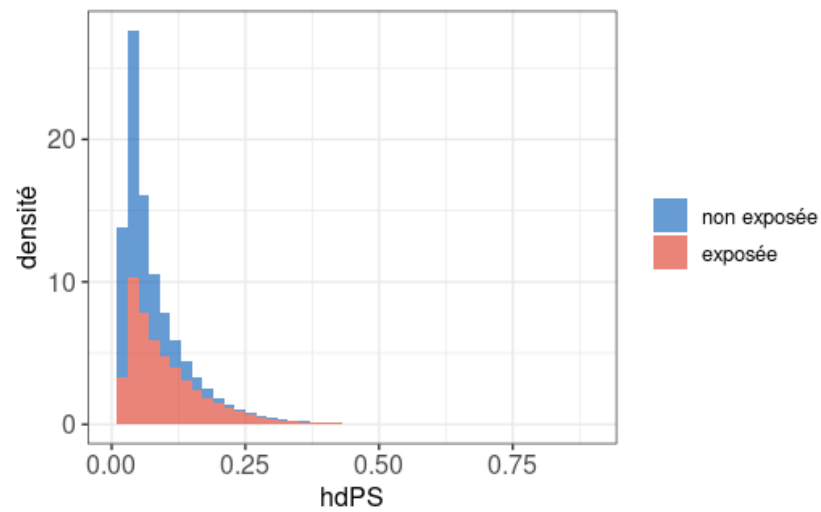
10e semaine de grossesse



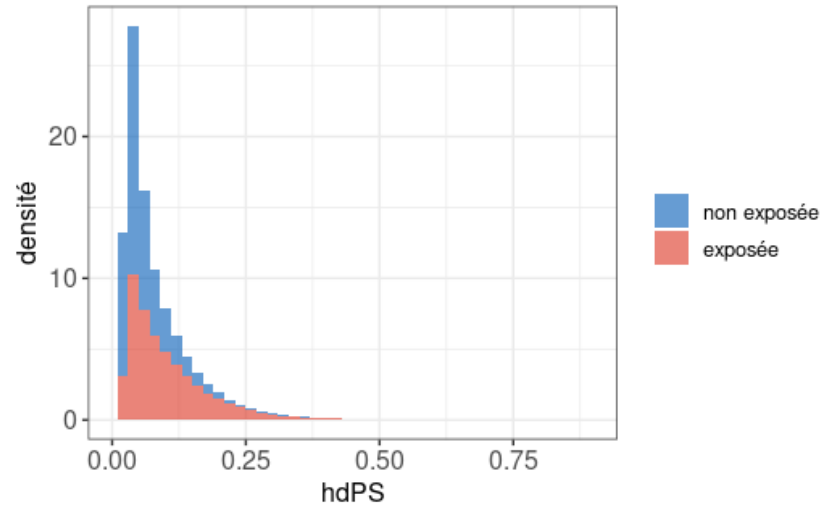
11e semaine de grossesse



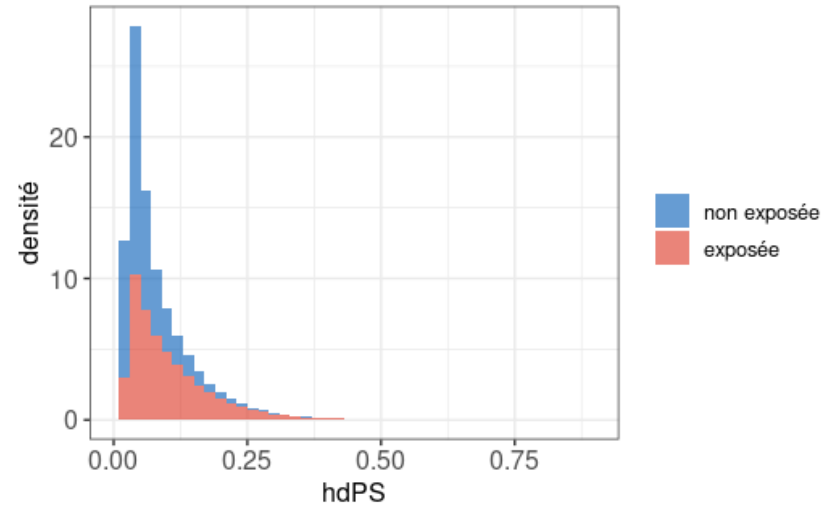
12e semaine de grossesse



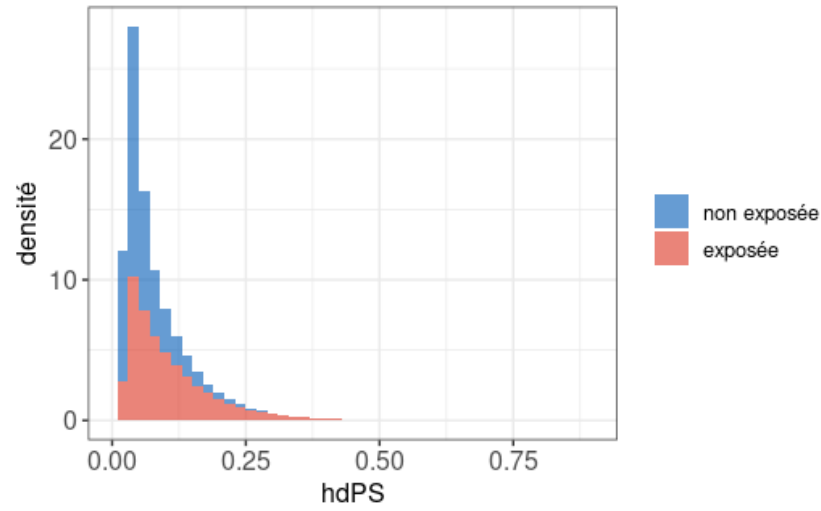
13e semaine de grossesse



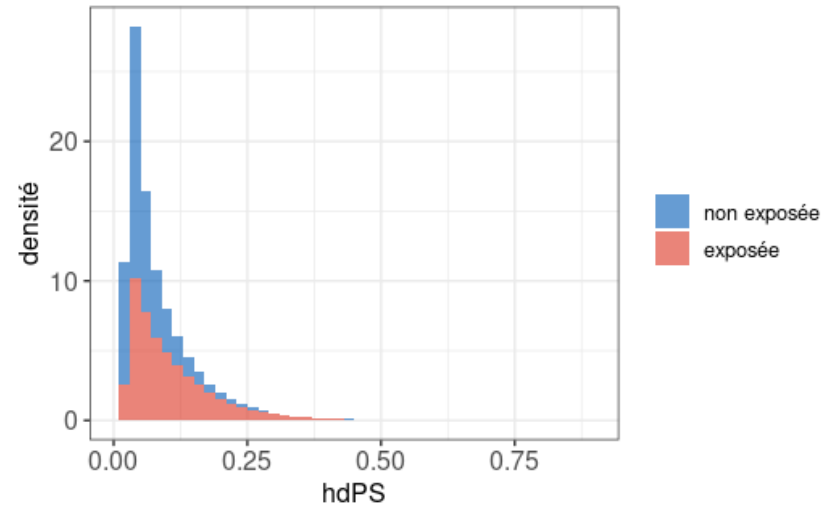
14e semaine de grossesse



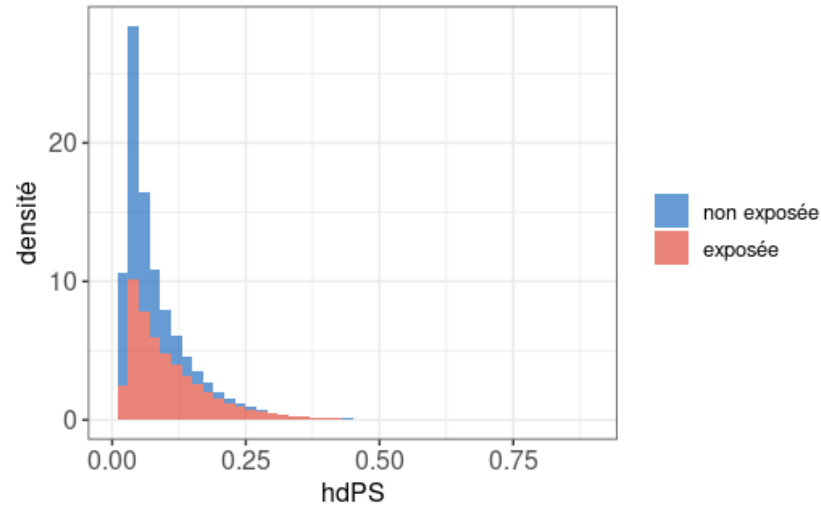
15e semaine de grossesse



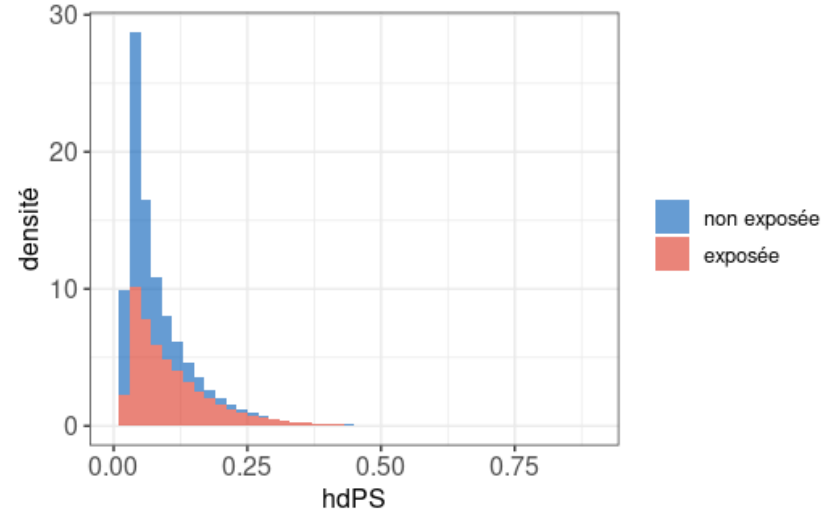
16e semaine de grossesse



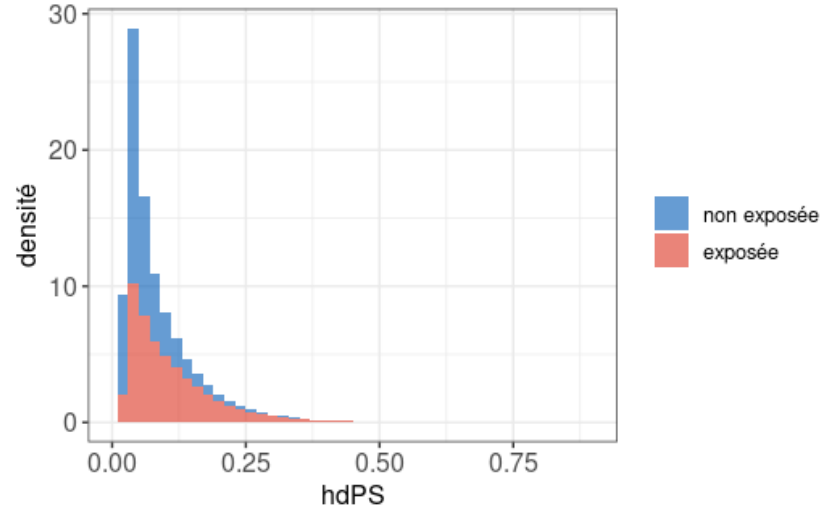
17e semaine de grossesse



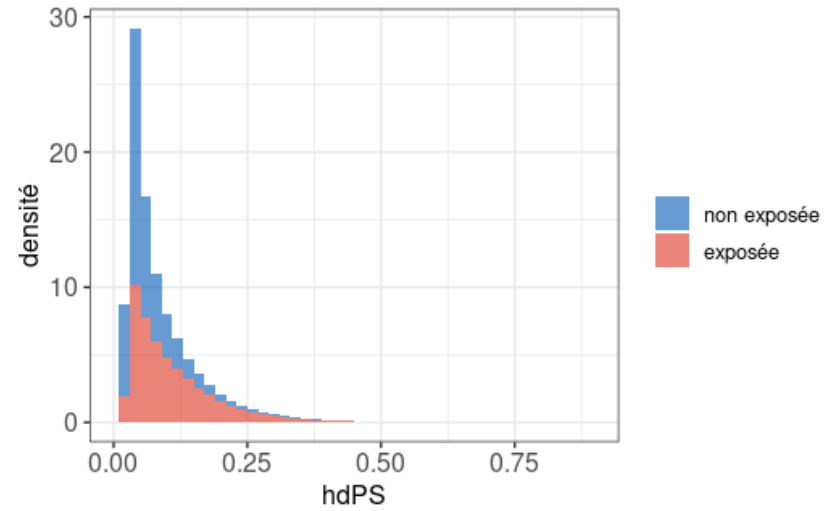
18e semaine de grossesse



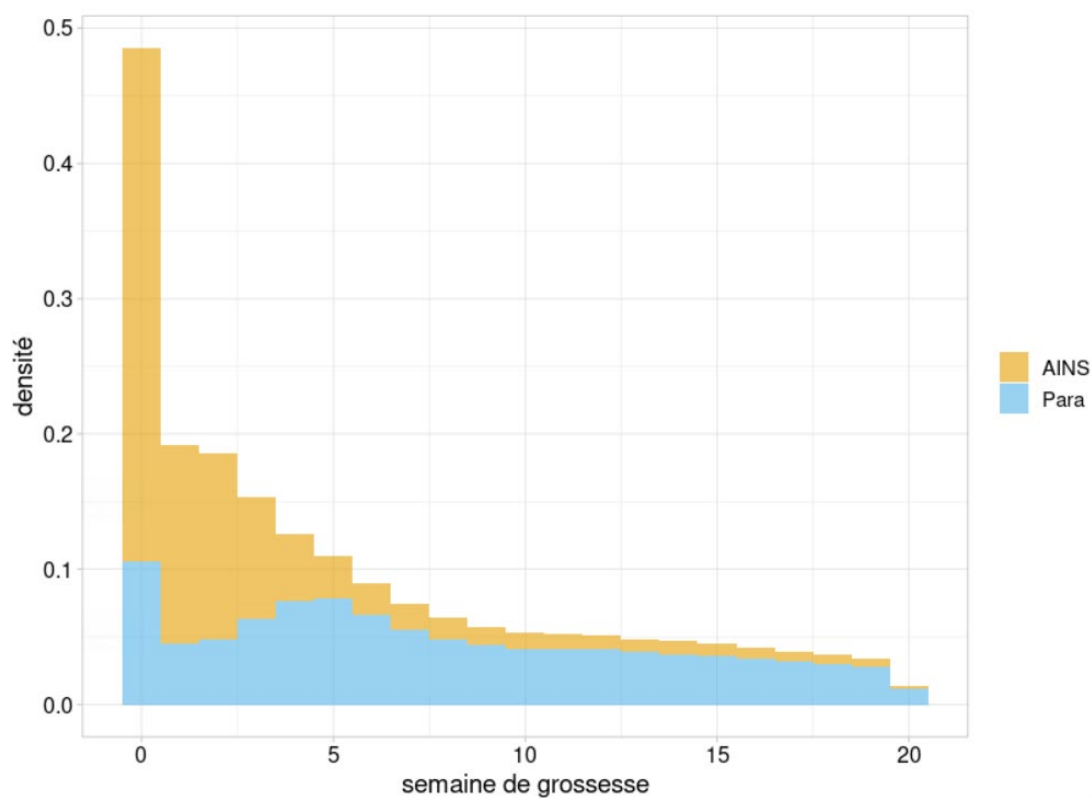
19e semaine de grossesse



20e semaine de grossesse



Annexe B.4 : Distribution des délais de première prescription de paracétamol et d'AINS après le début de grossesse



ANNEXE C : PUBLICATIONS ET SOUSSIONS

Annexe C.1 : Article soumis

Evaluation of Methods Adjusting for Indirectly Measured Confounding in Pharmacoepidemiology Using Healthcare Databases: An Empirical Study Concerning Drugs Inducing Prematurity

Chi-Hong Duong^{1*}, Sylvie Escolano¹, Romain Demailly^{1,2}, Anne Thiebaut¹, Jonathan Cottenet³, Catherine Quantin^{1,3,4}, Pascale Tubert-Bitter^{1†}, Ismail Ahmed^{1†} † These authors contributed equally to this work

¹ High-Dimensional Biostatistics for Drug Safety and Genomics, CESP, Université Paris-Saclay, UVSQ, Université Paris-Sud, Inserm, Villejuif, France.

² Faculty of Medicine and Midwifery, ETHICS EA 7446 Lille Catholic University, F-59000 Lille, France

³ Biostatistics and Bioinformatics (DIM), Dijon University Hospital, Dijon, France; Faculty of Medicine, University of Bourgogne-Franche-Comté, Dijon, France.

⁴ Clinical Epidemiology/Clinical Trials Unit, Clinical Investigation Center 1432, Dijon, France.

*** Correspondence:**
Chi-Hong Duong
hong.duong@inserm.fr

Key words: Machine learning, unmeasured confounders, propensity score, targeted learning, g-computation, preterm delivery

Abstract

Introduction: Unmeasured confounding in healthcare databases is a major issue in pharmacoepidemiologic studies. Extensions of causal inference methods to high-dimensional settings could help address this problem, but studies comparing their performance in real-world databases are still lacking. In the current study, we compare the ability to reduce the measured and indirectly measured confounding of three causal inference methods adapted to a real-world high-dimensional database using a machine learning LASSO algorithm: G-computation (GC), Targeted Maximum Likelihood estimation (TMLE) and Propensity Score (PS) with overlap (OW) or stabilized inverse probability treatment weighting (SIPTW).

Method: This empirical study was based on the French national healthcare claims database (SNDS), consisting of 2,172,702 pregnancies ≥ 22 weeks of gestation over the period 2011-2014. We used a set of 42 negative and 13 positive reference drugs related to prematurity risk, with the exposure window defined as the 3-month period prior delivery. For each reference drug, the log(OR) for prematurity, defined as deliveries occurred before 37th week of gestation, was estimated using each method. The proportions of positive associations (i.e. false positives for negative references and true positives for positive references) were calculated and compared between the methods.

Results: All methods performed significantly better in terms of false positives than a crude model not exploiting all the available variables in the SNDS. The lowest proportion of false positives was obtained with TMLE (45.2%), followed by GC (47.6%). The highest proportion of true positives was obtained with GC (92.3%).

Discussion: Our findings confirm the interest of using causal inference methods that exploit the wealth of data typically available in healthcare databases. GC was found to be the best in terms of performance and ease of implementation.

1 Introduction

In recent years, the use of healthcare databases has increased significantly in pharmacoepidemiologic studies. However, with observational studies and those using healthcare databases, ascertaining a causal effect of a treatment on a health outcome is hampered by the fact that observed associations are often biased due to a lack of exchangeability between compared groups.

Conventional methods like regression adjustment and propensity score (PS) deal with observed potential confounders.¹ However, they cannot mitigate the bias caused by unmeasured confounders that hamper pharmacoepidemiologic studies based on healthcare databases. In the context of high dimensionality of databases, one can capture information on unmeasured features from a large number of observed covariates and use this indirect information to attenuate the bias caused by these features.² In this setting, a set of observed covariates correlating with the unmeasured confounding acts as an unmeasured confounder proxy. The high-dimensional propensity score algorithm (hdPS) is a data-driven algorithm widely used nowadays for automated variable selection from a large number of observed covariates to improve confounding control.² Previous studies found that supplementing investigator-specified covariates with hdPS-identified confounder proxies improved confounding control.²⁻⁴

In addition to PS and hdPS, two other methods are increasingly used in causal inference: G-computation (GC) and targeted maximum likelihood estimation (TMLE).^{5,6} GC consists in predicting hypothetical outcomes of individuals from which causal parameters of interest (e.g. odds ratio, risk difference, relative risk) are estimated. TMLE combines two models, i.e. a PS exposure model and a GC outcome model through a so-called ‘targeting’ step and results in a doubly robust estimator of the treatment effect. Originally, one must make an assumption of no unmeasured confounding, along with some other assumptions, to causally interpret the results of these methods. Like hdPS, however, these methods can be used with data-driven screening algorithms for the automated search for confounding proxies to mitigate the violation of the ‘no unmeasured confounding’ assumption.⁷⁻¹⁰

Several studies have compared the performance of these methods based on simulation studies and have produced inconsistent results.^{7,10-13} While GC was shown to perform better in the studies of Chatton et al. and Kang et al.,^{7,12} Schuler et al. found that TMLE outperformed GC and inverse probability weighting under model misspecification.¹¹ Despite their obvious interest, most of these studies were conducted on simulated datasets that may differ significantly from those used in real practice, not only in terms of size but also in terms of exposure, outcome prevalence, and correlation structures between variables. In particular, to our knowledge, studies investigating this issue of indirectly measured confounding in high-dimensional settings focused on PS-based approaches.¹⁴ In this work, we compared the performance of PS, GC and TMLE applied to high-dimensional data based on an empirical study conducted on an extraction from the French national health data system (SNDS, *Système National des Données de Santé*) targeting all pregnancies ≥ 22 weeks of gestation over the period 2011-2014. The methods were compared on the basis of a recently published set of

drugs related to prematurity consisting of 42 negative and 13 positive reference drugs.¹⁵

2 Materials and methods

2.1 Data

We used an extraction from the SNDS which contains information about hospital and claimed ambulatory care in nearly the whole French population.¹⁶ The extraction concerned pregnant women from January 2011 to December 2014. Pregnancies were identified using International Classification of Diseases 10th Revision codes (ICD-10) and French national codes of clinical procedures (Classification Commune des Actes Médicaux). We excluded pregnancies of less than the 22nd week of gestation and those in which gestational age and number of days since the first day of the last menstruation were lacking.

Premature births were defined as all pregnancies in which deliveries occurred before the 37th week of gestation.

Our data included the following variables: all drug dispensations (coded according to the fifth level of the Anatomical Therapeutic Chemical classification, ATC), 16 common complications and diseases (ICD-10 codes) recorded at the end of pregnancies: multiple pregnancy, gestational diabetes, oligohydramnios, polyhydramnios, premature rupture of membranes, macrosomia, pre-eclampsia, hemorrhagic diseases, hyperemesis gravidarum, intrahepatic cholestasis of pregnancy, acute renal failure, acute liver failure, acute pancreatitis, thrombotic disorders, renal colic, infection; and the following socio-demographical covariates: age, status of Complementary Universal Health Care Coverage (CMU) restricted to people with low income, and Fdep, an index of socioeconomic status of the area of residence.¹⁷

Following Demailly et al., for each pregnancy, we considered three drug exposure periods: a 3-month period before conception (T_0), the first trimester of pregnancy (T_1), and a 3-month period before delivery (T_{BD}).¹⁵ A woman was considered as exposed to a drug in a given period if at least one dispensation of that drug was recorded during that period. Thus, there were $k \times 3$ period-specific binary variables created from k drugs. Drug exposures concerning fewer than 100 women were excluded from the analyses. The exposure period of interest was T_{BD} .

2.2 Statistical Methods

2.2.1 Dimension reduction

For each investigated drug, we reduced the dimension of the dataset and the computational burden using the Bross formula (Supplementary material, section S1), as proposed by Schneeweiss et al.: a subset of 500 covariates marginally associated with both prematurity and drug exposure of interest was selected from the set of 16 complications and diseases, and drug exposure covariates.^{2,18} These 500 preselected covariates as well as the three socio-demographic covariates were then used as predictors for the PS, GC and TMLE algorithms.

2.2.2 Propensity Score (PS)

As originally proposed, the hdPS algorithm consists of two main steps: a variable preselection step based on the Bross formula (i.e. the step we applied for dimension reduction) and a PS construction step based on all the preselected variables.² One of the limitations of hdPS is that the number of covariates included in the PS model is chosen arbitrarily. Previous studies showed that including too many covariates in the PS model can lead to overfitting and harm the properties of the estimator.^{19,20} Following Franklin et al., we further refined the set of 503 preselected covariates using a LASSO outcome regression model.²¹ This model helps to remove variables unassociated with prematurity, hence the instrumental variables. The regularization parameter of the LASSO was chosen using the Bayesian Information Criterion (BIC). The final subset of covariates was included in a classical logistic regression model to build the PS.

The PS was accounted for in the final outcome model using two weighting approaches proposed to address extreme PS that are very likely to occur in the event of rare exposures: (1) overlap weights (OW) which equal PS for unexposed and $1 - PS$ for exposed women; (2) stabilized inverse probability treatment weight (SIPTW) which equals $P(A = 0)/(1 - PS)$ for unexposed and $P(A = 1)/PS$ for exposed women, where A is a binary treatment, $P(A = 0)$ and $P(A = 1)$ are the marginal prevalence of unexposed and exposed groups respectively.²²⁻²⁴ Furthermore, for SIPTW, we applied a data-adaptive truncation for the lower bound of PS, $[\sqrt{n} \times \ln(n)]/5$, (n being the number of observations), which was shown to help in optimizing the bias-variance tradeoff.²⁵

2.2.3 G-computation (GC)

G-computation consists in estimating potential outcomes of each individual under specific interventions. Full details of the method can be found in Robins et al.⁵ In brief, GC involves fitting an outcome model (frequently referred to as the Q-model) on the entire sample and then predicting the potential outcomes (observed/factual and unobserved/counterfactual outcomes) of each individual under set values of the intervention. In our study, because all exposures of interest are binary, each individual has two predicted probabilities of preterm birth $\hat{P}(Y_0 = 1)$ and $\hat{P}(Y_1 = 1)$ had that individual be unexposed or exposed to the drug of interest at T_{BD} , respectively. The odds ratio was then calculated from the means of the two vectors of predicted

probabilities ($\hat{\pi}_0$ and $\hat{\pi}_1$) as $\frac{\frac{\hat{\pi}_1}{(1-\hat{\pi}_1)}}{\frac{\hat{\pi}_0}{(1-\hat{\pi}_0)}}$.

Many flexible, machine-learning algorithms can be used to estimate the Q-model from a large number of covariates. As for PS, we used the LASSO in combination with the BIC. For the drug of interest, the corresponding regression coefficient was estimated without regularization constraint in order to force the variable into the final model.

2.2.4 Targeted maximum likelihood estimation (TMLE)

Originally introduced by van der Laan et al.,⁶ TMLE combines the information about the outcome model (Q-model) and the PS exposure model to estimate the targeted

parameter, here the marginal odds ratio. The method is doubly robust, thereby giving an unbiased estimate if either the outcome or the exposure model is correctly specified.

Briefly, we implemented the TMLE by following these four steps:

- 1) estimate the PS;
- 2) calculate initial estimates of the potential outcomes (GC step);
- 3) update the initial GC estimates by adjusting a logistic regression model: $\text{logit}[E(Y|A, W)] = \text{logit}(Q_a) + \epsilon H_a$, where Q_a are the predicted outcomes from the GC model according to the treatments individuals actually received and W is the set of baseline covariates. $H_a = \frac{I(A=1)}{PS} - \frac{I(A=0)}{1-PS}$, with $I(\cdot)$ the indicator function, is referred to as the clever variable and can be used to integrate the estimated PS into the targeting model;
- 4) estimate the marginal odds ratio from the updated estimates of the potential outcomes by using the estimated ϵ .

The summary of all the studied methods is illustrated in Figure 1.

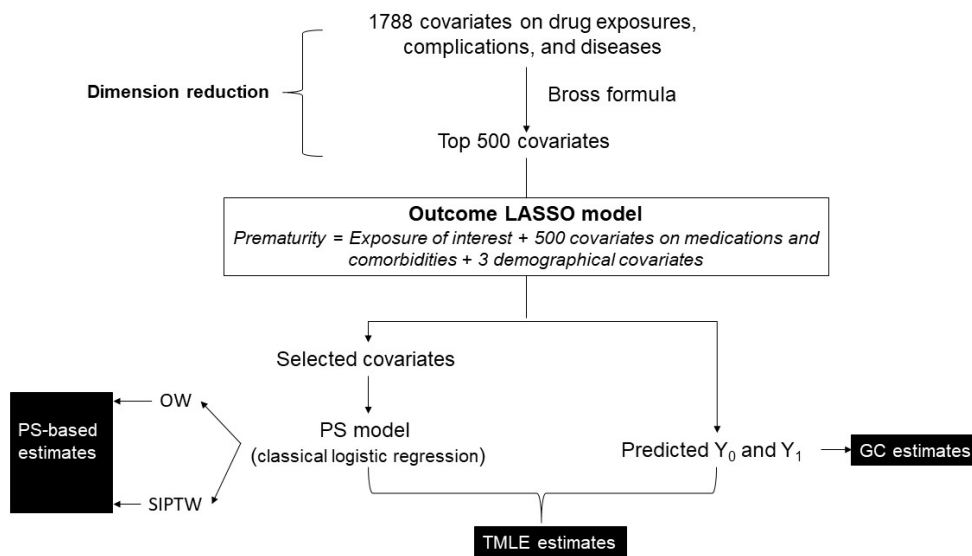


Figure 1 Summary of studied algorithms. Abbreviations: LASSO, Least Absolute Shrinkage and Selection Operator; PS, propensity score; GC, G-computation; TMLE, targeted maximum likelihood estimation; OW, overlap weight; SIPTW, stabilized inverse probability of treatment weighting.

Estimator variance was estimated by bootstrapping (500 iterations) for all methods according to the recommendations of Austin et al.²⁶ The 95% confidence intervals (95% CIs) were calculated based on asymptotic normal theory from the standard deviation of the bootstrap estimates.

2.3 Performance assessment

All the methods were implemented on a reference set of 42 negative and 13 positive reference drugs regarding the risk of preterm birth established in a previous study¹⁵. In the latter, automated pharmacovigilance signal detection was used to screen 400 drugs and generated 81 signals. These signals were further classified according to scientific literature and clinical expertise as “Spurious” (42 drugs), “Suspicious” (13 drugs) or “Inconclusive” (26 drugs). Here, our negative reference drugs correspond to the “Spurious” set from Demailly et al., i.e. drugs that do not increase the risk of preterm birth according to current medical knowledge (i.e. drugs that do not alter or might decrease the risk). Conversely, our positive reference drugs were defined as the “Suspicious” set from Demailly et al., i.e. drugs known to increase the risk of preterm birth. For each drug in the reference set, we estimated the odds ratio of preterm birth between women who were exposed to the drug of interest at least once at T_{BD} and those who were not exposed to the drug in the same period. The list of the reference drugs and their five-digit ATC codes can be found in table S1 and S2 in Supplementary material.

PS, GC and TMLE-based methods were also compared to a classical logistic regression model, referred to as the crude model, including the exposure of interest (at T_{BD}) along with the exposures to the same drug at T_0 , T_1 , and the three sociodemographic covariates.

For each of the 55 reference drugs, we estimated the 95%CI of the ORs with all methods. We then calculated the proportion of drugs for which one fell below, within or above the bounds of their 95%CIs. Methods were compared according to their proportion of false positives and true positives, defined as associations with lower bounds of the 95%CIs above one in the negative reference set and the positive reference set, respectively. As all methods do not target the same estimand, we did not directly compare their estimates; for each method, we compared the distribution of the estimated $\log(\text{OR})$ separately among the negative and positive references.

All statistical analyses were performed with the R statistical software (v3.4.4) using the `glmnet` (v2.0-16) and `adapt4pv` (v0.2-2) packages for BIC-LASSO regressions.

3 Results

The final dataset included 2,172,702 pregnancies after excluding 41 women with missing gestational age and number of days since the first day of the last menstruation. The median age was 30 years (interquartile range: 26 – 33). The prevalence of premature birth was 7.74%. There were 1772 period-specific drug exposure variables, and 16 complication and disease variables. Among the negative references, the most common drug was paracetamol (29.26%, $n = 635,788$) and the rarest one was nadroparin (0.01%, $n = 167$). Twenty-six out of 42 negative references were delivered to less than 1% of pregnancies. Among the positive references, oxazepam had the highest prevalence (0.23%, $n = 5015$) and mianserin had the lowest (0.01%, $n = 157$). The prevalence of all references can be found in the table S1 and S2 in Supplementary material.

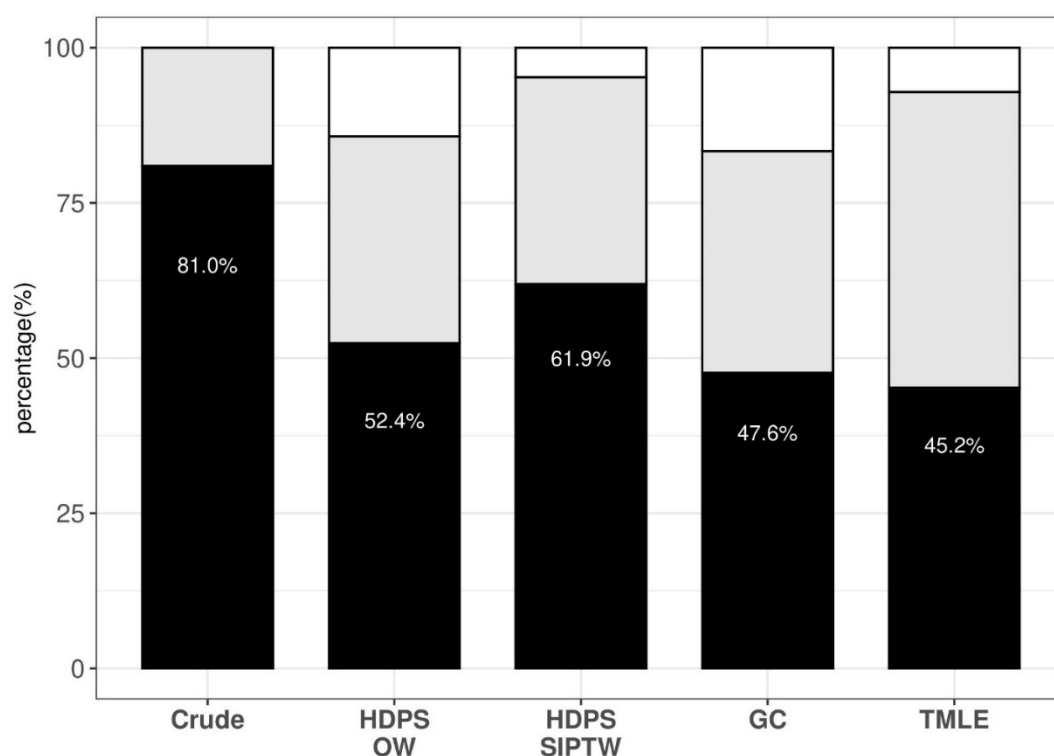


Figure 2 Proportion of negative references associated with (i) a lower bound of the 95% confidence interval (95%CI) above one (black), (ii) a 95%CI containing 1 (gray) and (iii) an upper bound of the 95%CI below one (white). Percentage represented in black bars corresponds to proportion of false positives.

Figure 2 shows that among the negative references, the crude model resulted in the highest proportion of false positives (81.0%) followed by PS-SIPTW (61.9%). Incorporating information about exposure on the outcome model through doubly robust estimation slightly decreased the proportion of false positives (47.6% versus 45.2% for GC and TMLE, respectively). Regarding the concordance between the methods, 14 false positives were identified by all the methods. The set of false positives identified by the crude model encompassed those of the other methods. (Figure S1 – Supplementary material).

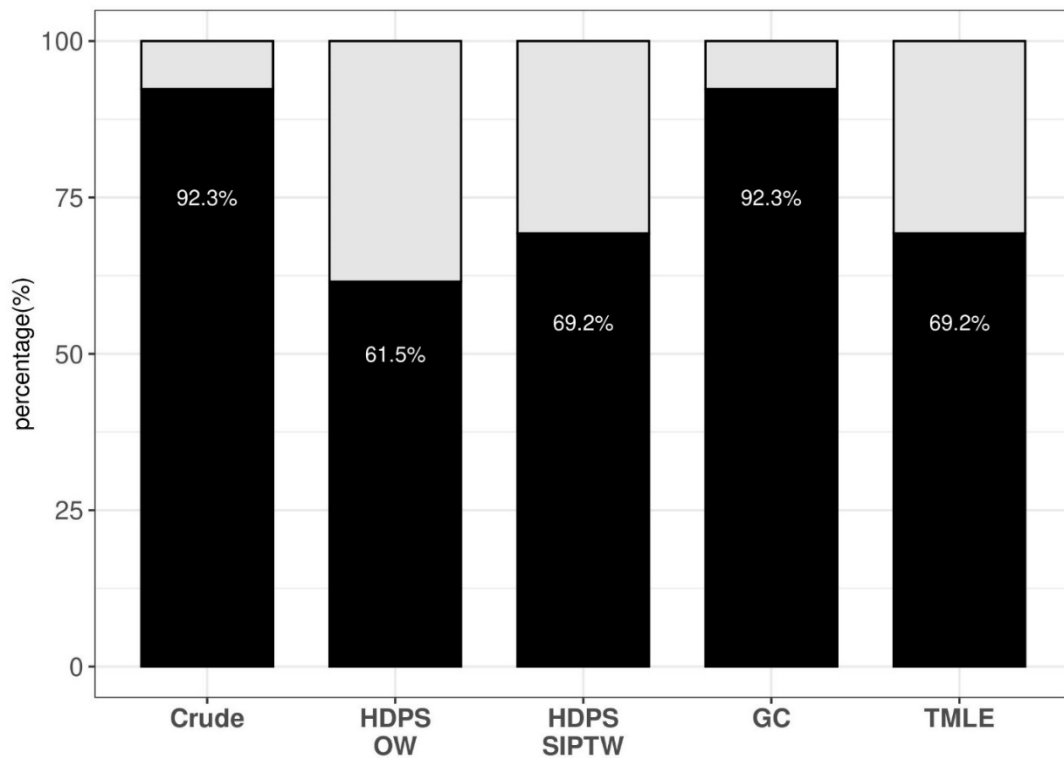


Figure 3 Proportion of positive references associated with (i) a lower bound of the 95% confidence interval (95%CI) above one (black), (ii) a 95%CI containing 1 (gray) and (iii) an upper bound of the 95%CI below one (white). Percentage represented in black bar corresponds to proportion of true positives.

As shown in Figure 3, among the positive references, the crude model and GC gave the highest proportion of true positives (92.3%). PS-OW resulted in the lowest proportion (61.5%) followed by PS-SIPTW (69.2%) and TMLE (69.2%). GC and the crude model identified the same set of true positives, and the latter included all the true positives identified by the other methods (Figure S2 – Supplementary material).

Figure 4 shows the distribution of the $\log(\text{OR})$ s according to the negative or positive status of the reference drugs for all methods. All methods discriminated between negative and positive references, but GC seemed to do so more markedly on average (Wilcoxon test p-values: Crude, 0.018; TMLE, 0.010; PS-SIPTW, 0.010; PS-OW, 0.014; GC, 0.001). $\log(\text{OR})$ estimates and their bootstrap 95%CI for all drugs and methods can be found in the figure S3 and S4 in Supplementary material.

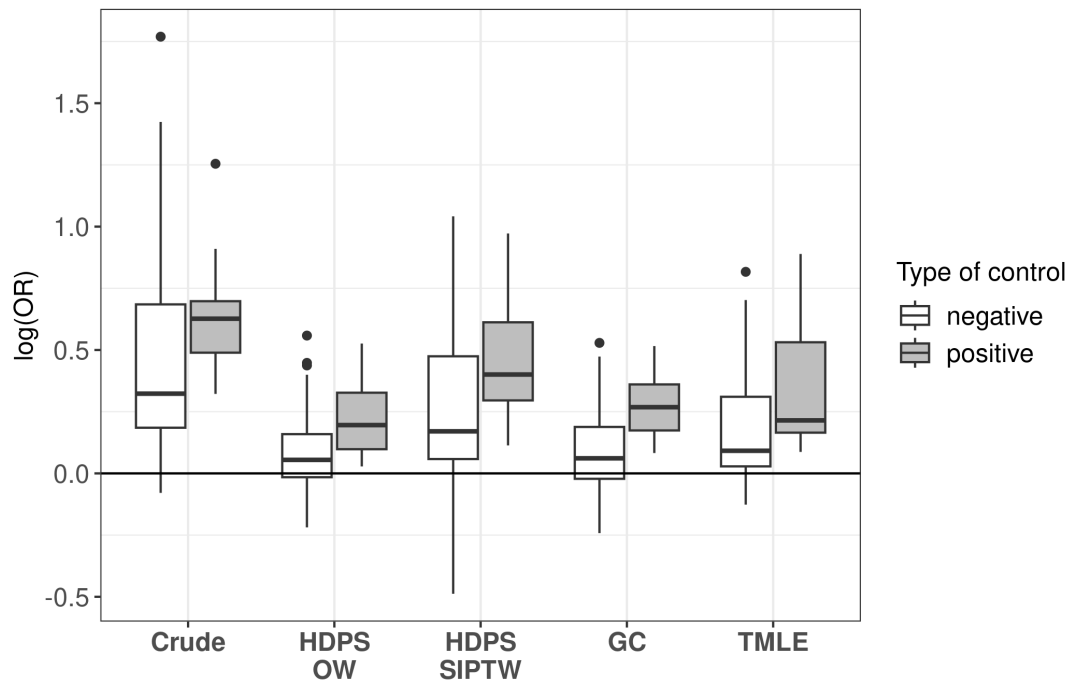


Figure 4 Box plot distribution of log(OR)s estimated in negative (white) and positive (gray) references.

4 Discussion

Despite the many advantages of healthcare databases, their use is hampered by unmeasured confounders. To address this issue, we compared the performance of PS, GC, and TMLE on a large SNDS extraction concerning pregnant women on the basis of a large reference set of negative and positive drugs. We compared the proportion of false positives and true positives across all negative and positive references respectively. Our results show the superiority of the three methods over the crude model. GC gave the best performance. Although it produced a slightly higher proportion of false positives than TMLE (Figure 2), it clearly stood out from the other methods in terms of the proportion of true positives (Figure 3).

The number of false positives was relatively high for all methods (Figure 2). This could be partly explained by the negative reference set used, which was constructed from spurious pharmacovigilance signals¹⁵ (see Performance assessment in the Methods section), for some of which substantial bias may remain even after adjustment. For example, progesterone and hydroxyprogesterone were incorrectly classified as positive by all methods. As these drugs are often prescribed to women at high risk of preterm birth to prevent preterm labour²⁷, there is a significant source of indication bias that was only partially addressed by the methods, which was reflected in the reduction of the OR estimates. Nevertheless, overall our results showed that PS, GC and TMLE-based methods improve the control of confounding.

Drug prevalence may be one of the reasons why drugs were classified as negative, as they tend to be associated with large confidence intervals. This is notably true for false negatives with the example of mianserin, which had the lowest prevalence in the positive references (0.01%) and was falsely classified as negative by all methods. It is noteworthy that the confidence intervals of GC were narrower than those of the other methods (Figure S4), which may explain why GC has more power to detect positive reference drugs with low prevalence, such as diazepam (0.05%), which was correctly classified as positive only by GC.

An individual was considered exposed if they had at least one prescription for the drug of interest during the period considered (i.e. T_0 , T_1 and T_{BD}). This choice was guided by the narrow observation window for exposure (3 months). There may be a risk of misclassification of exposure but it can be considered non-differential. In addition, there would also be a risk of misclassification if more prescriptions were required for an individual to be considered exposed, i.e. truly exposed individuals would be considered unexposed with a stricter definition.

The underperformance of TMLE relative to GC is not in agreement with some previous studies.^{11,13,28,29} However, none of them was conducted in high-dimensional settings and considered indirectly measured confounding. In addition, some of these studies compared TMLE and GC when the outcome model was mis-specified but the PS model was correctly specified, which disadvantaged the performance of GC more than TMLE.^{11,13} In our study, the underperformance of TMLE compared to GC is perhaps attributable to the difficulty of estimating PS in a context of rare drug exposures (Table S1 and S2 – Supplementary material). Indeed, PS was found to be the least efficient of the three methods, especially when SIPTW weighting was used. Yet TMLE uses SIPTW weighting. Moreover, in real data applications, the validity of models is always unknown, and the theoretical efficiency of TMLE become less evident when neither the outcome nor the exposure model is correctly specified. One study found that, in such a case, doubly robust estimators may underperform GC with a good predictive outcome model, in terms of bias, efficiency and robustness.¹² Similarly, Kreif et al.¹⁰ found that combining two mis-specified models did not reduce the bias compared to using single methods.

Our empirical study was particularly demanding from a computational point of view, requiring more than two weeks of parallelized calculation on a server equipped with 100 cores and 1 terabyte of random access memory. This is due to the size of the data, the reference set (55 drugs) and the number of methods evaluated. The number of 500 covariates selected using the Bross formula was chosen as a compromise between attempting to maximize the capture of potential confounders and computational cost. This choice is consistent with other studies that typically include 200-500 covariates to estimate a high-dimensional propensity score.³⁰ In addition, it has been found that varying the number of covariates between 100 and 1000 does not substantially change the effect estimates.³¹ The computational cost was also increased by using bootstrapping to estimate variances. Although variance estimators exist for some of the methods (e.g. for the TMLE), we used bootstrapping to ensure better comparability of performance. The computational cost also guided our decision to use LASSO

regression, which is a relatively low-cost machine-learning method. Regarding GC and TMLE, the use of more flexible methods (such as random forest, gradient tree boosting or neural network) as well as the combination of their predictions through the super learner for the outcome model could have improved their respective performances by limiting functional form misspecification. Nevertheless, it seems rather unlikely that this would have altered their comparative performances.^{10,32} With regard to the use of the BIC rather than cross-validation, this choice was driven by the objective of variable selection rather than prediction and not by computational considerations.

The reference set used in this work included 55 drugs, some of which are in the same therapeutic class. Although there may be some common adverse reactions at the therapeutic class level, this is not necessarily the case. In addition, different drugs in the same class may be prescribed preferentially to different populations (and therefore be related to different confounders) and may have different prevalence. Another feature of our reference set is that it includes a large proportion of rare drug exposures, which is to be expected in practice.

The evaluation of a large number of associations implies a certain standardization of the analyses, which may appear simplistic compared with carefully designed pharmacoepidemiologic studies focusing on a given association. In particular, our crude model only included a minimal and common set of covariates, rather than specific sets for each drug. Nevertheless, this crude model is useful as a baseline against which we can measure the benefit of methods based on automated screening for potential confounders. In addition, although a large number of covariates were considered, they are unlikely to be sufficient for unmeasured confounding control. Nevertheless, these limitations are not likely to invalidate our findings.

In this methodological work, we found that in a high-dimensional, real-world, complex setting, the use of methods that exploit the richness of data can improve confounding control. GC was found to offer better performance than HPDS and TMLE, as well as straightforward implementation. Despite their use, substantial residual confounding remains, which could be further reduced by combining them with more refined pharmacoepidemiologic designs.

List of abbreviations: GC: G-computation, TMLE: Targeted Maximum Likelihood Estimation, LASSO: Least Absolute Shrinkage and Selection Operator, PS: propensity score; OW: overlap weight, CI: confidence interval; SIPTW: stabilized inverse probability of treatment weighting.

Data availability statement

The study was approved by the French data protection agency (Commission Nationale de l'Informatique et des Libertés), but the data set is not publicly available because of data access permission policy prohibition.

Ethic statement

Not applicable

Author contributions

CHD contributed to study conception, data analysis, and article drafting.

SE contributed to acquisition and analysis of data.

CQ contributed to data acquisition and interpretation.

RD contributed to result interpretations.

AT and JC contributed to data analysis and interpretation.

PTB and IA contributed to study conception, data interpretation and analysis, and article drafting.

All authors contributed to the article reviewing and approved the final version of the paper.

Funding

This project has received funding from the French National Agency for Medicines and Health Products Safety (ANSM - Grant n° 2020S012) in the framework of the Scientific Interest Group EPI-PHARE, and from the French National Research Agency (Grant number: ANR-15-CE36-0006-01). This work represents the views of the authors and does not necessarily represent the position of the relevant health authorities.

Conflicts of interest

The authors declare that they have no conflict of interest in this research.

Consent to participate

Not applicable

Consent to publication

Not applicable

References

1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55. doi:10.1093/biomet/70.1.41
2. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional Propensity Score Adjustment in Studies of Treatment Effects Using Health Care Claims Data. *Epidemiology*. 2009;20(4):512-522. doi:10.1097/EDE.0b013e3181a663cc
3. Hallas J, Pottegård A. Performance of the High-dimensional Propensity Score in a Nordic Healthcare Model. *Basic Clin Pharmacol Toxicol*. 2017;120(3):312-317. doi:10.1111/bcpt.12716
4. Garbe E, Kloss S, Suling M, Pigeot I, Schneeweiss S. High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications. *Eur J Clin Pharmacol*. 2013;69(3):549-557. doi:10.1007/s00228-012-1334-2
5. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7(9):1393-1512. doi:10.1016/0270-0255(86)90088-6
6. Laan MJ van der, Rubin D. Targeted Maximum Likelihood Learning. *Int J Biostat*. 2006;2(1). doi:10.2202/1557-4679.1043
7. Chatton A, Le Borgne F, Leyrat C, et al. G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. *Sci Rep*. 2020;10(1):9219. doi:10.1038/s41598-020-65917-x
8. Le Borgne F, Chatton A, Léger M, Lenain R, Foucher Y. G-computation and machine learning for estimating the causal effects of binary exposure statuses on binary outcomes. *Sci Rep*. 2021;11(1):1435. doi:10.1038/s41598-021-81110-0
9. Austin PC. Using Ensemble-Based Methods for Directly Estimating Causal Effects: An Investigation of Tree-Based G-Computation. *Multivar Behav Res*. 2012;47(1):115-135. doi:10.1080/00273171.2012.640600
10. Kreif N, Gruber S, Radice R, Grieve R, Sekhon JS. Evaluating treatment effectiveness under model misspecification: A comparison of targeted maximum likelihood estimation with bias-corrected matching. *Stat Methods Med Res*. 2016;25(5):2315-2336. doi:10.1177/0962280214521341
11. Schuler MS, Rose S. Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies. *Am J Epidemiol*. 2017;185(1):65-73. doi:10.1093/aje/kww165

12. Kang JDY, Schafer JL. Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. *Stat Sci.* 2007;22(4):523-539. doi:10.1214/07-STS227
13. Ju C, Gruber S, Lendle SD, et al. Scalable collaborative targeted learning for high-dimensional data. *Stat Methods Med Res.* 2019;28(2):532-554. doi:10.1177/0962280217729845
14. Zhang L, Wang Y, Schuemie MJ, Blei DM, Hripcsak G. Adjusting for indirectly measured confounding using large-scale propensity score. *J Biomed Inform.* 2022;134:104204. doi:10.1016/j.jbi.2022.104204
15. Demailly R, Escolano S, Haramburu F, Tubert-Bitter P, Ahmed I. Identifying Drugs Inducing Prematurity by Mining Claims Data with High-Dimensional Confounder Score Strategies. *Drug Saf.* 2020;43(6):549-559. doi:10.1007/s40264-020-00916-5
16. Scailteux LM, Droitcourt C, Balusson F, et al. French administrative health care database (SNDS): The value of its enrichment. *Therapies.* 2019;74(2):215-223. doi:10.1016/j.therap.2018.09.072
17. Rey G, Jouglu E, Fouillet A, Hémon D. Ecological association between a deprivation index and mortality in France over the period 1997 – 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health.* 2009;9:33. doi:10.1186/1471-2458-9-33
18. Bross ID. Spurious effects from an extraneous variable. *J Chronic Dis.* 1966;19(6):637-647. doi:10.1016/0021-9681(66)90062-2
19. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163(12):1149-1156. doi:10.1093/aje/kwj149
20. Myers JA, Rassen JA, Gagne JJ, et al. Effects of Adjusting for Instrumental Variables on Bias and Precision of Effect Estimates. *Am J Epidemiol.* 2011;174(11):1213-1222. doi:10.1093/aje/kwr364
21. Franklin JM, Eddings W, Glynn RJ, Schneeweiss S. Regularized Regression Versus the High-Dimensional Propensity Score for Confounding Adjustment in Secondary Database Analyses. *Am J Epidemiol.* 2015;182(7):651-659. doi:10.1093/aje/kwv108
22. Li F, Morgan KL, Zaslavsky AM. Balancing Covariates via Propensity Score Weighting. *J Am Stat Assoc.* 2018;113(521):390-400. doi:10.1080/01621459.2016.1260466
23. Li L, Greene T. A Weighting Analogue to Pair Matching in Propensity Score Analysis. *Int J Biostat.* 2013;9(2):215-234. doi:10.1515/ijb-2012-0030

24. Sato T, Matsuyama Y. Marginal Structural Models as a Tool for Standardization. *Epidemiology*. 2003;14(6):680-686. doi:10.1097/01.EDE.0000081989.82616.7d
25. Gruber S, Phillips RV, Lee H, van der Laan MJ. Data-Adaptive Selection of the Propensity Score Truncation Level for Inverse-Probability-Weighted and Targeted Maximum Likelihood Estimators of Marginal Point Treatment Effects. *Am J Epidemiol*. 2022;191(9):1640-1651. doi:10.1093/aje/kwac087
26. Austin PC. Bootstrap vs asymptotic variance estimation when using propensity score weighting with continuous and binary outcomes. *Stat Med*. 2022;41(22):4426-4443. doi:10.1002/sim.9519
27. Norman JE. Progesterone and preterm birth. *Int J Gynaecol Obstet*. 2020;150(1):24-30. doi:10.1002/ijgo.13187
28. Zivich PN, Breskin A. Machine Learning for Causal Inference: On the Use of Cross-fit Estimators. *Epidemiol Camb Mass*. 2021;32(3):393-401. doi:10.1097/EDE.0000000000001332
29. Naimi AI, Mishler AE, Kennedy EH. Challenges in Obtaining Valid Causal Effect Estimates with Machine Learning Algorithms. *Am J Epidemiol*. Published online July 15, 2021:kwab201. doi:10.1093/aje/kwab201
30. Tazare J, Wyss R, Franklin JM, et al. Transparency of high-dimensional propensity score analyses: Guidance for diagnostics and reporting. *Pharmacoepidemiol Drug Saf*. 2022;31(4):411-423. doi:10.1002/pds.5412
31. Rassen JA, Blin P, Kloss S, et al. High-dimensional propensity scores for empirical covariate selection in secondary database studies: Planning, implementation, and reporting. *Pharmacoepidemiol Drug Saf*. 2023;32(2):93-106. doi:10.1002/pds.5566
32. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J Clin Epidemiol*. 2010;63(8):826-833. doi:10.1016/j.jclinepi.2009.11.020