Score de propension en grande dimension et régression pénalisée pour la détection automatisée de signaux en pharmacovigilance

Émeline Courtois

Thèse de doctorat

NNT: 2020UPASR009

université PARIS-SACLAY

Score de propension en grande dimension et régression pénalisée pour la détection automatisée de signaux en pharmacovigilance

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 570, santé publique (EDSP)
Spécialité de doctorat : Santé publique - biostatistiques
Unité de recherche : Université Paris-Saclay, UVSQ, Inserm, CESP, 94807, Villejuif, France
Référent: Université de Versailles -Saint-Quentin-en-Yvelines

Thèse présentée et soutenue en visioconférence totale le 14 décembre 2020, par

Émeline COURTOIS

Composition du jury:

Christophe Ambroise Professeur, Université d'Évry	Président
Marc Cuggia Professeur, Université de Rennes 1	Rapporteur
Vivian Viallon Maître de Conférence, Université Claude Bernard (Lyon), CIRC	Rapporteur
Niklas Norèn Directeur Scientifique, Organisation Mondiale de la Santé	Examinateur
Pascale Tubert-Bitter Directrice de Recherche, Inserm U1018	Directrice de thèse
Ismaïl Ahmed Chargé de Recherche, Inserm U1018	Co-encadrant de thèse

Remerciements

Tout d'abord, mes remerciements s'adressent à ma directrice de thèse : Pascale Tubert-Bitter, qui par sa bienveillance, sa rigueur, mais aussi (et surtout) son éternel optimisme a su me guider tout au long de cette thèse, tant sur l'aspect scientifique que pour la valorisation de ce travail. Je remercie tout autant mon co-encadrant de thèse : Ismaïl Ahmed. Ce sont nos échanges, ta patience et ton encadrement en général qui m'ont permis de progresser et de produire ce travail dont je suis fière. Au delà de tout ce que tu as pu m'apporter sur le plan professionnel, merci aussi pour tes nombreux conseils sur les jeux de société, livres et séries, et pour m'avoir fait découvrir le jeux de rôle en ligne. Là encore, merci de ta patience car certaines sessions furent laborieuses.

Merci à tous les deux de m'avoir fait confiance pour travailler avec vous dans le cadre de cette thèse. A vos côtés j'ai énormément appris sur la plan scientifique et sur la pratique de la recherche, tout cela grâce à votre accompagnement, à votre écoute, à votre bienveillance.

Je remercie M. Marc Cuggia et M. Vivian Viallon pour avoir accepté d'être les rapporteurs de cette thèse. Merci pour vos remarques pertinentes et précieuses sur ce travail. Je remercie également M. Chritophe Ambroise et M. Niklas Norèn d'avoir accepté d'apporter leur expertise à ce jury et de l'intérêt que vous avez porté à ce travail de thèse.

Merci à mes collègues Anne, Sylvie (grâce à laquelle je suis en bonne forme physique), Hervé, Étienne, Matthieu, Liliane, Hong, Stéphanie et Romain avec qui j'ai eu plaisir à travailler, à échanger autour de certaines problématiques de cette thèse ou d'autres thématiques scientifiques, à partager des moments plus légers, aussi.

Merci à ma famille, tout particulièrement mes parents Valérie et Dominique ainsi que mon frère Alban, pour leur soutien, leur écoute, leur présence, pour leurs encouragements mais surtout pour leur amour. A mon grand-père, Jacques, qui ne m'aura pas vu soutenir,

mais qui aurait sans doute dit à l'issue de cette soutenance quelque chose comme : "Bravo Mimi, mais moi je fais ça chaque matin avant de manger mes tartines". Je sais que cela aurait été sa façon de me dire qu'il était fier de moi, comme il était fier de tous ces petits enfants. Tu me manques énormément.

Merci également à ma belle-famille, Violaine, Olivier Capucine, Bertille et Hortense entres autres, pour leur soutien et leur affection depuis toutes ces années. Mention spéciale à ma très chère Hortense, qui en plus de nous avoir soulagé avec son compagnon Jérémy de notre chatte capricieuse Ernestine Cha Guevara, a relu dans l'urgence mon manuscrit de thèse pour y chercher les dernières fautes.

Merci à mes tous mes amis, avec qui j'ai pu me changer les idées à de nombreuses reprises. Merci à Karine, qui en plus d'être une amie formidable depuis quinze ans, m'a permis de voyager au Rwanda et en Martinique pendant cette thèse pour lui rendre visite. Merci à Jean-Sebastien (à qui je dois une citation depuis son mémoire de master sur les médecins vénitiens en Syrie au XVème et XVIème siècles), à Juliette, à Selim, à Sevan (qui a toujours su rester "injectif"), à Thomas, à Pauline, à Fred, à Cécile, à David, à Émilie, à Jérémy, à Léonore (qui à l'heure actuelle, a brillamment soutenu sa thèse de pharmacie), à Anthony et tous les autres dont l'amitié m'est très précieuse.

Enfin merci à l'amour de ma vie, mon meilleur ami, quasi co-auteur de cette thèse et qui est maintenant docteur en biostatistiques : Arthur. Merci pour tout, de ton amour bien sûr, mais aussi de ton soutien sans faille, de ton éternel optimisme, de ta douceur, de ton écoute, de ta curiosité. Sans toi tout cela n'aurai pas été possible. Avec toi je grandi et j'avance. Je suis très fière de nous, de nos accomplissements. Nous avons tenu le coup face à une pandémie mondiale et une rédaction de thèse quasi simultanée, nous pouvons affronter le monde. Je t'aime.

Valorisation scientifique

Publications

Acceptée et publiée dans une revue internationale avec comité de relecture

Courtois, É., Pariente, A., Salvo, F., Volatier, É., Tubert-Bitter, P., Ahmed, I. (2018). Propensity Score-Based Approaches in High Dimension for Pharmacovigilance Signal Detection: an Empirical Comparison on the French Spontaneous Reporting Database. *Frontiers in Pharmacology*, 9(September), 1010.

Soumise dans une revue internationale avec comité de relecture

Courtois, É., Tubert-Bitter, P., Ahmed, I.. New adaptive lasso approaches for variable selection in automated pharmacovigilance signal detection. Soumis auprès de la revue *Statistical Methods in Medical Research*.

Publication en lien avec la thèse, soumise dans une revue internationale avec comité de relecture

Volatier, É., Salvo, F., Pariente, A., Courtois, É., Escolano, S., Tubert-Bitter, P., Ahmed, I.. High-dimensional propensity score-adjusted case-crossover for discovering adverse drug reactions from computerized administrative healthcare databases. Soumis auprès de la revue *Statistics in Medicine*.

Communications orales

Congrès nationaux

- Approches basées sur le score de propension en grande dimension pour la détection de signaux en pharmacovogilance, Journée des Jeunes Chercheurs de la Société Française de Biométrie, Paris, 29 Mai 2018.
- Exploitation conjointe de plusieurs bases de données dans la détection de signaux en pharmacovigilance via l'utilisation du lasso pondéré, 51$^{\text{ème}}$ Journées de la Statistique de la Société Française de Statistiques, Nancy, 3-7 Juin 2019.

Congrès internationaux

- Propensity Score-Based Approaches in High Dimension For Pharmacovigilance Signal Detection, XXIXth International Biometric Conference, Barcelona, July 8-13th 2018.

Communications affichées

Congrès internationaux

- New automated signal detection methods in pharmacovigilance combining medico-administrative database controls and spontaneous reporting cases, 40th Annual Conference of the International Society for Clinical Biostatistics, Leuven, July 14-18th 2019.
- Adaptive lasso approaches for pharmacovigilance signal detection : new penalty weights for variable selection, 41st Annual Conference of the International Society for Clinical Biostatistics, Krakow, August 23-27th 2020 (conférence virtuelle, avec courte présentation à l'oral du poster).

Outils logiciels

Package R disponible sur le CRAN `adapt4pv`.

Table des matières

Liste des figures

Liste des tableaux

Liste des acronymes

AIC	Critère d'Information d'Akaike (*Akaike Information Criterion*).
ATC	Anatomique, Thérapeutique et Chimique.
BCPNN	*Bayesian Component Propagation Neural Network.*
BIC	Critère d'information bayésien (*Bayesian Information Criterion*).
BMA	Base Médico-Administrative.
BNPV	Base Nationale de Pharmacovigilance.
CCO	*Case-Crossover.*
CIM-10	Classification Internationale des Maladies (*International Classification of Diseases* (ICD)), $10^{\text{ème}}$ révision.
CISL	*Class Imbalanced Subsampling Lasso.*
DILI	Lésion hépatique d'origine médicamenteuse (*Drug-Induced Liver Injury*).
EGB	Échantillon Généraliste des Bénéficiaires.
EI	Évènement Indésirable.
EMA	*European Medecines Agency.*
FDA	*Food and Drug Administration.*
FDP	Proportion de Faux Positifs (*False Discovery Proportion*).
FDR	*False Discovery Rate.*
FWER	*Family-Wise Error Rate.*
GTB	*Gradient Tree Boosting.*
hdPS	*high-dimensional Propensity Score.*

IPTW *Inverse Probability of Treatment Weighting.*

MedDRA dictionnaire Médical des Affaires Réglementaires (*Medical Dictionary for Regulatory Activities*).

MSCCS *Multiple Self-Controlled Case Series.*

MW *Matching Weights.*

OHDSI *Observational Health Data Sciences and Informatics.*

OMOP *Observational Medical Outcomes Partnership.*

OMS Organisation Mondiale de la Santé.

PMSI Programme de Médicalisation des Systèmes d'Information.

PPV Valeur Prédictive Positive (*Positive Predictive Value*).

PRR *Proportional Reporting Ratio.*

PS Score de Propension (*Propensity Score*).

PT *Preferred Term.*

RCP Résumé des Caractéristiques du Produit.

RFET *Reporting Fisher Exact Test.*

ROR *Reporting Odds Ratio.*

SCCS *Self-Controlled Case Series.*

SMD *Standard Mean Difference.*

SNDS Système National des Données de Santé.

SNIIRAM Système National d'Information Inter Régimes de l'Assurance Maladie.

SSA *Sequence Symmetry Analysis.*

UMC *Uppsala Monitoring Center.*

Chapitre 1

Introduction

1.1 Développement du médicament et pharmacovigilance

Une fois que l'intérêt thérapeutique d'une molécule a été mis en évidence par la recherche fondamentale, celle-ci va passer par différentes phases d'étude avant de devenir éventuellement un « vrai » médicament. On distingue deux phases différentes dans le développement des médicaments : la phase pré-clinique et la phase clinique. La phase pré-clinique permet de tester la molécule sur des cellules en culture et chez l'animal. Lors de cette phase, les mécanismes d'action, d'absorption, de transformation et d'élimination de la molécule sont évalués, ainsi que sa toxicité, sa mutagénicité, sa cancerogénicité et sa tératogénicité. Une fois ces informations recueillies, la molécule, si elle est sélectionnée, peut passer en phase de développement clinique chez l'homme. L'évaluation clinique repose sur trois phases :

- La phase I a pour objectif de tester la toxicité du médicament. La molécule est administrée à un petit nombre de volontaires, sains ou malades et son devenir dans l'organisme en fonction du temps est évalué.
- La phase II a pour objectif de déterminer la posologie optimale du médicament et ses éventuels effets indésirables. Elle porte sur des volontaires malades (entre 100 et 300). Dans un premier temps, une dose minimale efficace pour laquelle des effets indésirables ne sont pas observés ou sont minimes est déterminée. A partir de cette

dose, l'éventuel bénéfice thérapeutique du médicament est évalué.

- La phase III a pour objectif de tester l'efficacité du médicament. Elle porte sur plusieurs centaines voire milliers de malades, en fonction de la maladie considérée. Cette phase de développement clinique du médicament dure plusieurs années. Lors de cette phase, l'efficacité du médicament est comparée à celle du traitement de référence, s'il y en a un, ou à un placébo.

A la fin des ces différentes étapes, le médicament sera mis sur le marché s'il a été démontré que son bénéfice pour la santé est supérieur au risque potentiel qu'il représente. La phase IV débute une fois le médicament commercialisé. Cette phase a pour but de détecter les effets indésirables rares, ou les complications tardives liées à son administration [1].

Les conditions d'exposition à un médicament en vie réelle étant très différentes de celles des essais cliniques réalisés lors des différentes phases, les effets secondaires et en particulier les effets indésirables des médicaments sont souvent repérés après leur introduction sur le marché. En effet, lors de ces essais les populations considérées sont plus homogènes que la population générale. Elles ont des caractéristiques spécifiques et une taille limitée. De plus la notion de sujet malade est étroitement définie : elle ne repose que sur la maladie étudiée. Enfin, le recul sur les durées d'exposition reste court. Ainsi, la survenue d'effets indésirables peut être due, entre autres, à une interaction complexe avec des sous-catégories de la population générale, ou à une longue période de latence après l'exposition. La pharmacovigilance post-commercialisation a pour but de détecter le plus précocement possible les effets indésirables qui n'ont pas été identifiés lors du développement des médicaments. Les différentes étapes de développement du médicament et de surveillance une fois celui-ci commercialisé sont schématisées en figure 1.1.

1. Dossier d'information sur le développement du médicament de l'Inserm.

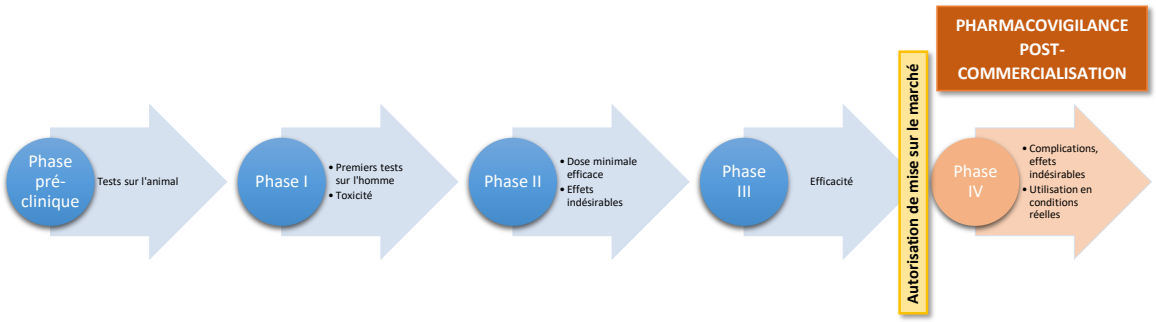


FIGURE 1.1 – Développement des médicaments et surveillance après leur introduction sur le marché.

1.2 Pharmacovigilance : bases de données et méthodes

La plupart des systèmes de pharmacovigilance s'appuient sur de grandes bases de données qui regroupent des notifications individuelles d'évènements indésirables (EI) soupçonnés d'être liés à la prise de médicaments. Un EI est défini comme un évènement fâcheux qui peut survenir lors d'un traitement avec un produit pharmaceutique, mais dont le lien de causalité n'a pas été établi. Les notifications peuvent être rapportées par des praticiens de santé ou des consommateurs auprès des autorités sanitaires, ou d'autres organisations compétentes. Généralement, une notification contient des informations concernant le patient, le produit suspecté (nom, posologie, date de début de traitement, etc...) et le ou les effets indésirable(s) constaté(s) (date d'apparition, évolution, etc...). De nombreuses agences nationales et supranationales comme l'Agence américaine des produits alimentaires et médicamenteux (*Food and Drug Administration*, FDA), l'Agence européenne des médicaments (*European Medecines Agency*, EMA), l'Observatoire d'Uppsala (*Uppsala Monitoring Center*, UMC) en charge de la pharmacovigilance pour l'Organisation Mondiale de la Santé (OMS), disposent de tels systèmes de notifications spontanées. En France, la base nationale de pharmacovigilance (BNPV) est gérée par l'Agence Nationale de Sécurité du Médicament et des Produits de Santé. Cette base de données est constituée des notifications que doivent rapporter les praticiens de santé auprès des 31 centres régionaux de pharmacovigilance partenaires. Construite en 1979, la BNPV a été mise à jour en 1985-1986 pour permettre les notifications en ligne. Sa structure a été modifiée en 1994-1995 afin de comprendre de nouveaux champs d'information (THIESSARD *et al.*,

2005). Le nombre de notifications enregistrées annuellement a fortement augmenté ces dernières années avec environ 21 000 notifications renseignées en 2010 contre 61 000 en 2017.

Plusieurs outils de détection automatisée de signaux ont été mis au point pour exploiter ces grands ensembles de données afin de mettre en évidence des couples (médicament, EI) suspects. Les signaux ainsi générés doivent être étudiés de manière plus approfondie par des experts ou faire l'objet d'études complémentaires pour être confirmés. Ce travail étant très chronophage, il est important de générer un nombre raisonnable de signaux avec le moins de fausses associations possibles pour cette étape d'analyse supplémentaire.

Les méthodes de détection utilisées en routine par les agences de santé publique sont appelées méthodes de disproportionnalité. Elles reposent sur une représentation agrégée des données de notifications spontanées, c'est-à-dire sous la forme d'une table de contingence qui a autant de lignes que de médicaments différents renseignés et autant de colonnes que d'EI différents renseignés. Chaque cellule contient le nombre de notifications correspondant au couple (médicament, EI) considéré. Une même notification pouvant mentionner plusieurs médicaments et plusieurs EI, elle peut donc entrer dans les comptages de plusieurs cellules de la table de contingence. Comme un grand nombre de couples (médicament, EI) ne sont jamais notifiés, la table de contingence est extrêmement creuse, c'est-à-dire que beaucoup de ses cellules sont à zéro.

Ces méthodes de disproportionnalité reposent sur une stratégie en trois étapes qui consistent, pour chaque couple (médicament, EI), à (i) déterminer le nombre de notifications impliquant ce couple ; (ii) construire une mesure qui quantifie l'écart entre ce nombre et le nombre de notifications attendu sous l'hypothèse d'indépendance entre le médicament et l'EI ; et (iii) décider qu'un signal est émis si cette mesure dépasse une certaine valeur seuil. Cette procédure est conduite simultanément pour tous les différents couples (médicament, EI) (Ahmed *et al.*, 2015).

On peut distinguer deux grandes familles de méthodes de disproportionnalité : les méthodes fréquentistes telles que le *Reporting Odds Ratio* (ROR) (van Puijenbroek *et al.*, 2002) ou le *Proportional Reporting Ratio* (PRR) (Evans *et al.*, 2001) utilisé par l'EMA, et les méthodes bayésiennes comme le *Gamma Poisson Shrinker* (Dumouchel, 1999) et le *Multi-item Gamma Poisson Shrinker* utilisé par la FDA (Szarfman *et al.*, 2004),

ou le *Bayesian Component Propagation Neural Network* (BCPNN) (Bate *et al.*, 1998) utilisé par l'UMC jusqu'en 2014 (Norén *et al.*, 2006; Caster *et al.*, 2017). A présent, l'UMC utilise pour la détection de signaux vigiRank : un modèle prédictif qui tient compte d'un ensemble d'informations caractérisant la notification d'un couple (médicament, EI), dont notamment la valeur de l'*Information Component*, la mesure de disproportionnalité issue de la méthode BCPNN (Caster *et al.*, 2014). Plus récemment, les méthodes de disproportionnalité ont été étendues pour prendre en compte la multiplicité des comparaisons effectuées et ainsi fournir des seuils de détection basés sur des estimations du *False Discovery Rate* (FDR) (Benjamini et Hochberg, 1995) et un classement alternatif des signaux (Ahmed *et al.*, 2009, 2010). D'autres méthodes qui reposent sur des tests du rapport de vraisemblance ou sur le *sequential probability ratio test* ont également été proposées (Huang *et al.*, 2011; Chan *et al.*, 2017; Ding *et al.*, 2020).

Le fait de considérer la forme agrégée des données implique cependant que les méthodes de disproportionnalité souffrent d'un biais appelé biais de masquage. Ce biais se produit lorsqu'un EI donné est particulièrement notifié avec un certain médicament. Il devient alors plus difficile de détecter un signal impliquant un autre médicament avec cet EI. De plus, ces méthodes ne tiennent pas compte de la coprescription des médicaments (Almenoff *et al.*, 2005; Harpaz *et al.*, 2012; Pariente *et al.*, 2012; Arnaud *et al.*, 2016).

1.3 Prise en compte de la confusion mesurée et régression pénalisée

Afin de pallier les biais dont souffrent les méthodes de disproportionnalité, des méthodes de détection qui considèrent la forme individuelle des données de notifications et qui reposent sur des modèles de régression logistique multiple ont été proposées récemment (Caster *et al.*, 2010; Harpaz *et al.*, 2013; Ahmed *et al.*, 2018). Comme la régression permet d'estimer l'effet d'une variable sur la réponse d'intérêt conditionnellement aux autres variables explicatives, la confusion mesurée due aux coprescriptions est prise en compte par cette technique. De plus, le terme d'intercept dans la régression logistique traduit la fréquence de la réponse d'intérêt lorsque toutes les variables expli-

catives incluses dans le modèle sont nulles. Cette fréquence de « fond » de la réponse d'intérêt est une estimation plus fine que celle calculée par les méthodes de disproportionnalité, puisqu'elle prend en compte l'effet d'un ensemble de prédicteurs, et empêche ainsi le masquage de certaines associations avec la réponse à cause d'associations plus fortes (Caster *et al.*, 2010). Néanmoins, l'utilisation d'une régression logistique dans le cadre de la grande dimension, où des milliers de variables explicatives sont présentes, est problématique. Elle peut entraîner des problèmes de convergence des algorithmes, des variances de coefficient estimées élevées, et des temps de calcul importants (Genkin *et al.*, 2007). Harpaz *et al.* (2013) propose d'opérer un filtrage des variables basé sur leurs fréquences de co-notification avec l'événement d'intérêt. Seules ces variables sont intégrées dans la régression logistique par la suite. Les autres méthodes de détection proposées ont recours à des régressions logistiques multiples pénalisées avec une pénalisation de type lasso.

La régression lasso (Tibshirani, 1996) consiste à maximiser la log-vraisemblance d'un modèle de régression, pénalisée par une fonction linéaire de la norme L_1 des coefficients de régression. En plus d'être adapté au contexte de la grande dimension, ce type de pénalisation permet de réduire à exactement zéro certains coefficients associés aux variables explicatives et ainsi d'opérer une sélection de variables. Le degré de pénalité appliqué est contrôlé par un paramètre classiquement noté λ. En pratique, le choix d'une valeur de λ « optimale » est une tâche ardue qui dépend principalement de l'objectif et de la modélisation. Différentes stratégies ont été proposées dans la littérature.

La validation croisée est la méthode la plus couramment employée (Stone, 1974). Pertinente quand l'objectif visé est un objectif de prédiction, elle consiste à estimer une erreur de prédiction pour chaque valeur de λ considérée en rééchantillonnant les données. Toujours dans un but de prédiction, des critères d'information utilisés pour la sélection de modèle comme le critère d'information d'Akaike (*Akaike Information Criterion*, AIC) (Akaike, 1973) ou le C_p de Mallows (Mallows, 1973) ont été définis dans le cadre de la régression pénalisée lasso. Dans une optique de sélection de variables, il a été proposé d'avoir recours au critère d'information bayésien (*Bayesian Information Criterion*, BIC) (Schwarz, 1978), un critère d'information bien connu pour choisir de manière consistante le vrai modèle. Tout comme pour l'AIC et le C_p de Mallows, le λ optimal est celui qui minimise la valeur du BIC. Toujours à des fins de sélection de variables, une autre

approche, appelée *stability selection*, a été proposée. Elle combine le sous-échantillonage par *bootstrap* et des algorithmes de sélection de variables tels que le lasso (Meinshausen et Bühlmann, 2010).

De nombreuses extensions du lasso ont été développées pour répondre à des problèmes de plus en plus complexes, et la littérature statistique dans ce domaine est foisonnante. Zou (2006) a proposé le lasso adaptatif (*adaptive lasso*), une extension du lasso qui possède de meilleures propriétés en termes de sélection de variables. En effet il a été montré que la sélection de variables effectuée par le lasso pouvait ne pas être consistante dans certaines situations (Fan et Li, 2001). Le lasso adaptatif pallie les faiblesses du lasso en introduisant des pénalités propres à chaque variable explicative.

En pharmacovigilance, l'objectif poursuivi est celui de la sélection de variables, dans la mesure où l'on souhaite identifier les médicaments réellement associés à l'EI étudié. En outre, on suppose que le nombre de ces médicaments soit relativement faible en comparaison du nombre de médicaments présents dans la base. Les méthodes de détection reposant sur le lasso ont abordé la question du choix de λ de différentes manières. Caster *et al.* (2010) ont choisi une valeur de λ identique pour l'ensemble des EI évalués, et ce dernier a été fixé de manière à générer un nombre de signaux égal à celui de la méthode de disproportionnalité BCPNN. Ahmed *et al.* (2018) ont développé la méthode *Class Imbalanced Subsampling Lasso* (CISL), une variation de *stability selection* adaptée au cas où la réponse d'intérêt est rare, comme cela peut être le cas en pharmacovigilance. Les seuils de détection proposés dans ce travail ont été établis sur la base d'estimations du FDR obtenues à partir de résultats de simulations s'appuyant sur des données réelles.

1.4 Intérêt des bases médico-administratives pour la pharmacovigilance

En plus des avancées méthodologiques pour la détection de signaux sur les bases de notifications spontanées, il y a eu ces dernières années un intérêt croissant pour l'exploitation des bases médico-administratives (BMA) pour répondre à des questions de recherche en santé, et en particulier en pharmacovigilance. Ces bases de données contiennent des

informations de santé collectées à des fins administratives, comme le remboursement des consommations de soins. La représentativité, la taille des populations couvertes, leur exhaustivité et les nombreuses informations longitudinales disponibles pour chaque individu sont autant d'arguments à l'avantage de ces bases de données. Ainsi, elles pourraient permettre de pallier certains biais inhérents aux bases de pharmacovigilance comme la sous notification (Bégaud *et al.*, 2002), ou le manque de représentativité due à une probabilité de notification qui varie notamment en fonction de la gravité de l'EI, du délai de commercialisation de certains médicaments ou de leur « notoriété » (Weber, 1984).

En France, le Système National des Données de Santé (SNDS) couvre la quasi-totalité de la population française et est maintenu par la Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés. Il contient entre autres des informations précises relatives aux séjours hospitaliers et aux remboursements des différentes consommations de soins en ville, dont les remboursements de médicaments. A partir des données du SNDS a été créé l'Échantillon Généraliste des Bénéficiaires (EGB) : un échantillon représentatif au $1/97^{\text{ème}}$ de la population française plus facilement accessible à la communauté des chercheurs et dont les données sont mises à jour trimestriellement pour les consommations de soins en ville, et annuellement pour les hospitalisations.

Bien que le potentiel des BMA pour la surveillance en pharmacovigilance soit immense, l'exploitation des ces données, qui ne sont pas collectées pour faire de la surveillance biomédicale, est un défi méthodologique. Contrairement aux données de notifications, aucune expertise médicale n'intervient dans le recueil des données. Un projet de grande ampleur mené par le consortium *Observational Medical Outcomes Partnership* (OMOP) a vu le jour aux États-Unis avec pour objectif d'évaluer les performances des méthodes disponibles pour la détection de signaux en pharmacovigilance. L'OMOP a ainsi publié dans un numéro spécial du journal scientifique *Drug Safety* (Ryan *et al.*, 2013c; Ryan et Schuemie, 2013) les résultats d'une étude qui consistait à comparer un vaste panel de méthodes. Ces dernières étaient des méthodes de disproportionnalité ou étaient dérivées d'approches classiquement utilisées en pharmacoépidémiologie, comme le design cas-témoins, celui de cohorte ou encore la série de cas (*Self-Controlled Case Series*, SCCS), une méthode autocontrôlée basée sur les cas uniquement (Farrington, 1995). Les résultats de cette vaste comparaison ont permis de mettre en avant les bonnes performances de cette dernière ap-

proche. En plus de l'implémentation classique de la SCCS qui ne considère qu'une seule exposition médicamenteuse comme variable explicative, une extension de cette méthode qui permet de prendre en compte les coprescriptions a été mise en œuvre : la *Multiple Self-Controlled Case Series* (MSCCS). De par la grande dimension des données considérées, des régressions pénalisées ont été utilisées. L'OMOP a eu recours a une pénalisation de norme L_2 où le paramètre de pénalité était sélectionné par validation croisée, mais il a été proposé une version de la MSCCS avec une pénalité de type lasso (Simpson *et al.*, 2013), où λ était également sélectionné par validation croisée.

Une alternative aux régressions multiples dans la prise en compte des coprescriptions a également été testée par l'OMOP pour la détection de signaux : il s'agit du score de propension (*propensity score*, PS) en grande dimension (Ryan *et al.*, 2013a). Cette méthode consiste à sélectionner de manière automatique des variables à inclure dans la construction d'un score résumé qui mesure la propension à être traité par un médicament d'intérêt, et à utiliser ce score dans l'analyse (Schneeweiss *et al.*, 2009). Le PS en grande dimension est de plus en plus employé pour la conduite d'études pharmacoepidémiologiques à partir de BMA. Son utilisation à des fins de détection automatique de signaux fait l'objet de recherche récentes (Demailly *et al.*, 2020).

Par ailleurs, plusieurs méthodes de détection de signaux ont été développées afin d'exploiter la complémentarité des BMA et des notifications spontanées. Ainsi, à partir de données américaines, Harpaz *et al.* (2013) ont mis en œuvre des méthodes de disproportionnalité dans les deux différentes sources de données et ont défini un signal s'il était généré à la fois dans les bases de notifications spontanées et dans les BMA. En évaluant manuellement la pertinence des signaux générés, ils ont montré que cette stratégie était plus performante qu'une détection réalisée sur la base des notifications spontanées seule. Dans leur travail, Pacurariu *et al.* (2015) ont nuancé ce constat. A partir de données européennes, ils ont également mis en œuvre des méthodes de disproportionnalité sur une BMA et une base de notifications spontanées. Il ont considéré les signaux générés uniquement dans les notifications spontanées, uniquement dans la BMA et dans les deux bases de données respectivement. Pour chacun de ces signaux ils ont procédé à une revue de la littérature afin de se prononcer sur leur pertinence. En comparant les performances qui découlent des trois ensembles de signaux distincts, ils arrivent à la conclusion que l'apport

des BMA pour la détection dépend principalement de la fréquence (taux d'incidence) de l'évènement d'intérêt. Li *et al.* (2015) ont proposé une stratégie plus élaborée qui consiste à combiner les p-valeurs obtenues à partir d'une détection dans les deux types de bases de données. Ces p-valeurs sont déterminées à partir de tests d'hypothèse unilatéraux où la statistique de test est estimée en ajustant sur des variables sélectionnées par le lasso comme associées à l'utilisation du médicament. A l'aide d'un ensemble de signaux de référence déterminé dans le cadre du projet OMOP, ils ont pu établir que leur stratégie de combinaison de signaux était plus performante qu'une détection dans les notifications spontanées seules, à partir du moment ou la BMA considérée est d'une taille suffisante.

Les BMA sont une source d'informations prometteuse pour la pharmacovigilance. Les efforts de recherche méthodologique dans ce domaine en sont encore à leurs débuts. Exploiter de manière conjointe les BMA et les notifications spontanées semblent être une stratégie pertinente pour tirer parti des avantages de ces deux types de données.

1.5 Objectifs de la thèse

Que ce soit pour les bases de notifications spontanées ou les BMA, le développement de nouvelles méthodes de détection de signaux est essentiel pour améliorer la réactivité et l'efficacité des systèmes de surveillance en pharmacovigilance. Il s'agit de mettre au point des outils statistiques qui génèrent un nombre raisonnable de signaux à analyser par des experts, tout en fournissant une liste de signaux pertinents avec le moins de fausses associations possible.

L'objectif général de cette thèse est double. Le premier objectif est de proposer et évaluer de nouvelles méthodes statistiques pour la détection de signaux sur les données de notifications spontanées. Deux directions ont été prises. L'une porte sur la proposition d'approches développées pour l'analyse des BMA : l'utilisation du score de propension en grande dimension. L'autre prolonge les développements méthodologiques autour de la sélection de variables avec le lasso et propose d'exploiter le lasso adaptatif. Ces deux propositions méthodologiques font l'objet des chapitres 2 et 3 de ce manuscrit.

Le deuxième objectif, faisant l'objet du chapitre 4, vise à proposer des stratégies de détection exploitant les deux sources de données que sont les notifications spontanées et

les BMA. Nous avons, dans un premier temps, évalué les performances d'une détection basée sur l'EGB à partir d'une étude empirique. Dans un second temps, nous avons considéré une approche basée sur le lasso adaptatif afin d'intégrer dans la détection sur les notifications spontanées l'information apportée par un groupe contrôle constitué dans l'EGB.

L'évaluation de nouvelles méthodes de détection de signaux peut être conduite sur la base d'études de simulations ou d'études empiriques sur données réelles. Nous avons ici utilisé ces deux approches. Dans le premier cas, la difficulté consiste à proposer un modèle de simulation qui rende compte de la complexité des données. Dans le second cas, la difficulté réside dans le fait de devoir évaluer la pertinence d'un grand nombre de signaux générés. Cette évaluation, en plus d'être très chronophage, nécessite une expertise pharmacologique pointue. Cette difficulté a été contournée ces dernières années par la constitution d'ensembles de référence, c'est-à-dire d'ensembles de couples (médicament, EI) avec un lien avéré. Pour permettre l'évaluation quantitative de méthodes de détection de signaux, il est nécessaire que ces ensembles soient de taille conséquente. Parmi les ensembles de référence souvent utilisés dans littérature, on trouve celui établi par l'OMOP qui contient 400 couples en lien avec quatre événements indésirables : les lésions hépatiques aiguës, les lésions rénales aiguës, les infarctus du myocarde aigus et les hémorragies gastro-intestinales (Ryan *et al.*, 2013b). Les études empiriques conduites tout au long de ce travail de thèse se sont appuyées sur un autre ensemble de référence établi plus récemment et relatif aux lésions hépatiques d'origine médicamenteuse (*Drug-Induced Liver Injury*, DILI) (Chen *et al.*, 2011, 2016).

Chapitre 2

Score de propension en grande dimension pour la détection de signaux à partir des notifications spontanées

2.1 Introduction

Afin de pallier certains biais inhérents aux méthodes de disproportionnalité, des approches de détection de signaux qui s'appuient sur la forme individuelle des données de notifications spontanées ont été proposées. Les données se présentent sous la forme de deux matrices binaires qui ont toutes deux le même nombre de lignes, égal au nombre de notifications enregistrées, et respectivement autant de colonnes que de médicaments ou d'évènements indésirables (EI) renseignés dans la base. Ces deux matrices sont de grande dimension et ont la particularité d'être extrêmement creuses. En effet, le nombre de médicaments et d'EI mentionnés dans une notification reste très faible comparé au nombre de médicaments et d'EI différents présents. Ainsi l'observation est une notification, les variables explicatives sont les nombreuses indicatrices de présence de médicaments et la réponse d'intérêt est la présence ou l'absence d'un EI donné. Une régression logistique multiple est alors utilisée pour régresser la réponse d'intérêt par rapport à toutes les variables médicament. Pour conduire une détection complète, il faut mener cette analyse

pour tous les EI présents dans la base de données.

Étant donné le nombre important de variables présentes, Caster *et al.* (2010) ont proposé d'utiliser une régression pénalisée de type lasso. Cette méthode est particulièrement adaptée au cadre de la grande dimension. De par le type de pénalité appliqué, elle permet d'obtenir des modèles parcimonieux en réduisant à exactement zéro certains coefficients de régression. Les signaux sont alors définis comme les couples formés par l'EI considéré et les variables ayant un coefficient de régression associé positif. Caster *et al.* (2010) ont fixé le degré de pénalisation de la régression, qui influe directement sur le nombre de variables ayant des coefficients non nuls et donc *a fortiori* positifs, en se comparant au nombre de signaux générés par une méthode de disproportionnalité. Ahmed *et al.* (2018) ont par la suite proposé une approche de détection basée sur une variation de *stability selection* (Meinshausen et Bühlmann, 2010) qui repose sur une procédure de sous-échantillonnage déséquilibré.

Une stratégie alternative pour faire face au grand nombre de variables présentes est d'avoir recours à la méthodologie du score de propension (PS). Cette méthode permet de résumer l'information dans un score de synthèse défini comme la probabilité d'être exposé à un médicament d'intérêt conditionnellement aux variables observées. Largement utilisé en pharmacoépidémiologie, il permet d'étudier l'association entre un EI et un médicament donné en se rapprochant des conditions d'expérience des essais randomisés. Les variables à inclure dans le modèle d'estimation du PS sont sélectionnées à partir de la littérature, ou de connaissances d'experts. Le but de cette sélection est d'inclure les facteurs de confusion qui interviennent dans la relation entre l'EI et le médicament, ainsi que les prédicteurs de l'EI. Une fois le score estimé, il est intégré dans le modèle de régression sur l'évènement d'intérêt. Il permet de réduire le biais dans l'estimation de la relation entre l'EI et le médicament considéré en prenant en compte la confusion mesurée.

Récemment, l'idée a émergé d'utiliser le PS dans l'exploitation de grandes bases de données de santé, en particulier des bases médico-administratives (BMA). Les BMA, qui n'ont pas été conçues pour répondre à des questions de recherche biomédicale, ne contiennent pas, *a priori*, tout ou partie des variables de confusion associées à l'exposition et à l'évènement d'intérêt. Ainsi l'hypothèse qui est faite derrière l'utilisation de PS dans ce cadre, qui est celui de la grande dimension, est que la multitude des informations

présentes permettent de mesurer indirectement ces facteurs de confusions. La principale difficulté dans la mise en œuvre de la méthode du PS en grande dimension réside dans la sélection des variables à inclure dans le modèle d'estimation du PS. Dans le cadre de la pharmacoépidémiologie, Schneeweiss *et al.* (2009) ont proposé l'algorithme du score de propension en grande dimension (*high-dimensional Propensity Score*, hdPS) pour sélectionner de manière automatique les variables à inclure dans le modèle. Par la suite, d'autre méthodes de sélection de variables et d'estimation du PS en grande dimension ont été proposées (Mccaffrey *et al.*, 2004; Franklin *et al.*, 2015, 2017; Ju *et al.*, 2019)

Dans le cadre de la pharmacovigilance, il s'agit de construire un score par exposition médicamenteuse considérée. Cette stratégie, encore peu explorée, a donné lieu à de récents développements méthodologiques dans le cadre des bases médico-administratives (Demailly *et al.*, 2020). Dans le contexte des données de notifications spontanées, Tatonetti *et al.* (2012) ont étudié l'utilisation d'une stratégie basée sur le PS et ont illustré son intérêt en la comparant à une méthode de disproportionnalité. Dans leur travail, les variables à inclure dans les scores étaient déterminées selon des critères empiriques.

Dans ce chapitre, nous présentons plusieurs méthodes basées sur le score de propension en grande dimension pour la détection de signaux à partir des données de notifications spontanées. Nous considérons quatre méthodes d'estimation du PS ainsi que trois stratégies d'intégration des scores estimés dans l'analyse de l'effet de l'exposition médicamenteuse sur l'EI d'intérêt. Nous comparons ces méthodes à des approches basées sur des régressions lasso. Cette étude comparative empirique est effectuée à partir une extraction de la BNPV sur la période 2000-2016 en utilisant un large ensemble de signaux de référence concernant un effet indésirable commun : *drug-induced liver injury* (DILI) (Chen *et al.*, 2011, 2016).

Dans un premier temps, nous présentons les méthodes de détection qui reposent sur la régression pénalisée lasso. Nous proposons également de formaliser une méthode de détection basée sur la régression lasso et qui repose sur le BIC. Dans un second temps, nous présentons la méthodologie du score de propension dans le cas classique puis dans le cas de la grande dimension. Enfin, nous présentons les méthodes de détection de signaux basées sur le score de propension en grande dimension que nous proposons. Les performances de toutes les méthodes de détection considérées sont comparées sur données réelles.

2.2 Notations

Sauf indications contraires, on notera N le nombre d'observations et P le nombre de variables explicatives, ce qui équivaut dans notre cas au nombre de médicaments différents présents dans la base de données. On définit les indices $i \in \{1, ..., N\}$ et $j \in \{1, ..., P\}$. On notera $\mathbf{X}$ la matrice de données binaires à N lignes et P colonnes qui contient les informations relatives aux expositions médicamenteuses renseignées dans la base de données. Dans la suite X_j fera référence à la $j^{\text{ème}}$ colonne de la matrice $\mathbf{X}$, qui indique l'absence ou la présence du médicament j. On notera $\mathbf{x}_j$ le vecteur de taille N qui correspond à la valeur observée de la variable aléatoire X_j sur l'ensemble des observations. De manière plus générale, un vecteur en gras sera un vecteur de taille N. On notera x_i le vecteur de taille P qui correspond aux valeurs observées des variables pour l'observation i. Enfin x_{ij} fera référence à la valeur observée de X_j pour la $i^{\text{ème}}$ observation.

La réponse d'intérêt traduit la présence ou l'absence d'un EI donné. Il s'agit d'une colonne de la matrice qui contient les informations relatives aux EI. Comme on ne considère qu'un EI d'intérêt, nous ne définissons pas cette matrice. Ainsi, on note Y notre réponse d'intérêt pour la désigner de manière générale, et $\mathbf{y}$ correspond à la valeur observée de Y sur les N observations.

2.3 Méthodes de détection basées sur des régressions logistiques pénalisées

2.3.1 Définition de la régression logistique lasso

Le modèle de régression logistique multiple de notre réponse d'intérêt est défini par :

$$\text{logit}(P(y_i = 1|x_i)) = \beta_0 + \sum_{j=1}^{P} \beta_j \; x_{ij}, \tag{2.1}$$

où β_0 est le terme d'intercept et $\beta \in \mathbb{R}^P$ le vecteur des coefficients de régression associés aux variables explicatives. Classiquement, la méthode du maximum de vraisemblance est utilisée pour estimer les paramètres (β_0, β). Bien que nous ne soyons pas dans la situation où P est grand devant N, P est très grand, ce qui peut causer des problèmes numériques

pour l'estimation des coefficients de régression.

La régression pénalisée lasso introduit une contrainte sur la norme L_1 des coefficients de régression. Elle consiste ainsi à estimer les coefficients de régression par

$$(\widehat{\beta_0}, \widehat{\beta})_\lambda = \text{argmax}_{(\beta_0,\beta)}\{l((\beta_0, \beta), \mathbf{y}, \mathbf{X}) - \text{pen}(\lambda)\}, \tag{2.2}$$

où l est la log-vraisemblance du modèle (2.1), et pen(λ) est définie par :

$$\text{pen}(\lambda) = \lambda|\beta|_1 = \lambda \sum_{j=1}^{P} |\beta_j|. \tag{2.3}$$

Dans la suite, on note $\widehat{\beta}_\lambda$ le vecteur de taille P dans $(\widehat{\beta_0}, \widehat{\beta})_\lambda$. Grâce à la pénalité (2.3), certains coefficients de $\widehat{\beta}_\lambda$ sont réduits à exactement zéro : les variables associées à ces coefficients ne sont pas « retenues » dans le modèle. Le paramètre λ permet de contrôler le degré de pénalisation qui influe directement sur le nombre de coefficients estimés non nuls : plus λ augmente, plus le nombre de composantes nulles dans $\widehat{\beta}_\lambda$ augmente. On notera λ_k, pour $k \in \{1, .., K\}$, une valeur de pénalisation testée dans la régression lasso avec $\lambda_1 < .. < \lambda_k < .. < \lambda_K$. Le vecteur de taille P des coefficients de régression estimés associé à λ_k sera noté $\widehat{\beta}_{\lambda_k}$. Un ensemble de variables dites « actives » est défini pour chacune de ces valeurs de pénalité :

$$\mathcal{S}(\lambda_k) = \{X_j : \widehat{\beta}_{\lambda_k,j} \neq 0;\ j = 1, ..., P\},$$

avec $\widehat{\beta}_{\lambda_k,j}$ le coefficient de régression associé à la variable X_j pour la valeur de pénalité λ_k, $k \in \{1, .., K\}$.

2.3.2 Le lasso logistique pour la détection de signaux : premiers travaux

Dans leur travail, Caster *et al.* (2010) ont développé une méthode de détection de signaux basée sur la régression lasso logistique. Ils ont mis en œuvre autant de régressions lasso que d'EI différents présents dans la base de données sur laquelle ils travaillaient (à savoir la base de pharmacovigilance de l'OMS), ce qui représentait plus de 2000 modèles

avec plus de 14 000 variables explicatives dans chaque modèle. Pour un EI donné, les signaux étaient définis comme les couples formés par cet EI et les variables appartenant à l'ensemble des variables actives associées positivement à l'EI. La valeur de pénalité λ, identique pour toutes les régressions lasso implémentées, a été fixée pour que le nombre total de signaux générés soit équivalent à celui obtenu par la méthode de disproportionnalité BCPNN.

2.3.3 *Stability selection* et *Class-Imbalanced Subsampling Lasso*

Pour contourner le problème du choix du paramètre λ, la méthode *stability selection* proposée par Meinshausen et Bühlmann (2010) dans le cadre plus général des algorithmes de sélection de variables, peut être utilisée. Cette méthode consiste dans un premier temps à perturber les données en les sous-échantillonnant plusieurs fois. Dans leur travail, Meinshausen et Bühlmann (2010) proposent d'implémenter un sous échantillonnage de taille $\frac{N}{2}$ et sans remise, afin de se rapprocher d'un échantillonnage de type *bootstrap* (qui est un échantillonnage aléatoire avec remise de taille égale au nombre d'observations), tout en étant computationnellement plus efficace. Une régression lasso est ensuite implémentée sur chacun de ces sous-échantillons. Pour chaque variable X_j, on calcule pour une valeur de pénalité donnée λ_k la probabilité de sélection définie par :

$$\widehat{\pi}_j^{\lambda_k} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}[\widehat{\beta}^b_{\lambda_k,j} \neq 0],$$

où B est le nombre de sous-échantillons considérés, et $\widehat{\beta}^b_{\lambda_k,j}$ le coefficient de régression associé à la variable X_j sur le sous-échantillon b pour la valeur de pénalité λ_k. Les variables retenues avec cette approche sont toutes les variables X_j telles que $\max_{\lambda_k \in \{\lambda_1,..,\lambda_K\}} \widehat{\pi}_j^{\lambda_k} \geq \pi_{thr}$. Meinshausen et Bühlmann (2010) discutent de la valeur seuil π_{thr} en fonction du contrôle du *Family-Wise Error Rate* (FWER).

Ahmed *et al.* (2018) ont proposé une variation de cette méthode pour prendre en compte le grand déséquilibre de la réponse binaire observée qu'il y a dans les bases de pharmacovigilance : l'algorithme *Class-Imbalanced Subsampling Lasso* (CISL). Ainsi dans CISL, les sous-échantillons sont déterminés de manière non équiprobable et avec remise pour permettre une meilleure représentation des individus qui ont experimenté l'EI d'in-

térêt. Tout comme pour *stability selection*, des régressions logistiques de type lasso sont implémentées sur chacun de ces B sous-échantillons. On note E la taille maximale des modèles étudiés avec le lasso logistique. La procédure CISL consiste à calculer la quantité :

$$\widehat{\pi}_j^b = \frac{1}{E} \sum_{\eta=1}^{E} \mathbb{1}[\widehat{\beta}_{\eta,j}^b > 0], \tag{2.4}$$

où $\widehat{\beta}_{\eta,j}^b$ est le coefficient de régression pénalisée associé à la variable X_j estimé sur le sous-échantillon b pour une valeur de pénalité menant à un ensemble de variables actives de taille η. La distribution empirique de $\widehat{\pi}_j^b$ est déterminée sur les B sous échantillons pour chaque variable. Une variable est sélectionnée avec la méthode CISL si un certain quantile de la distribution de $\widehat{\pi}_j^b$ est non nul. Dans leur travail, Ahmed *et al.* (2018) ont réalisé des études de simulation pour guider l'utilisateur sur le choix de ce quantile. Les couples définis par l'évènement d'intérêt et les variables sélectionnées avec CISL sont considérées comme des signaux puisque l'aspect délétère de l'association avec la réponse est pris en compte dans la condition de positivité sur $\widehat{\beta}_j$ dans (2.4).

2.3.4 Utilisation du BIC pour la sélection variables avec le lasso

Dans leurs travaux, Caster *et al.* (2010) précisent que la très grande dimension des données considérées (qui comprenaient plus de quatre millions de notifications) ne leur avait pas permis de sélectionner le paramètre λ par validation croisée. Cette méthode consiste à diviser le jeu de données en n_f sous-ensembles disjoints, ou *folds*, de même taille. Chaque *folds* est utilisé $n_f - 1$ fois pour estimer le modèle et une fois pour l'évaluer via une mesure de performance prédictive. Dans le contexte de la régression lasso, la validation croisée est effectuée pour tester les modèles correspondant aux différentes valeurs de λ considérées. Cette approche, majoritairement utilisée dans la littérature, est reconnue comme pertinente quand on cherche à répondre à un objectif de prédiction. Or, les enjeux de la pharmacovigilance s'inscrivent d'avantage dans un objectif de sélection de variables.

Nous avons voulu dans ce travail mettre en œuvre un critère de sélection de λ qui serait plus approprié à cet objectif et qui ne serait pas dépendant d'une autre méthode de détection. Ainsi, nous avons choisi de considérer le critère BIC, classiquement utilisé pour la sélection de modèles. Dans un premier temps, une régression lasso est implémentée en

considérant différentes valeurs du paramètre de pénalisation $\lambda_1, ..., \lambda_K$. Pour chacun des ensembles de variables actives, une régression non pénalisée est mise en œuvre. Ainsi, le BIC d'un modèle candidat issu de la régression lasso est défini comme suit :

$$\text{BIC}_k = -2\ l_k + \text{df}(\lambda_k)\ \ln(N), \tag{2.5}$$

où l_k est la log-vraisemblance du modèle de régression logistique non pénalisée qui comprend comme variables explicatives l'ensemble $\mathcal{S}(\lambda_k)$, et $\text{df}(\lambda_k) = |\mathcal{S}(\lambda_k)|$. Si des valeurs de pénalité différentes conduisent au même ensemble de variables actives, alors les modèles non pénalisés correspondants sont les mêmes, tout comme leur BIC. Ainsi, cette approche, que nous appellerons lasso-bic, sélectionne l'ensemble de variables actives qui conduit au modèle logistique non pénalisé minimisant le BIC défini en (2.5). On considère comme signaux les couples formés par l'évènement d'intérêt et les variables associées positivement à la réponse dans le modèle minimisant le BIC.

2.4 Méthodes de détection basées sur le score de propension en grande dimension

2.4.1 Méthodologie du score de propension

Définition

Le score de propension a été défini par Rosenbaum et Rubin (1983) comme la probabilité d'être exposé à un traitement donné conditionnellement aux variables observées. Il résume l'information sur le mécanisme d'attribution d'un traitement d'intérêt en un score. Initialement, le PS a été défini pour un traitement à deux niveaux d'exposition (traité et non traité). Conditionnellement au PS, l'exposition au traitement et les variables observées sont indépendantes. Sous les hypothèses de validité du PS que sont : la positivité, l'échangeabilité et la cohérence (présentées dans le tableau 2.1, (Cole et Hernan, 2008)), et à condition que le modèle d'estimation du score soit bien spécifié, le PS permet d'induire l'équilibre des variables observées entre le groupe des individus traités et non traités. Ainsi, le PS est utilisé dans le contexte des études observationnelles comme

un score d'équilibre qui permet aux essais non randomisés de reproduire les conditions d'une expérience randomisée en rendant les sujets exposés et non exposés comparables. Largement utilisé en pharmacoépidémiologie, il permet de réduire le biais dans le calcul de la relation entre une pathologie d'intérêt et le traitement considéré.

Le PS est inconnu et doit être estimé à partir des données, généralement à l'aide d'une régression logistique modélisant la probabilité d'être traité en fonction des facteurs de confusion supposés et mesurés. Chaque patient se voit alors attribuer une probabilité prédite d'être traité, calculée à partir du modèle de régression.

Positivité (*positivity*)	Il y a des individus exposés et non exposés à tous les niveaux des facteurs de confusion
Echangeabilité (*exchangeability*)	Il n'y a pas de facteurs de confusion non mesurés
Cohérence (*consistency*)	Le devenir potentiel d'un individu, s'il recevait l'exposition effectivement observée, est le devenir observé

Tableau 2.1 – Hypothèses de validité pour l'utilisation du score de propension (Cole et Hernan, 2008).

Construction du score de propension

Il est recommandé d'inclure dans le modèle d'estimation du PS les variables qui sont associées à la réponse d'intérêt et à l'exposition (les facteurs de confusion) ainsi que celle associées uniquement à la réponse, appelées prédicteurs. En revanche, il est conseillé de ne pas inclure de variables instrumentales, c'est-à-dire les variables associées uniquement à l'exposition (Brookhart *et al.*, 2006). En pratique, il est souvent recommandé de faire une sélection des variables à inclure dans le modèle d'estimation du PS à partir de connaissances d'experts ou à partir de la littérature.

Utilisation du score de propension

Une fois le PS estimé, il existe quatre familles de méthodes décrites dans la littérature pour prendre en compte le score dans l'analyse de l'effet du médicament sur la réponse (Austin, 2011) :

- l'ajustement (Vansteelandt et Daniel, 2014),

- la stratification (Rosenbaum et Rubin, 1984),
- l'appariement (Austin, 2009; Abadie et Imbens, 2016),
- la pondération sur le PS (Lunceford et Davidian, 2004).

L'ajustement sur le PS consiste à considérer le score comme une variable d'ajustement dans l'analyse. Le PS est considérée comme une variable explicative en plus de l'exposition dans la régression sur la réponse d'intérêt. Dans la stratification sur le PS, la population est divisée en plusieurs groupes selon les quantiles de la distribution du score. Classiquement, les quintiles de la distribution du score sont utilisés comme seuils pour définir ces groupes afin d'avoir cinq strates de tailles identiques. L'analyse de l'effet de l'exposition est ensuite effectuée parmi ces sous-groupes, puis les estimations de l'effet du traitement sont regroupées via une moyenne pondérée. L'appariement sur le score de propension consiste à apparier un ou plusieurs individus non traités qui ont des valeurs proches de PS estimé. L'appariement le plus fréquemment implémenté est un appariement sans remplacement (un individu ne peut être apparié qu'une fois) 1:1, où l'individu exposé et l'individu non exposé ont leur PS compris dans un même intervalle de largeur égal à 0,2 fois l'écart type du PS. Il a été également étudié la possibilité d'apparier des cas (individu présentant l'évènement d'intérêt) avec des témoins (individus qui ne présentent pas l'évènement d'intérêt) sur le PS (Pfeiffer et Riedl, 2015). La pondération sur le PS permet de construire une pseudo population où les caractéristiques entre les sujets traités et non traités sont équilibrées. Chaque individu se voit attribuer un poids déterminé en fonction de son PS estimé. Ce poids est ensuite intégré dans la régression modélisant le lien entre l'EI et l'exposition médicamenteuse, il mesure la contribution de l'individu dans l'estimation de l'effet de l'exposition. La pondération la plus classiquement utilisée est celle de l'*Inverse Probability of Treatment Weighting* (IPTW) (Austin et Stuart, 2015).

2.4.2 Le score de propension en grande dimension

Avec l'émergence de l'exploitation de grandes bases de données de santé pour la recherche médicale, l'idée d'utiliser la méthodologie du score de propension dans le cas de la grande dimension a vu le jour. Ainsi, le score de propension en grande dimension a été initialement introduit dans le cadre des BMA par Schneeweiss *et al.* (2009). Ces bases

de données disposent de peu d'informations cliniques car elles n'ont pas été construites pour répondre à des questions de recherche. L'hypothèse qui est faite derrière l'utilisation du PS dans ce cadre est qu'à défaut de disposer de l'ensemble variables de confusion entre le médicament et l'évènement d'intérêt, le grand nombre de variables disponibles dans ces bases peuvent permettre de recouvrir l'information manquante.

Dans le cadre de la grande dimension, une sélection *a priori* des variables à inclure dans le modèle d'estimation de PS n'est pas envisageable. Dans leur travail, Schneeweiss *et al.* (2009) présentent l'algorithme *high-dimensional Propensity Score* (hdPS) qui permet de sélectionner ces variables de manière automatique. Cette sélection repose sur une évaluation empirique de la prévalence des variables candidates qui sont transformées en variables binaires. L'algorithme ordonne les variables selon leur « potentiel de confusion » défini à partir de la formule de Bross (1966). En reprenant les notations de Schneeweiss *et al.* (2009), on défini la quantité :

$$\text{Bias}_{\text{M}} = \begin{cases} \dfrac{P_{C1}(RR_{CD}-1)+1}{P_{C0}(RR_{CD}-1)+1} & \text{si } RR_{CD} \geq 1 \\ \dfrac{P_{C1}\left(\frac{1}{RR_{CD}}-1\right)+1}{P_{C0}\left(\frac{1}{RR_{CD}}-1\right)+1} & \text{si } RR_{CD} < 1, \end{cases}$$

avec P_{C0} et P_{C1} la prévalence de la variable candidate chez les non exposés et chez les exposés respectivement, et RR_{CD} le risque relatif entre la variable candidate et l'évènement d'intérêt. Les variables sont ordonnées selon $\log(\text{Bias}_M)$ dans l'ordre décroissant. L'enquêteur doit choisir les ν variables les mieux classées. Dans leur travail, Schneeweiss *et al.* (2009) propose d'inclure ces ν variables dans le modèle d'estimation du PS et d'utiliser une régression logistique pour estimer le PS. ν est alors choisi pour être un nombre raisonnable de variables à inclure dans une régression logistique classique. Pour une valeur de ν plus grande, il a été proposé d'estimer le PS avec des algorithmes de *machine learning* (Demailly *et al.*, 2020; Volatier *et al.*). L'algorithme hdPS a aussi été utilisé pour sélectionner des variables à inclure directement dans une régression logistique pénalisée sur l'évènement d'intérêt (Franklin *et al.*, 2015).

La régression lasso a également été utilisée afin de sélectionner les variables à inclure dans un PS en grande dimension. La réponse d'intérêt est alors l'exposition, et le paramètre λ est choisi par validation croisée (Franklin *et al.*, 2017; Schneeweiss *et al.*,

2017). L'estimation du PS est ensuite faite à partir des coefficients de régression obtenus par maximisation de la vraisemblance pénalisée.

Des méthodes de *machine learning* ont été proposées comme des alternatives intéressantes à la régression logistique pour l'estimation des PS (Mccaffrey *et al.*, 2004; Setoguchi *et al.*, 2010; Lee *et al.*, 2010; Westreich *et al.*, 2010). En particulier, les arbres de régression boostés ont été identifiés comme une méthode potentiellement prometteuse. Le *boosting* est une technique ensembliste qui peut être appliquée à de nombreuses méthodes d'apprentissage statistique (Schapire, 1990). Elle consiste à agréger un ensemble d'apprenants (*learner*) faibles construits séquentiellement sur des observations pondérées, ces poids successifs étant fixés sur la base de l'erreur de prédiction ou de classification de l'itération précédente. Les observations mal prédites se voient attribuer un poids plus important, à l'inverse des observations bien prédites. Ainsi, les futurs apprenants faibles se concentrent davantage sur les observations mal prédites aux itérations précédentes. Dans le cas des arbres de décision, ces apprenants sont typiquement des arbres de régression ou de classification construits avec peu de variables. Plusieurs paramètres sont à prendre compte dans la mise en œuvre des arbres de régression boostés comme le nombre d'arbres, la profondeur de chaque arbre (c'est-à-dire leur nombre de nœuds) et le taux d'apprentissage ζ de l'algorithme.

Il a aussi été proposé d'avoir recours à l'algorithme du *Super Learner* pour estimer les PS (Ju *et al.*, 2019). Cet algorithme permet de créer une combinaison pondérée de nombreux algorithmes de *machine learning* pour construire l'estimateur optimal en termes de minimisation d'une fonction de perte spécifiée. Dans le travail de Ju *et al.* (2019), les paramètres de la combinaison des différents algorithmes sont estimés par validation croisée, et la fonction de perte considérée est l'opposée de la log-vraisemblance. En plus de la régression lasso, Ju *et al.* (2019) ont ajouté hdPS à la liste des algorithmes d'apprentissage considérés par le *Super Learner*.

2.4.3 Score de propension en grande dimension et pharmacovigilance

Appliquée dans le domaine de la pharmacovigilance, la méthodologie du score de propension en grande dimension consiste à construire un PS pour chaque exposition médicamenteuse considérée. A notre connaissance, Tatonetti *et al.* (2012) sont les premiers à avoir mis en œuvre une stratégie de détection qui repose sur un PS en grande dimension dans le cadre des données de notifications spontanées. Dans leurs travaux, les variables à inclure dans le score d'une exposition d'intérêt étaient choisies parmi les autres médicaments et parmi les autres EI que celui considéré. Les variables étaient sélectionnées selon leur fréquence de co-notification avec le médicament d'intérêt. Une fois les scores estimés avec des régressions logistiques, les individus exposés et non exposés étaient appariés selon leur PS.

Construction des scores de propension en grande dimension

Dans ce travail, nous avons cherché à construire un PS pour chaque exposition médicamenteuse présente dans la base de données. Les variables à inclure dans le modèle de PS sont sélectionnées parmi les autres expositions médicamenteuses. Nous avons considéré quatre approches pour l'estimation de PS en grande dimension.

La première approche mise en œuvre repose sur l'algorithme hdPS. Pour une exposition, nous avons considéré les $\nu = 20$ premières variables sélectionnées par cette algorithme.

Nous avons également testé deux méthodes d'estimation des scores qui reposent sur la régression logistique lasso. Nous avons utilisé les méthodes de sélection de variables présentées respectivement en section 2.3.4 et 2.3.3 : lasso-bic et CISL. Avec lasso-bic, les variables retenues pour la construction du score sont toutes celles présentes dans le modèle qui minimise le BIC. Dans CISL, la condition sur $\widehat{\beta}_j$ dans (2.4) d'être strictement positif a été remplacée par une condition de non nullité. Toutes les variables dont le quantile à 10% de leur distribution est supérieur à zéro ont été sélectionnées.

Une fois les variables à inclure dans les PS en grande dimension obtenues avec hdPS, lasso-bic et CISL, les scores ont été estimés à l'aide de régressions logistiques multiples classiques.

La quatrième approche mise en œuvre pour l'estimation du PS repose sur l'utilisation d'arbres de régression boostés implémentés à l'aide de l'algorithme *gradient tree boosting*, (GTB) (Hastie *et al.*, 2009). Nous avons fixé le nombre d'arbres à 200, la profondeur maximale de ces arbres à 6 et la valeur de ζ à 0,1.

Les régressions lasso ont été implémentées à l'aide de la version 2.0-10 du package R `glmnet`. Nous avons limité le nombre maximum de variables dans la régression lasso à 150 quand le BIC est utilisé pour sélectionner λ et à 50 avec CISL, évitant ainsi des instabilités numériques dans l'estimation des scores avec la régression logistique. Les arbres de régressions boostées ont été implémentés à l'aide de la version 0.6–4 du package R `xgboost`.

Utilisation des scores de propension en grande dimension

Une fois les différents PS estimés, nous avons examiné deux méthodes de prise en compte des scores sur les quatre présentées dans la section 2.4.1. Nous avons d'abord procédé à un ajustement sur le score de propension. En notant $\widehat{e}_{ij}$ le PS estimé associé à l'exposition médicamenteuse j de l'individu i, nous avons implémenté le modèle de régression logistique suivant :

$$\text{logit}(P(y_i = 1|x_{ij}, \widehat{e}_{ij})) = \beta_0 + \beta_j x_{ij} + \widetilde{\beta}_j \widehat{e}_{ij}.$$

Nous avons implémenté la pondération IPTW, où le poids attribué à l'individu i selon le PS associé à X_j est défini par :

$$w_{ij}^{IPTW} = \frac{x_{ij}}{\widehat{e}_{ij}} + \frac{1 - x_{ij}}{1 - \widehat{e}_{ij}}.$$

Cette pondération va avoir tendance à donner un poids élevé aux individus non traités qui ont des valeurs de PS élevées et aux individus traités qui ont des valeurs de PS faibles.

Comme certaines expositions médicamenteuses sont très peu notifiées, et ont ainsi très peu de notifications en commun avec l'EI d'intérêt, apparier les individus sur le PS peut conduire à des pertes dommageables d'individus qui ont fait l'expérience à la fois de l'exposition médicamenteuse et de la réponse. En gardant cette contrainte à l'esprit,

nous avons implémenté un autre type de pondération qui tend à mimer un appariement sur le PS en recréant une pseudo population moyenne qui serait celle d'un appariement classique 1:1. Ce type de poids sont appelés *Matching Weights* (MW) (Li et Greene, 2013; Franklin *et al.*, 2017) et sont définis par :

$$w_{ij}^{MW} = \frac{\min(\widehat{e}_{ij}, 1 - \widehat{e}_{ij})}{x_{ij}\widehat{e}_{ij} + (1 - x_{ij})(1 - \widehat{e}_{ij})}.$$

Contrairement à l'appariement qui écarte les sujets non appariés, la pondération avec les MW n'exclut aucun individu « entièrement », à la place elle réduit le poids de certains d'entre eux. Cette pondération va avoir tendance à donner un poids faible aux individus traités qui ont des valeurs de PS élevées et aux individus non traités qui ont des valeurs de PS faibles. Les individus qui reçoivent des poids élevés avec la pondération IPTW se voient attribuer avec cette pondération un poids égal à un, la valeur maximale des MW.

Nous avons implémenté des régressions logistiques pondérées avec ces deux types de pondération, avec comme seule variable explicative l'exposition d'intérêt. La vraisemblance d'une telle régression est définie par :

$$L = \prod_{i=1}^{N} P(y_i = 1|x_{ij})^{w_{ij}^{\gamma} y_i} (1 - P(y_i = 1|x_{ij}))^{w_{ij}^{\gamma}(1-y_i)},$$

avec $\gamma = IPTW$ ou MW. Toutes les régressions logistiques non pénalisées ont été implémentées à l'aide de la version 0.3-2 du package R `speedglm`.

Correction de tests multiples

Pour toutes ces méthodes de détection basées sur le PS en grande dimension, la règle de décision pour générer un signal est celle appliquée dans le cadre des tests d'hypothèses. Pour tenir compte de la multiplicité des hypothèses testées, nous avons employé une procédure de correction de tests multiples qui a été proposée dans le cadre de la pharmacovigilance (Ahmed *et al.*, 2010). Cette procédure repose sur l'estimation du FDR à partir du *location based estimator* proposé par Dalmasso *et al.* (2005) et prend en compte le fait que les tests réalisés soient unilatéraux.

2.5 Cadre de la comparaison

Par abus de langage, on parlera de signaux pour désigner les variables qui sont sélectionnées avec les méthodes de détection présentées ci-dessus. Il est sous entendu que les signaux sont les couples formées par ces variables et l'évènement indésirable considéré.

2.5.1 Critères de détection

Nous avons implémenté la méthode CISL pour la détection de signaux en considérant les quantiles à 5% et 10% de la distribution de la quantité (2.4). Les approches de détection qui découlent de ces différentes valeurs de quantiles sont dans la suite désignées par CISL-5% et CISL-10%.

Le seuil de détection pour le FDR est fixé à 5% pour les méthodes basées sur le PS en grande dimension. Les termes adjustPS-hdPS, mwPS-hdPS, iptwPS-hdPS, adjustPS-BIC, mwPS-BIC, iptwPS-BIC, adjustPS-CISL, mwPS-CISL, iptwPS- CISL et adjustPS-GTB, mwPS-GTB, iptwPS-GTB font référence aux approches mises en œuvre selon la manière dont le PS a été estimé (lasso-bic, CISL, GTB ou hdPS) et selon la manière dont le score a été pris en compte dans l'analyse (ajustement, pondération MW, ou IPTW). Afin d'implémenter ces approches pour toutes les expositions renseignées, nous avons parallélisé nos calculs.

En plus des méthodes présentées ci-dessus, nous avons inclus dans la comparaison une approche dite univariée, qui peut être assimilée à une méthode de disproportionnalité, le ROR, afin de servir de niveau de référence. Cette approche repose sur la mise en œuvre de régressions logistiques univariées où Y est régressé sur X_j pour $j \in \{1, .., P\}$:

$$\text{logit}(P(y_i = 1|x_{ij})) = \beta_0 + \beta_j x_{ij}. \tag{2.6}$$

La règle de décision pour générer un signal avec cette méthode est la même que celle utilisée pour les méthodes de détection basées sur le PS. Dans la suite, le terme Univ servira à désigner cette méthode.

2.5.2 Évaluation des performances

Pour évaluer les performances de toutes les approches de détection considérées, nous avons comparé la capacité de chaque méthode à détecter les vrais signaux et à ne pas détecter les faux signaux à partir d'un ensemble de signaux de référence présenté en section 2.5.4. Ces performances sont mesurées en termes de sensibilité et de spécificité, de valeur prédictive positive (*Positive Predictive Value*, PPV) et de proportion de faux positifs (*False Discovery Proportion*, FDP). Ces critères sont présentés dans le tableau 2.2. Ils sont évalués uniquement à partir des signaux dont le statut réel est connu, ce qui n'est pas le cas de tous les signaux générés.

		Statut réel		
		Témoin positif	Témoin négatif	
Statut prédit	Prédit positif	Vrai positif	Faux positif	$\text{FDP} = \frac{\lvert\text{Faux positif}\rvert}{\lvert\text{Prédit positif}\rvert}$ $\text{PPV} = \frac{\lvert\text{Vrai positif}\rvert}{\lvert\text{Prédit positif}\rvert}$
	Prédit négatif	Faux négatif	Vrai négatif	
		$\text{Sensibilité} = \frac{\lvert\text{Vrai positif}\rvert}{\lvert\text{Témoin positif}\rvert}$	$\text{Spécificité} = \frac{\lvert\text{Vrai négatif}\rvert}{\lvert\text{Témoin négatif}\rvert}$	

Tableau 2.2 – Critères d'évaluation des performances des méthodes de détection.

Nous avons également évalué pour chaque approche sa capacité à fournir un classement pertinent des signaux générés afin de pouvoir comparer si, à effectifs de signaux générés égaux, une approche détecte plus de vrais positifs et moins de faux positifs par rapport à une autre. Ainsi, les signaux générés sont ordonnés selon un critère propre à chaque approche, puis on représente le nombre de vrais positifs et de faux positifs détectés en fonction du nombre de signaux générés. Avec cette représentation graphique, une méthode montre de bonnes performances si la courbe obtenue

- est proche de l'identité pour la détection de vrais positifs,
- est proche de l'axe des abscisses pour la détection de faux positifs.

Pour lasso-bic, les signaux sont rangés selon leur p-valeur dans le modèle de régression logistique classique qui minimise le BIC. Pour CISL-5% et CISL-10%, les signaux sont

ordonnés selon leur valeur de quantile 5% et 10% de la quantité (2.4). Pour toutes les approches basées sur le PS en grande dimension, ainsi que pour l'approche univariée, les signaux sont ordonnés selon les valeurs du FDR estimé.

2.5.3 Base française des notifications spontanées

Nous avons implémenté les différentes méthodes de détection de signaux présentées sur les données issues de la BNPV. La classification Anatomique, Thérapeutique et Chimique (ATC) sert à coder l'ensemble des médicaments. Elle repose sur cinq niveaux différents de classement hiérarchisés. Dans la BNPV, le niveau de classification le plus fin est utilisé : un code est alors composé de sept caractères qui permettent de préciser le système d'organe cible, les propriétés thérapeutiques, pharmacologiques et chimiques du produit. Les évènements indésirables sont codés selon la terminologie du dictionnaire médical des affaires réglementaires (*Medical Dictionary for Regulatory Activities*, MedDRA) qui comprend cinq niveaux de précision. Dans la BNPV le niveau de termes préférentiels (*Preferred Term*, PT) est utilisé, ce qui correspond au quatrième niveau de précision. Ce code à huit chiffres se rapporte à un symptôme, une maladie, un diagnostic, une indication thérapeutique, une intervention chirurgicale ou médicale, ou une caractéristique d'antécédent médical, social ou familial. Pour cette étude, nous avons exclu (i) la phytothérapie, l'homéothérapie, les compléments alimentaires, l'oligothérapie et les inhibiteurs enzymatiques ; (ii) les surdosages ou les erreurs de médication.

Nous avons travaillé sur une extraction de la BNPV sur la période allant du 1er janvier 2000 au 31 décembre 2016, ce qui représente 382 484 notifications avec 5 906 EI différents et 2 344 médicaments différents selon leurs classifications respectives. La majorité (60,93%) des notifications ne mentionnait qu'un seul médicament, et 56,24% ne mentionnaient qu'un seul EI. La valeur médiane du nombre de notifications pour les médicaments était de 33 (Q1 = 5 et Q3 = 179). La valeur médiane du nombre de notifications pour les EI était de 4 (Q1 = 1 et Q3 = 20).

2.5.4 Ensemble de signaux de référence

Afin d'établir les performances des méthodes de détection considérées, nous avons utilisé l'ensemble de signaux de référence constitué par Chen *et al.* (2011) et qui concerne l'évènement DILI : DILIrank. Cet ensemble a été construit en analysant les résumés des caractéristiques du produit (RCP) des médicaments qui ont été approuvés par la FDA. Ces RCP contiennent des informations relatives aux indications, posologies et effets indésirables connus ou suspectés des médicaments. Ces informations sont structurées dans différentes sections. La figure A.1 en annexe est un exemple de RCP de la FDA non remplie.

Une liste de mot-clés liés à l'évènement DILI a été déterminée. Si aucun mot-clé n'apparaissait dans le RCP d'un médicament, alors le médicament était considéré comme non associé à une DILI. En revanche, si un mot-clé était identifié alors le médicament était classé comme moyennement ou fortement associé à une DILI, en fonction de la section de de RCP dans laquelle le mot-clé apparaissait. Chaque mot-clé s'est vu également attribuer un niveau de sévérité lié à une DILI via un score allant de zero à huit : si un mot-clé avait un score de huit alors il était associé à une DILI grave. En 2016, cette classification a été raffinée en incluant des informations provenant d'autres sources de données afin d'évaluer la relation de cause à effet entre chaque médicament considéré et un événement DILI. Seuls les médicaments confirmés comme associés de manière causale à une DILI ont été retenus (Chen *et al.*, 2016). De manière générale, il y avait corrélation entre force d'association avec une DILI et gravité de l'évènement. Autrement dit, si un médicament était fortement lié à une DILI, alors il s'agissait très souvent d'une forme sévère de DILI. La table A.1 en annexe fournit la liste des médicaments étudiés. Pour chaque médicament, il est précisé son code ATC, la section du RCP dans laquelle un mot-clé apparaissait, la sévérité attribuée à ce mot-clé et le niveau d'association de ce médicament avec une DILI. Cette table est issue des travaux de Chen *et al.* (2011) et Chen *et al.* (2016). Dans la construction de l'ensemble de référence, nous avons considéré comme témoins négatifs les médicaments « non liés à une DILI », et comme témoins positifs les médicaments « fortement liés à une DILI ».

En collaboration avec Antoine Pariente et Franscesco Salvo de l'unité Inserm 1219 (Bordeaux Population Health, équipe Médicament et santé des populations, Bordeaux,

France), nous avons converti les mots-clés utilisés pour définir l'évènement DILI en codes PT de la classification MedDRA. Le tableau A.2 en annexe présente ces différents codes, leur signification ainsi que les concepts plus généraux auxquels ils sont rattachés, appelés requêtes standardisées dans le dictionnaire MedDRA. Au total, 91 codes PT ont été considérés comme traduisant un évènement DILI. Parmi ces codes, 54 étaient présents dans la BNPV. Chaque notification spontanée impliquant au moins un de ces 54 codes a été considérée comme une DILI déclarée, ce qui s'est traduit par 22 440 notifications de cas de DILI dans la BNPV.

Pour éviter les problèmes numériques dus à la très faible notification de certains médicaments, nous avons limité la comparaison aux médicaments qui ont plus de trois notifications en commun avec une DILI, et qui ont plus de dix notifications au total. Cette restriction appliquée, nous avons considéré 922 médicaments différents sur les 2344 médicaments présents à l'origine dans la BNPV. Sur ces 922 médicaments, notre ensemble de référence comprenait 90 témoins négatifs et 114 témoins positifs.

2.6 Résultats de la comparaison

2.6.1 Performances en termes de détection

Pour chaque méthode de détection, le tableau 2.3 présente le nombre de signaux générés, le nombre de vrais et faux positifs détectés en plus des critères définis en section 2.5.2. Toutes les statistiques résumées sont dérivées de l'ensemble de signaux de référence DILIrank.

Les approches adjustPS sont celles qui ont généré le plus de signaux comparées aux autres approches basées sur le PS : entre 273 signaux pour adjustPS-GTB et 310 pour adjustPS-hdPS. Toutes les approches basées sur la pondération IPTW ont conduit au plus petit nombre de signaux générés avec 35, 63, 70, 34 signaux pour iptwPS-BIC, iptwPS-CISL, iptwPS-GTB, et iptwPS-hdPS respectivement. Les approches qui reposent sur la pondération MW ont généré un nombre intermédiaire de signaux avec 147, 121, 136, et 139 signaux pour mwPS- BIC, mwPS-CISL, mwPS-GTB, et mwPS-hdPS respectivement. Les approches basées sur des régressions multiples pénalisées : lasso-bic, CISL-5% et CISL-10% ont généré entre 99 et 173 signaux. L'approche univariée est celle qui a généré le plus

grand nombre de signaux : 359.

Les approches basées sur des régressions lasso ont montré les FDP les plus faibles : CISL-5% et CISL-10% n'ont fait aucune fausse découverte et lasso-bic avait une FDP de 3%. Les approches mwPS ont eu des résultats similaires avec une FDP d'environ 2%, à l'exception de mwPS-GTB qui a montré une FDP supérieure à 5%. L'approche univariée a montré les plus mauvaises performances en termes de fausses découvertes avec une FDP à 18%. Les approches basées sur l'ajustement sur le PS ont montré des performances légèrement meilleures que l'approche univariée avec des FDP entre 15% pour adjustPS-BIC et 10% pour adjustPS-hdPS respectivement. Selon la méthode utilisée pour estimer les scores, les approches basées sur la pondération IPTW ont obtenu des résultats variés avec des FDP allant de 17% pour iptwPS-CISL à 6% pour iptwPS-hdPS.

Les approches adjustPS et Univ ont obtenu les meilleures performances en termes de sensibilité : entre 65% pour adjustPS-GTB et 75% pour Univ. Cependant ces approches ont les plus faibles valeurs de spécificité : entre 78% pour Univ et 89% pour adjustPS-hdPS. A contrario, les approches basées sur des régressions lasso et les approches mwPS ont montré les meilleures valeurs de spécificité : entre 96% pour mwPS-GTB et jusqu'à 100% pour CISL-5% et CISL-10%. Les approches basées sur les régressions multiples pénalisées avaient une sensibilité comprise entre 37% pour CISL-5% et 56% pour lasso-bic. Les approches mwPS avaient une sensibilité comprise entre 43% pour mwPS-CISL et mwPS-GTB, et 46% pour mwPS-hdPS. Les approches iptwPS avaient une bonne spécificité mais une très faible sensibilité, variant de 11% pour iptwPS-BIC à 22% pour iptwPS-GTB.

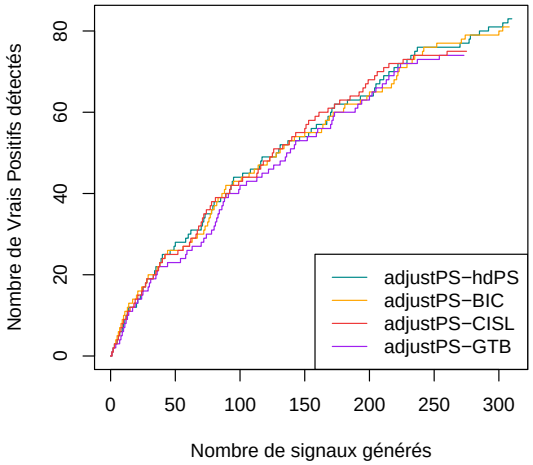
Méthode	Nombre de signaux	Nombre de signaux à statut connu (positif ou négatif)	Vrais signaux positifs			Faux signaux positifs		
			n	PPV (%)	Sensibilité (%)	n	FDP (%)	Spécificité (%)
Univ	359	105	86	81,90	75,44	19	18,10	78,89
lasso-bic	173	66	64	96,97	56,14	2	3,03	97,78
CISL-5%	99	43	43	100,00	37,72	0	0,00	100,00
CISL-10%	109	48	48	100,00	42,11	0	0,00	100,00
adjustPS-hdPS	310	93	83	89,25	72,81	10	10,75	88,89
mwPS-hdPS	139	54	53	98,15	46,49	1	1,85	98,89
iptwPS-hdPS	34	16	15	93,75	13,16	1	6,25	98,89
adjustPS-BIC	308	96	81	84,38	71,05	15	15,62	83,33
mwPS-BIC	147	53	52	98,11	45,61	1	1,89	98,89
iptwPS-BIC	35	14	13	92,86	11,40	1	7,14	98,89
adjustPS-CISL	275	86	75	87,21	65,79	11	12,79	87,78
mwPS-CISL	121	50	49	98,00	42,98	1	2,00	98,89
iptwPS-CISL	63	17	14	82,35	12,28	3	17,65	96,67
adjustPS-GTB	273	85	74	87,06	64,91	11	12,94	87,78
mwPS-GTB	136	52	49	94,23	42,98	3	5,77	96,67
iptwPS-GTB	70	28	25	89,29	21,93	3	10,71	96,67

TABLEAU 2.3 – Performances des approches basées sur le score de propension en grande dimension (dont le nom commence par adjustPS, mwPS ou iptwPS), les régressions pénalisées (lasso-bic, CISL-5%, CISL-10%) et de la méthodes univariée (Univ). Le nombre de vrais signaux positifs est le nombre de témoins positifs détectés par la méthode. Le nombre de faux signaux positifs est le nombre de témoins négatifs détectés par la méthode. PPV : Valeur Prédictive Positive, FDP : Proportion de Fausses Découvertes

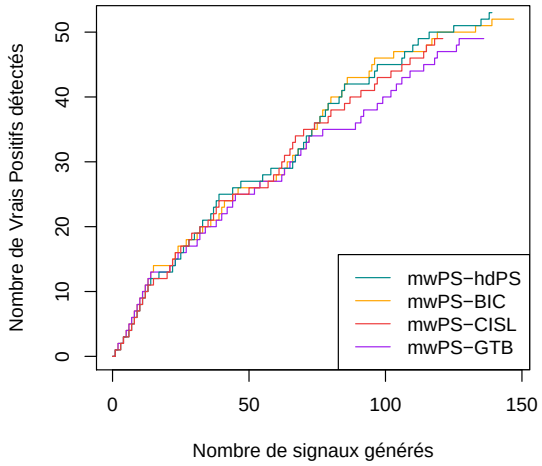
2.6.2 Performances en termes de classement des signaux

Les figures 2.1 et 2.2 représentent respectivement le nombre de vrais et de faux positifs détectés en fonction du nombre de signaux générés par les approches basées sur le PS en fonction de la méthode utilisée pour estimer le score. Les signaux dont le statut est inconnu d'après DILIrank sont pris en compte dans le nombre de signaux générés. Pour les approches adjustPS et mwPS, la manière dont les scores ont été estimés a peu influencé les résultats en termes de classement de signaux. Néanmoins, les scores estimés par GTB ont conduit à de moins bonnes performances pour le classement des vrais positifs pour ces deux familles d'approches : adjustPS-GTB a montré un classement moins pertinent à partir de 50 signaux générés environ (figure 2.1 (a)), et mwPS-GTB à partir de 80 signaux générés environ (figure 2.1 (b)). En ce qui concerne le classement des faux positifs, mwPS-GTB est l'approche qui a eu les moins bonnes performances (figure 2.2 (b)) en détectant des faux positifs pour un nombre inférieur de signaux générés par rapport à mwPS-hdPS, mwPS-BIC et mwPS-CISL. AdjustPS-GTB a montré un classement des faux positifs assez proche de celui d'adjustPS-CISL. Les classement fournis par ces méthodes ont été tous deux moins pertinents que ceux fournis par adjustPS-hdps et adjustPS-BIC (pour moins de 280 signaux générés, figure 2.2 (a)). Pour la pondération IPTW sur le PS, on observe une plus grande variabilité des performances entre iptwPS-BIC, iptwPS-CISL, iptwPS-GTB et iptwPS-hdPS. Contrairement à l'ajustement et à la pondération MW, c'est l'approche qui repose sur les scores estimés par GTB qui a classé le mieux les vrais positifs (figure 2.1 (c)). Les scores estimés par lasso-bic et hdPS montrent des performances assez proches. En revanche les scores estimés par CISL ordonnent le moins bien les vrais positifs. Concernant les faux positifs, iptwPS-CISL a montré le classement le moins pertinent en détectant un faux positif pour un nombre faible de signaux générés (figure 2.2 (c)). IptwPS-BIC, iptwPS-GTB et iptwPS-hdPS ont montré de meilleures performances. Ces trois approches ont fourni des classement semblables en détectant leur premier faux positif pour environ 30 signaux générés.

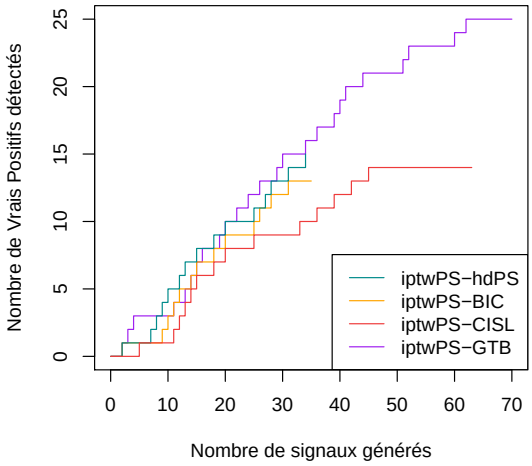
La figure 2.3 représente le nombre de vrais et de faux positifs détectés en fonction du nombre de signaux générés par les approches basées sur le PS, où les scores sont estimés avec lasso-bic. Sauf pour l'approche par pondération IPTW, dont les performances ont été décevantes tant en termes de classement de vrais positifs que de faux positifs, estimer



(a)

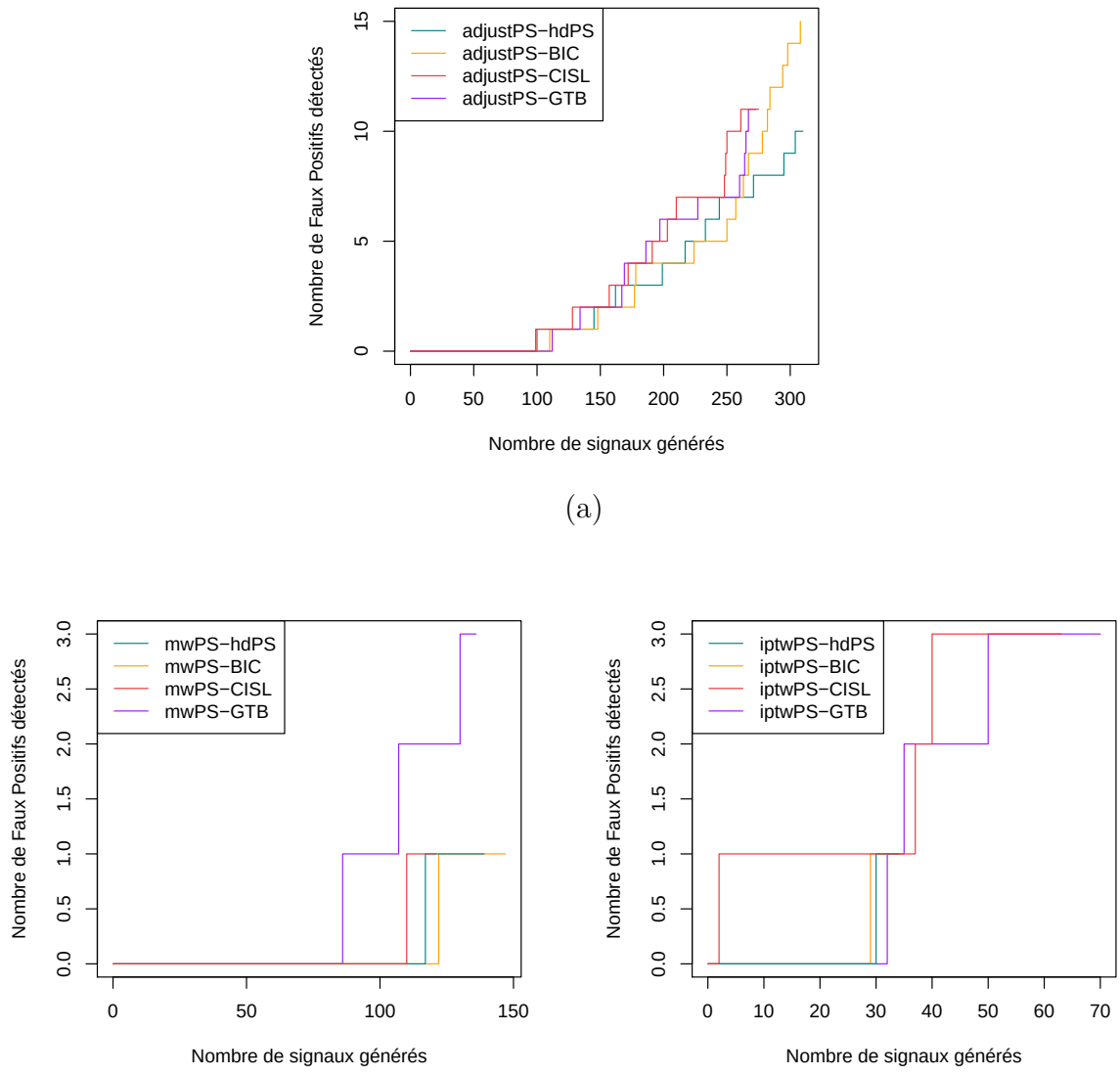


(b)



(c)

FIGURE 2.1 – Nombre de vrais positifs détectés en fonction du nombre de signaux générés par les approches basées sur le score de propension en grande dimension via (a) ajustement (adjustPS-hdPS, adjustPS-BIC, adjustPS-CISL, adjustPS-GTB) ; (b) pondération avec les poids *matching weights* (mwPS-hdPS, mwPS-BIC, mwPS-CISL, mwPS-GTB) ; (c) pondération avec les poids *inverse probability of treatment weighting* (iptwPS-hdPS, iptwPS-BIC, iptwPS-CISL, iptwPS-GTB) et pour différentes méthodes d'estimation des scores. Pour toutes ces approches, les signaux sont ordonnés selon les p-valeurs corrigées pour la multiplicité des tests.

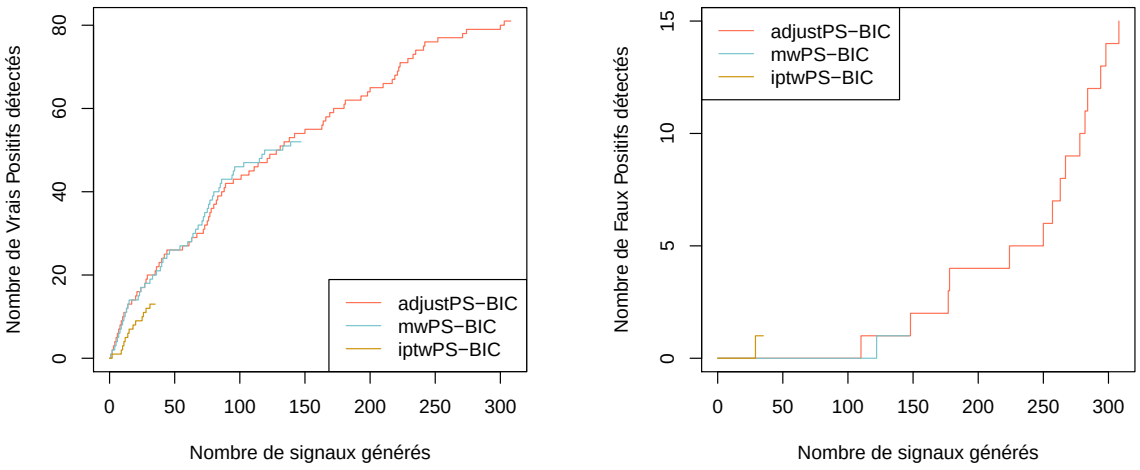


(a)

(b)

(c)

FIGURE 2.2 – Nombre de faux positifs détectés en fonction du nombre de signaux générés par les approches basées sur le score de propension en grande dimension via (a) ajustement (adjustPS-hdPS, adjustPS-BIC, adjustPS-CISL, adjustPS-GTB) ; (b) pondération avec les poids *matching weights* (mwPS-hdPS, mwPS-BIC, mwPS-CISL, mwPS-GTB) ; (c) pondération avec les poids *inverse probability of treatment weighting* (iptwPS-hdPS, iptwPS-BIC, iptwPS-CISL, iptwPS-GTB) et pour différentes méthodes d'estimation des scores. Pour toutes ces approches, les signaux sont ordonnés selon les p-valeurs corrigées pour la multiplicité des tests.



(a) (b)

FIGURE 2.3 – Nombre de (a) vrais positifs et (b) faux positifs détectés selon le nombre de signaux générés par les approches basées sur le score de propension : ajustement, pondération avec les poids *matching weights*, pondération *inverse probability of treatment weighting* où les scores sont estimés avec lasso-bic : adjustPS-BIC, mwPS-BIC, iptwPS-BIC.
Pour toutes ces approches, les signaux sont ordonnés selon les p-valeurs corrigées pour la multiplicité des tests.

le score avec lasso-bic a conduit à des classements pertinents pour les autres approches basées sur le PS. En comparant les performances d'adjustPS-BIC et mwPS-BIC à nombres égaux de signaux générés, on constate que les classement fournis par ces approches sont assez proches. Néanmoins, à partir de 60 signaux générés, mwPS-BIC a détecté plus de vrais positifs qu'adjustPS-bic. MwPS-BIC a détecté son premier faux positif pour un nombre de signaux générés légèrement plus important qu'adjustPS-BIC.

Les figures B.1, B.2 et B.3 en annexe représentent le nombre de vrais et de faux positifs détectés en fonction du nombre de signaux générés par les approches basées sur le PS où les scores sont estimés avec hdPS, CISL et GTB respectivement. Quelle que soit la méthode utilisée pour estimer les PS, les approches basées sur l'ajustement et sur la pondération MW ont montré des classements pertinents des signaux générés. A nombres égaux de signaux générés, les approches mwPS ont eu tendance a détecter quelques vrais positifs supplémentaires par rapport aux approches adjustPS. Les approches mwPS-hdPS et mwPS-CISL ont fourni un classement des faux positifs légèrement plus pertinent qu'adjust-hdPS et qu'adjust-CISL respectivement. En revanche, cela n'a pas été le cas lorsque le PS a été

estimé avec GTB : mwPS-GTB a détecté son premier faux négatif pour un nombre de signaux générés plus faible qu'adjustPS-GTB. Pour toutes les stratégies d'estimation du score, l'approche basée sur la pondération IPTW a montré des performances bien moins bonnes que ses concurrentes.

La figure B.4 en annexe montre les mêmes représentations graphiques de classement des signaux pour les approches basées sur des régressions lasso. Ces trois approches ont montré des performances assez similaires avec des classements pertinents des signaux générés. Pour environ 50 signaux générés, lasso-bic a montré néanmoins des performances légèrement meilleures en détectant plus de vrais positifs. Contrairement à CISL-5% et CISL -10%, lasso-bic a détecté des faux positifs. On peut cependant remarquer que son premier faux positif est apparu pour plus de 80 signaux générés.

Au vu des résultats exposés et dans un souci de clarté, nous avons décidé de ne considérer dans la suite que l'approche qui repose sur les poids MW avec des scores estimés avec lasso-bic (mwPS-BIC), lasso-bic et Univ.

2.6.3 Comparaison des approches mwPS-BIC, lasso-bic et Univ

Les figures 2.4 (a) et 2.4 (b) représentent respectivement le nombre de vrais positifs et de faux positifs détectés selon le nombre de signaux générés par ces trois approches. En ce qui concerne la détection de vrais positifs, lasso-bic a fourni le meilleur classement des signaux et Univ le pire. MwPS-BIC a eu des performances comparables à celle de lasso-bic. En ce qui concerne la détection de faux positifs, mwPS-BIC a obtenu les meilleurs résultats puisqu'il a détecté son premier faux positif pour un nombre de signaux générés plus important. Univ a généré beaucoup plus de signaux que mwPS-BIC et lasso-bic, mais son taux de détection des vrais positifs a été plus faible pour les derniers signaux générés. En effet, 36% de ses 150 premiers signaux générés étaient de vrais positifs, contre 15% de ses 209 derniers signaux. La même tendance a été observée pour la détection de faux positifs : Univ a détecté 2% de faux positifs dans ses 150 premiers signaux générés et environ 7% dans ses 209 derniers signaux.

Nous avons voulu regarder plus en détail les signaux générés par ces trois différentes méthodes de détection en se plaçant à nombres de signaux générés comparables. Comme

Univ, mwPS-BIC et lasso-bic ont généré respectivement 359, 173 et 147 signaux, nous avons considéré pour ces trois approches leurs 147 premiers signaux générés. La figure 2.4 (c) montre la répartition des 147 premiers signaux générés par Univ, lasso-bic et mwPS-BIC. La figure 2.4 (d) représente les effectifs de notifications observés des 147 premiers signaux générés par Univ, lasso-bic et mwPS-BIC par rapport aux effectifs attendus, en fonction de la méthode par laquelle ils ont été générés. Si on note N_j le nombre de notifications qui impliquent le médicament X_j et N_{DILI} le nombre de notifications qui impliquent un évènement DILI, l'effectif attendu du couple (X_j, DILI) sous l'hypothèse d'indépendance est défini par $(N_j \times N_{DILI})/N$. Parmi les signaux considérés, 111 signaux ont été générés par les trois approches. La figure 2.4 (d) montre que ces signaux sont caractérisés par un rapport de risque observé plus élevé (qui correspond au ratio de l'effectif observé sur l'effectif attendu) et/ou un plus grand support des données (càd des effectifs observés importants) en comparaison des signaux détectés uniquement par une approche. La figure 2.4 (d) montre également que les méthodes ont eu tendance à générer différents types de signaux : en particulier, lasso-bic a mis en évidence certains signaux avec un support relativement faible des données tandis que Univ a mis en évidence des signaux avec un faible risque observé mais un grand nombre de notifications. MwPS-BIC a eu un comportement intermédiaire.

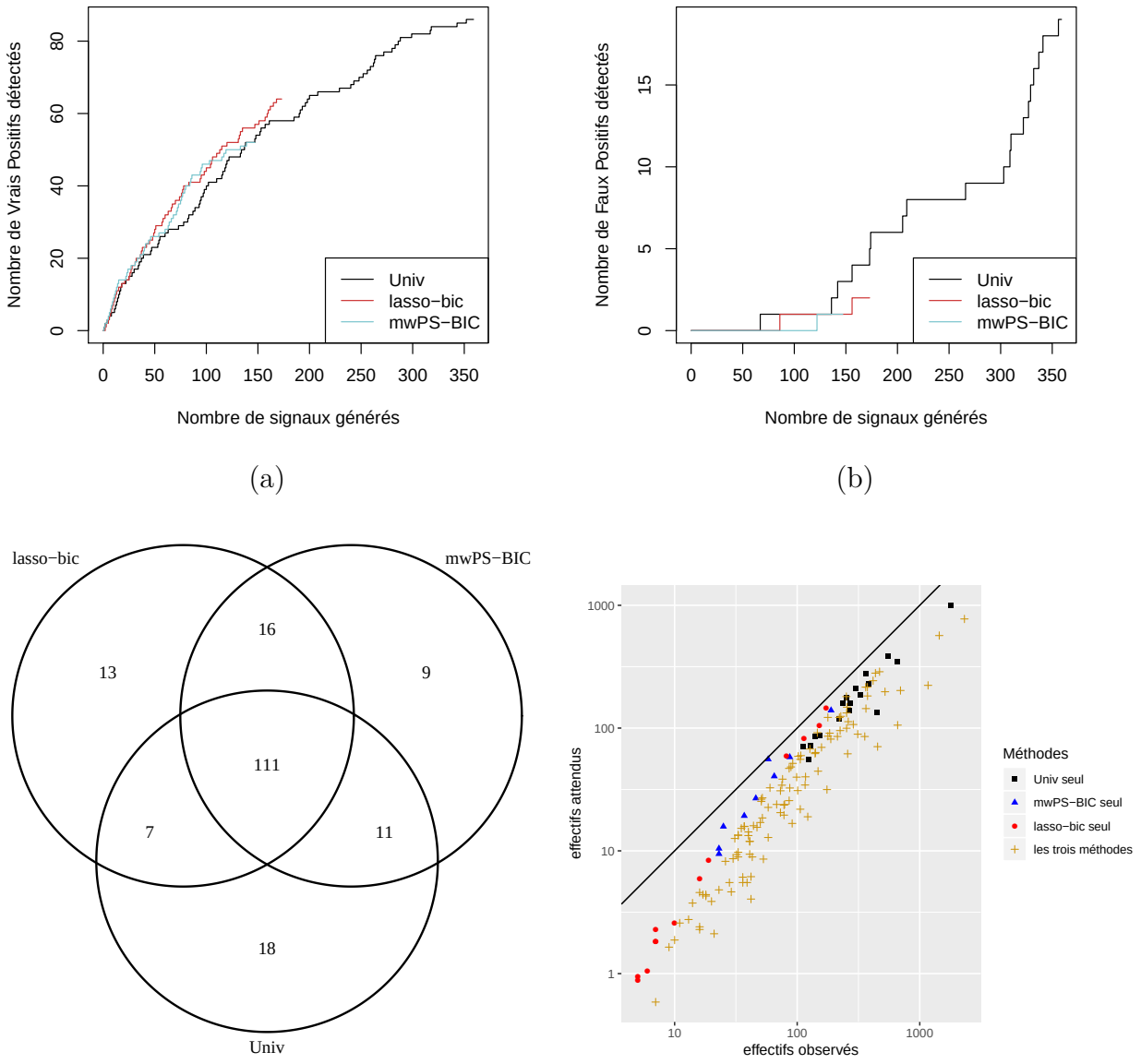


FIGURE 2.4 – Nombre de vrais positifs (a) et faux positifs (b) détectés selon le nombre de signaux générés par lasso-bic, mwPS-BIC et Univ.
Pour lasso-bic, les signaux sont ordonnés selon leur p-valeur dans le modèle de régression non pénalisée.Pour mwPS-BIC et Univ, les signaux sont ordonnés selon les p-valeurs corrigées pour la multiplicité des tests.
(c) Répartition des 147 premiers signaux générés par les méthodes Univ, lasso-bic et mwPS-BIC.
(d) Effectifs attendus en fonction des effectifs observés de notifications des signaux générés par Univ seul, lasso-bic seul, mwPS-BIC seul et par les trois méthodes, en considérant leurs 147 premiers signaux générés.

2.7 Discussion

Cette étude nous a permis d'explorer l'intérêt du score de propension en grande dimension dans le cadre de la détection de signaux en pharmacovigilance sur les bases de notifications spontanées. En considérant plusieurs approches pour la construction des scores de chaque médicament de la base de données, et plusieurs manières de prendre en compte ces scores dans les analyses, nous avons pu mettre en évidence les approches qui se sont révélées être les plus pertinentes pour la détection. Pour évaluer les performances de nos approches ainsi que des approches concurrentes basées sur des régressions pénalisées, nous avons utilisé un ensemble de signaux de référence relatif aux lésions hépatiques d'origine médicamenteuse : l'ensemble DILIrank.

L'approche basée sur le PS qui a montré les meilleurs résultats est celle qui s'appuie sur la pondération avec les poids MW. Ses performances en termes de détection des vrais/faux signaux sont très proches de celles des approches basées sur des régressions pénalisées, qui sont les plus performantes. En termes de classement des signaux générés, elle a eu des performances proches de celles de lasso-bic, qui a fourni un classement pertinent. Les approches mwPS ont un comportement intermédiaire entre les approches basées sur des régressions multiples pénalisées et l'approche univariée dans le type de signaux générés.

L'ajustement sur le PS n'est pas la méthodologie qui a obtenu les meilleures performances dans notre étude. Les approches adjustPS ont un comportement assez similaire à celui de l'approche univariée. En particulier, elles génèrent un très grand nombre de signaux par rapport aux autres approches basées sur le PS et aux approches basées sur des régressions pénalisées. Elles montrent également une faible spécificité et une bonne sensibilité. Néanmoins, ces méthodes fournissent un classement des signaux générés pertinent.

Les approches basées sur la pondération IPTW ont donné des résultats extrêmement médiocres. Contrairement aux poids MW, les poids dans IPTW ne sont pas normalisés et peuvent potentiellement être très grands pour les individus non exposés ayant une valeur de PS proche de zéro. Cette instabilité numérique due à des poids élevés avec IPTW a déjà été signalée dans la littérature (Yoshida *et al.*, 2017). Pour éviter ce problème, une solution consiste à procéder à une troncature des poids : les poids supérieurs à une valeur

donnée se voient attribuer cette valeur. On procède de la même manière pour les valeurs de poids faibles (Seeger *et al.*, 2017). Plus récemment, d'autres poids calculés à partir du PS ont été proposés : les *overlap weights* (Li *et al.*, 2019). Tout comme les poids MW, ils sont construits de manière à être compris entre zero et un. Dans leur travail Li *et al.* (2019) montrent que cette pondération produit une estimation des effets du traitement sur la réponse d'intérêt moins biaisée et de variance moindre par rapport aux estimations obtenues avec IPTW (avec et sans troncature). Il pourrait donc être intéressant d'évaluer les performances d'une telle pondération dans le cadre de la pharmacovigilance.

Outre l'algorithme de sélection de variables bien connu hdPS, nous avons mis en œuvre trois autres méthodes d'estimation du PS dans ce cadre : deux basées sur des régressions lasso et un algorithme d'apprentissage automatique basé sur des arbres de régression. Un inconvénient de l'algorithme hdPS dans le contexte de la pharmacovigilance est que le PS obtenu avec cette méthode est construit à partir de l'EI considéré. Lorsque l'on examine plusieurs milliers d'EI, cette tâche peut s'avérer très chronophage. Au contraire, les trois autres méthodes d'estimation présentent un avantage en termes de calcul : une fois le PS estimé pour un médicament donné, il peut être utilisé pour tester son association avec n'importe quel autre EI. Néanmoins, ne pas impliquer la réponse d'intérêt dans le processus de sélection des variables à inclure dans le modèle d'estimation du PS peut amener à choisir des variables instrumentales. Comme signalé par Brookhart *et al.* (2006), ajouter de telles variables dans la construction du PS augmente la variance de l'estimation de l'effet de l'exposition médicamenteuse sur la réponse. Au vu des résultats de notre comparaison empirique, nous sommes cependant assez confiant dans l'utilisation de ces algorithmes pour la construction des scores. En comparant les résultats des approches issues des différents scores obtenus, nous avons constaté que la variabilité des performances s'explique majoritairement par la manière de prendre en compte le score dans l'analyse et non la méthode d'estimation de celui-ci.

Une autre source de variabilité dans l'estimation de l'effet de l'exposition est liée à l'étape d'estimation du PS. La majorité des études conduites à l'aide du PS considèrent le PS comme une valeur théorique et non une valeur estimée. La variabilité issue de l'étape d'estimation du score n'est alors pas prise en compte dans l'estimation de la

variance de l'effet de l'exposition sur la réponse d'intérêt. Pour remédier à cela, plusieurs estimateurs de la variance de l'effet de l'exposition ont été définis pour les différentes stratégies de prise en compte du PS (Williamson *et al.*, 2012; Li et Greene, 2013; Zou *et al.*, 2016; Abadie et Imbens, 2016). Le contexte de la pharmacovigilance est exploratoire. Les méthodes développées ici ont surtout pour but de détecter un effet délétère d'une exposition médicamenteuse. Elles n'ont pas vocation à estimer la magnitude de cet effet. Néanmoins, mettre en œuvre ces estimateurs proposés et ainsi prendre en compte l'étape d'estimation du PS serait une perspective prometteuse de ce travail : une meilleure estimation de la variance amènerait à une détection plus fiable avec nos approches.

Pour attester de la qualité d'un score de propension, on mesure sa capacité à induire l'équilibre des distribution des variables observées entre les individus traités et non traités qui ont des valeurs de PS proches. La mesure d'équilibre la plus utilisée est celle de la différence de moyennes standardisées (*Standardized Mean Difference*, SMD) (Austin, 2011). Généralement, la SMD est définie pour chaque variable observée comme la différence standardisée des moyennes (pour une variable continue) ou des proportions (pour une variable binaire) entre des populations d'individus traités et non traités appariés sur le PS (Austin, 2011). Ainsi, une valeur est calculée pour chacune des variables dont on souhaite tester l'équilibre induit par le PS. La moyenne de toutes ces valeurs (l'*Average Standard Mean Difference*) est utilisée comme résumé de l'information portée par les différentes valeurs des SMD. D'autres métriques comme la distance de Mahalanobis (Caruana *et al.*, 2015), la distance de Kolmogorov–Smirnov, la distance de Lévy ou encore le coefficient de superposition (*overlapping coefficient*) ont été proposées (Ali *et al.*, 2014). Certains de ces critères ont été utilisés dans des études pharmacoépidémiologiques basées sur le PS en grande dimension et menées sur des bases médico-administratives (Groenwold *et al.*, 2011). Un développement pertinent de ce travail serait de mettre en œuvre ces mesures d'équilibre pour comparer plus finement les scores obtenus par nos différentes méthodes d'estimation des scores. Se poserait alors la question, dans le cadre de la pharmacovigilance, de trouver une manière de résumer toutes ces mesures sur chacun des scores estimés associés aux différents médicaments.

L'utilisation du BIC comme critère pour sélectionner le paramètre de pénalisation a été largement étudié. Zou *et al.* (2007) ont recommandé d'avoir recours à ce critère quand l'objectif visé est un objectif de sélection de variables. Dans leur travail, le BIC est calculé à partir de la vraisemblance obtenue par maximisation de la vraisemblance pénalisée et le nombre de degrés de liberté du modèle est approché par le nombre de variables ayant des coefficients de régression estimés non nuls. Des variations du BIC ont été développées spécifiquement dans le cadre des régressions pénalisées comme l'*extended BIC* (Chen et Chen, 2008), le *modified BIC* (Wang *et al.*, 2009) ou encore le *high dimensional BIC* (Wang et Zhu, 2011). Dans ce travail, nous avons utilisé le BIC dans un cadre de sélection de modèles où la liste des modèles candidats était déterminée à l'aide de la régression lasso. Une extension intéressante de ce travail serait de considérer d'autres critères basés sur le BIC pour choisir la pénalité à appliquer dans la régression lasso.

L'une des principales difficultés dans le développement de méthodes de détection de signaux est de disposer d'un ensemble fiable et suffisamment large de signaux de référence permettant d'évaluer les performances des méthodes envisagées. Nous avons utilisé ici l'ensemble DILIrank relatif à un événement indésirable commun, ce qui n'est pas le cas pour tous les EI de la base de données.

Les critères qui permettent d'évaluer les performances des méthodes testées sont obtenus à partir de l'ensemble de signaux de référence. Bien que cet ensemble soit de taille importante, les performances des approches peuvent être surestimées. En effet, il est raisonnable de penser que parmi tous les médicaments dont le statut d'association avec l'évènement DILI n'est pas connu, seul un petit nombre est associé de façon délétère à une DILI. Or, plus de la moitié des signaux générés par chaque approche ont un statut inconnu.

Un dernier aspect à prendre en considération dans le développement de méthodes de détection de signaux concerne le coût en termes de temps de calcul. Les méthodes basées sur les régressions pénalisées sont nettement plus lourdes de ce point de vue que les méthodes traditionnelles de disproportionnalité, car il faut les appliquer aux milliers d'EI présents dans les bases de données. Les approches proposées basées sur le PS sont égale-

ment très coûteuses en temps de calcul dans l'étape d'estimation des scores pour chaque exposition. Par exemple, lorsque les scores sont estimés à l'aide de la méthode lasso-bic, cela revient à mettre en œuvre autant de régressions pénalisées qu'il y a de médicaments renseignés. Dans ce travail, c'est le fait de pouvoir paralléliser les tâches d'estimation qui a rendu cette étape moins chronophage. En revanche, une fois les scores estimés, les approches basées sur l'ajustement ou la pondération sur le PS sont compétitives avec les approches basées sur des régressions pénalisées en termes de temps de calcul.

2.8 Conclusion

Les résultats suggèrent que la méthodologie proposée basée sur le PS en grande dimension est un complément intéressant aux autres méthodes de détection existantes sur les bases des notifications spontanées. Dans un sens, elle combine les principaux atouts des approches de régression univariée et multiple : elle permet à la fois de prendre en compte la co-notification en résumant l'information dans un score, et d'avoir recours à la théorie des tests d'hypothèses multiples en ce qui concerne le seuil de détection. Néanmoins, toutes les méthodes de détection développées n'aboutissent pas aux mêmes performances. L'approche basée sur le PS ayant les meilleurs résultats en termes de fausses découvertes est celle qui repose sur la pondération sur les poids MW. Ces performances restent néanmoins sujettes à caution étant donné qu'elles dépendent grandement de l'ensemble de signaux de référence choisi.

Bien que l'on ne puisse pas se reposer sur un critère d'erreur statistique pour la sélection de variables avec les approches basées sur les régressions pénalisées, ces dernières ont montré les meilleures performances dans cette comparaison empirique. En outre ces approches sont beaucoup moins coûteuses en temps de calcul par rapport aux approches qui reposent sur le PS en grande dimension.

Dans la suite, nous avons ainsi décidé de poursuivre le développement méthodologique pour la détection de signaux sur les bases de notifications spontanées autour des régressions pénalisées.

2.9 Valorisation de l'étude

Ce travail a fait l'objet d'une publication dans *Frontiers in Pharmacology* (cf. annexe D.1). Il a été également présenté lors d'un congrès national, la journée des Jeunes Chercheurs de la Société Française de Biométrie (mai 2018), et lors d'un congrès international, l'*International Biometric Conference* (juillet 2018).

Chapitre 3

Le lasso adaptatif pour la détection de signaux à partir des notifications spontanées

Nous avons montré lors de notre première étude que l'utilisation du score de propension en grande dimension pour la détection de signaux en pharmacovigilance présente un certain intérêt. Cette étude a également mis en avant les très bonnes performances des méthodes de détection basées sur la régression pénalisée de type lasso. Comme cela a été dit dans l'introduction (voir section 1.5), l'objectif de la pharmacovigilance peut être abordé sous l'angle de la sélection de variables : parmi la multitude de variables médicaments on vise à choisir les quelques uns potentiellement associés à un évènement indésirable d'intérêt. Dans ce chapitre, nous décidons de poursuivre le développement méthodologique autour des régressions pénalisées avec l'utilisation d'une extension du lasso qui possède de bonnes propriétés en termes de sélection de variables : le lasso adaptatif. Ces développements méthodologiques seront appliqués dans le cadre des données de notifications spontanées.

3.1 Introduction

Les propriétés asymptotiques du lasso pour la sélection de variables ont été largement étudiées (Knight et Fu, 2000; Meinshausen et Bühlmann, 2006). En particulier, Fan

et Li (2001) ont défini les propriétés que doit remplir une procédure statistique optimale. Entre autres, une telle procédure doit (i) identifier l'ensemble des vrais prédicteurs, (ii) être non biaisée. Fan et Li (2001) ont montré dans leur travail que le lasso ne satisfait pas ces propriétés oracles. Ainsi, la sélection de variables opérée par le lasso peut ne pas être consistante.

Proposé par Zou (2006), le lasso adaptatif (*adaptive lasso*) est une extension du lasso qui vise à améliorer les propriétés de sélection de variables de ce dernier. Elle consiste à introduire des valeurs de pénalité propres à chaque variable explicative dans la fonction de pénalisation de norme L_1 du lasso (voir équation (2.3)) afin de pénaliser différemment les variables du modèle. Ces différentes valeurs de pénalité sont déterminées à partir des données, d'où le fait que l'on parle aussi de pénalité adaptative ou de poids adaptatifs pour les désigner. Sous certaines conditions sur les poids adaptatifs, le lasso adaptatif satisfait les propriétés oracles. Dans le travail de Zou (2006), ces poids étaient construits dans le cadre classique en recourant au maximum de vraisemblance. Des propositions ont été faites par la suite pour construire ces poids dans le contexte de la grande dimension. En particulier, Bühlmann et van de Geer (2011) ont utilisé l'estimation obtenue avec une première régression lasso pour en déduire un vecteur de poids adaptatifs à intégrer dans le lasso adaptatif.

Dans ce travail, nous présentons une nouvelle stratégie de détection automatisée de signaux basée sur le lasso adaptatif, ce qui n'a jamais encore été exploré. Nous proposons deux nouveaux poids adaptatifs dérivés de (i) une régression logistique lasso pour laquelle le paramètre de régularisation est choisi à l'aide du BIC, et (ii) la méthode CISL, toutes deux présentées au chapitre précédent (voir sections 2.3.4 et 2.3.3). Nous comparons nos deux propositions à d'autres stratégies de détection de signaux qui reposent sur

- des mises en œuvre plus classiques du lasso adaptatif en grande dimension, (Bühlmann et van de Geer, 2011; Huang *et al.*, 2008a,b) ;
- des régressions lasso en considérant plusieurs critères pour le choix de la pénalisation (validation croisée, BIC, ou permutations (Ayers et Cordell, 2010; Sabourin *et al.*, 2015)) ;
- CISL
- le score de propension en grande dimension, présenté au chapitre précédent.

Cette comparaison est réalisée via des simulations qui exploitent la matrice d'exposition de la BNPV afin de préserver la structure des données de pharmacovigilance, en particulier leur sparcité. Ce travail de simulations est complété par une application aux données réelles, où l'on utilise le même ensemble de signaux référence que celui du chapitre précédent : DILIrank.

3.2 Lasso adaptatif en grande dimension

Afin de pallier le fait que la sélection de variables opérée par le lasso peut ne pas être consistante, Zou (2006) a proposé une nouvelle version du lasso : le lasso adaptatif qui consiste à introduire dans la fonction de pénalisation des valeurs de pénalité propres à chaque variable. Ainsi, dans le lasso adaptatif, la fonction de pénalisation dans (2.2) est définie par

$$\text{pen}(\lambda) = \lambda \sum_{j=1}^{P} w_p |\beta_j|.$$

La pénalité appliquée à la variable X_j est donc égale à $\lambda_j = \lambda \times w_j$. Une augmentation de la valeur du poids w_j implique alors une augmentation de la pénalisation appliquée à la variable X_j. Ainsi, elle a moins de chance d'avoir un coefficient estimé non nul et d'être sélectionnée par cette procédure.

Dans son travail, Zou construit les poids adaptatifs à partir d'un estimateur consistant de β^*, le vecteur des vrais coefficients de régression (inconnus). Dans le cas logistique, il s'agit de considérer les estimations produites par maximum de vraisemblance $\widehat{\beta}_{mle}$. Ainsi, pour une variable X_j, son poids adaptatif associé est défini par

$$w_j = \frac{1}{|\widehat{\beta}_j^{mle}|^\gamma},$$

avec $\gamma > 0$. En attribuant une pénalité plus forte aux variables qui ont des coefficients de régression faibles, et une pénalité plus faible à celles qui ont des coefficients élevés, le lasso adaptatif satisfait les conditions oracles.

Néanmoins, dans le contexte de la grande dimension, il n'est pas trivial de trouver des estimateurs consistants pour construire les poids adaptatifs, puisque l'estimation classique

par maximum de vraisemblance n'est pas envisageable. Zou propose d'utiliser les estimations issues d'une régression *ridge* (Hoerl et Kennard, 2000), tout en précisant que les propriétés des poids adaptatifs dérivés de ces coefficients restent à démontrer.

Une autre version du lasso adaptatif appropriée au contexte de la grande dimension a été proposée par Bühlmann et van de Geer (2011) dans le cas linéaire. Cette approche repose sur une stratégie en deux étapes : les coefficients estimés par une première régression lasso sont utilisés pour définir les poids dans le lasso adaptatif, en fixant $\gamma = 1$. Huang *et al.* (2008a) ont proposé la même stratégie dans le cas logistique. Ils ont également montré dans un autre travail que dans le cas linéaire, les poids adaptatifs dérivés des coefficients de régression univariés possèdent de bonnes propriétés sous certaines conditions (Huang *et al.*, 2008b).

Dans la pratique, la question de la sélection du paramètre de pénalité pour le lasso subsiste avec le lasso adaptatif. Dans les propositions faites dans le cas de la grande dimension évoquées ci-dessus, la validation croisée est utilisée.

3.3 Nouveaux poids adaptatifs dans le cadre de la grande dimension

Nous proposons deux nouveaux poids dans l'utilisation du lasso adaptatif dans le but de définir des méthodes de détection de signaux en pharmacovigilance. L'objectif de ces nouveaux poids est d'ajuster la pénalité appliquée aux variables médicaments afin d'améliorer la sélection du bon sous-ensemble de médicaments associés à un EI donné. Nous proposons également d'avoir recours au critère BIC et non à la validation croisée, comme cela a été proposé dans les mises en œuvre du lasso adaptatif en grande dimension présentées en section 3.2. Avoir recours au critère BIC nous semble plus pertinent, la validation croisée étant peu appropriée au contexte de la sélection de variables.

Le premier poids adaptatif proposé est défini à partir des estimations obtenues avec une régression lasso où le BIC est utilisé comme critère de décision pour choisir la valeur de la pénalité. Il s'agit de l'approche présentée en section 2.3.4, appelée lasso-bic. En désignant par $\widehat{\beta}^{lbic}$ le vecteur de taille P des coefficients de régression non pénalisés estimés avec lasso-bic (les composantes de ce vecteur correspondant aux variables non retenues dans

le modèle minimisant le BIC étant nulles), nous définissons le poids adaptatif associé à la variable X_j dans l'étape du lasso adaptatif par :

$$w_j^{lb} = \begin{cases} \dfrac{1}{|\widehat{\beta}_j^{lbic}|} & \text{si } \widehat{\beta}_j^{lbic} \neq 0 \\ \infty & \text{sinon.} \end{cases} \tag{3.1}$$

Ainsi, une variable qui n'a pas été sélectionnée par le lasso-bic dans la première étape est automatiquement exclue dans l'étape du lasso adaptatif. Par construction, les variables sélectionnées avec le lasso adaptatif sont un sous-ensemble de celles sélectionnées à l'aide du BIC dans la première régression lasso.

Le deuxième poids adaptatif proposé est dérivé de la méthode CISL présentée en section 2.3.3. Dans un premier temps, CISL est implémenté en considérant une condition non nulle sur les coefficients de régression dans le calcul de la quantité (2.4), au lieu de la condition de positivité initiale :

$$\widehat{\tau}_j^b = \frac{1}{E} \sum_{\eta=1}^{E} \mathbb{1}[\widehat{\beta}_{\eta,j}^b \neq 0], \tag{3.2}$$

où les notations sont les mêmes que celles utilisées en section 2.3.3. Ainsi, la quantité (3.2) mesure dans quelle proportion une variable X_j a été sélectionnée parmi les modèles candidats de complexité 1 à E issus de la régression lasso, sur le sous-échantillon b. On note $\widehat{\tau}_j$ le vecteur de taille B obtenu à l'issue de la procédure CISL. On définit le poids adaptatif associé à la variable X_j dans l'étape du lasso adaptatif par :

$$w_j^{cisl} = \begin{cases} \dfrac{1}{B} & \text{si } \forall b \in \{1, ..., B\}\ \widehat{\tau}_j^b > 0, \\ \infty & \text{si } \forall b \in \{1, ..., B\}\ \widehat{\tau}_j^b = 0, \\ 1 - \dfrac{1}{B} \sum_{b=1}^{B} \mathbb{1}[\widehat{\tau}_j^b > 0] & \text{sinon.} \end{cases} \tag{3.3}$$

Ainsi, plus $\widehat{\tau}_j$ a de composantes non nulles, plus la variable X_j est apparue dans la liste de modèles candidats fournie par le lasso sur les B sous-échantillons, et plus le poids associé à cette variable est faible. Les variables qui ne sont sélectionnées par aucune des B régressions lasso sont exclues à l'étape du lasso adaptatif.

Toujours dans le but de favoriser les performances du lasso adaptatif en termes de sélection de variables, nous avons choisi d'utiliser le BIC tel que défini dans la section 2.3.4 pour identifier le sous-ensemble de variables final à sélectionner dans l'étape du lasso adaptatif. Tout comme pour la méthode de détection appelée lasso-bic au chapitre précédent, nous déclarons comme signaux les variables associées positivement à la réponse dans le modèle logistique non pénalisé qui minimise le BIC. Dans ce qui suit, nous appellerons les méthodes de détection ainsi définies adapt-bic et adapt-cisl.

3.4 Méthodes de détection compétitrices

3.4.1 Méthodes de détection basées sur le lasso adaptatif en grande dimension

On se propose de considérer des approches de détection basées sur les mises en œuvre du lasso adaptatif pour la grande dimension évoquées en section 3.2.

La première approche repose sur la proposition de Bühlmann et van de Geer (2011). Elle consiste à utiliser les coefficients de régression obtenus avec régression lasso afin de déterminer les poids adaptatifs. A la fois pour cette étape et pour l'étape du lasso adaptatif, le paramètre de pénalisation λ est sélectionné par validation croisée. En notant $\widehat{\beta}^{lcv}$ le vecteur de taille P des coefficients de régression du lasso logistique déterminé par validation croisée, le poids adaptatif associé à la X_j dans le lasso adaptatif est défini par :

$$w_p^{lcv} = \begin{cases} \dfrac{1}{|\widehat{\beta}_j^{lcv}|} & \text{si } \widehat{\beta}_j^{lcv} \neq 0 \\ \infty & \text{sinon.} \end{cases}$$

Ainsi, une variable qui n'a pas été sélectionnée par validation croisée dans la première régression lasso est automatiquement exclue par le lasso adaptatif. Par construction, les variables sélectionnées avec le lasso adaptatif sont un sous-ensemble de celles sélectionnées à l'aide de la validation croisée dans la première régression lasso.

La deuxième approche, basée sur le travail de Huang *et al.* (2008b), consiste à déterminer les poids adaptatifs à partir des coefficients obtenus avec des régressions univariées sur la réponse d'intérêt. En notant $\widehat{\beta}_j^{univ}$ le coefficient univarié associé à la variable X_j,

on définit son poids associé dans le lasso adaptatif par :

$$w_j^{univ} = \frac{1}{|\widehat{\beta}_j^{univ}|}.$$

Comme dans le travail de Huang *et al.* (2008b), on détermine la valeur du paramètre de pénalité dans le lasso adaptatif par validation croisée.

Dans la suite, les approches de détection qui résultent de ces deux implémentations du lasso adaptatif en grande dimension seront désignées sous les appellations adapt-cv et adapt-univ. Sont considérés comme signaux avec ces approches toutes les variables qui ont un coefficient de régression associé strictement positif dans le lasso adaptatif.

3.4.2 Méthodes de détection basées sur la régression lasso

Afin de quantifier l'intérêt de l'utilisation du lasso adaptatif par rapport au lasso classique en pharmacovigilance, nous avons implémenté des approches de détection de signaux basées sur la régression lasso logistique (présentée en section 2.3.1). En plus du lasso-bic défini en section 2.3.4 et de CISL présenté en section 2.3.3, nous avons choisi de mettre en œuvre deux autres stratégies pour sélectionner le paramètre de pénalisation dans la régression lasso : la validation croisée et une méthode basée sur des permutations proposée par Sabourin *et al.* (2015) à partir de l'idée formulée par Ayers et Cordell (2010). Pour de multiples permutations aléatoires de la réponse, cette méthode consiste à identifier la plus petite pénalité associée à un ensemble de variables actives vide dans la régression lasso, puis à considérer la médiane de ces valeurs de pénalité. En notant π une permutation de $\{1, ..., N\}$, on définit $\mathbf{y}_{\pi_l} = (y_{\pi(1)}, ..., y_{\pi(N)})$ une version permutée de la réponse d'intérêt $\mathbf{y}$ avec $1 \leq l \leq L$. Une régression lasso est implémentée pour chaque permutation, où $\mathbf{y}_{\pi_l}$ est régressé sur le jeu de données original $\mathbf{X}$. Pour chacune de ces L régressions lasso, la plus petite valeur du paramètre de pénalité aboutissant à un ensemble de variables actives vide est sélectionnée. On note cette valeur $\lambda_{max}(\mathbf{y}_{\pi_l})$ pour la permutation l. La valeur médiane de $(\lambda_{max}(\mathbf{y}_{\pi_1}), ..., \lambda_{max}(\mathbf{y}_{\pi_L}))$ est finalement utilisée comme paramètre de pénalité dans une régression lasso effectuée avec la réponse d'intérêt non permutée. Comme suggéré dans leur travail, nous avons fixé $L = 20$.

Dans ce qui suit, lasso-cv et lasso-perm feront respectivement référence à l'approche impliquant la validation croisée ou des permutations. Pour la validation croisée, nous avons fixé le nombre de *folds* à cinq et nous avons utilisé comme mesure de performance la déviance, définie comme l'écart entre la log-vraisemblance du modèle évalué par rapport à la log-vraisemblance du modèle vide. Pour ces deux approches, les signaux sont les variables qui ont un coefficient de régression associé strictement positif.

3.4.3 Méthodes de détection basées sur le score de propension en grande dimension

Nous avons également voulu comparer toutes ces approches de détection basées sur les régressions pénalisées avec les approches basées sur le score de propension en grande dimension définies au chapitre précédent.

Pour chaque variable médicament, un score de propension a été construit avec la méthode du lasso-bic en sélectionnant les variables à inclure dans les modèles d'estimation parmi tous les autres médicaments renseignés. Nous avons pris en compte ces scores dans le modèle de régression final par ajustement et pondération IPTW (Austin, 2011) et MW (Li et Greene, 2013). Pour tenter d'améliorer les performances médiocres de la méthode par pondération IPTW observées dans le chapitre précédent, nous avons également considéré une approche avec troncature de ces poids afin de limiter l'impact des poids extrêmes. Cela consiste à attribuer aux individus dont le poids correspondant est inférieur au $r^{\text{ème}}$ percentile ou supérieur au $(1-r)^{\text{ème}}$ percentile de la distribution des poids, la valeur du $r^{\text{ème}}$ ou du $(1-r)^{\text{ème}}$ percentile respectivement (Franklin *et al.*, 2017). Ici, nous avons choisi de fixer $r = 2{,}5\%$.

Pour tenir compte des tests multiples, nous avons utilisé la procédure proposée par Benjamini et Yekuteli (2001) pour contrôler le FDR sous des hypothèses de dépendance entre les tests réalisés. Cette procédure consiste dans un premier temps à ordonner les p-valeurs des P tests réalisés dans l'ordre croissant : $p_{(1)} \leq ... \leq p_{(P)}$, avec $p_{(1)}$ la plus petite des P p-valeurs obtenues et $p_{(P)}$ la plus grande. Pour un seuil α donné, on cherche

la plus grande valeur de h telle que

$$p_{(h)} \leq \frac{h}{P \times c(P)}\alpha$$

avec $c(P) = \sum_{j=1}^{P} \frac{1}{j}$. Les hypothèses nulles des tests associés aux p-valeurs $p_{(1)}, ..., p_{(h)}$ sont alors rejetées. Les variables médicaments sélectionnées par nos approches basées sur le PS en grande dimension sont les variables $X_{(1)}, ..., X_{(h)}$. Nous avons fixé $\alpha = 5\%$. Dans ce qui suit, ps-adjust, ps-iptw, ps-iptwT et ps-mw feront respectivement référence à l'ajustement sur le PS, la pondération sur le PS avec les poids IPTW, IPTW avec troncature, et MW.

3.5 Simulations

Nous avons réalisé une étude de simulations pour évaluer les performances des méthodes de détection basées sur le lasso adaptatif que nous proposons, présentées en section 3.3, et toutes les autres approches présentées en section 3.4. Les performances des approches sont évaluées en termes de FDR, qui est notre critère d'intérêt principal, et de sensibilité pour un large éventail de scénarios de simulations.

3.5.1 Modèles et plans de simulations

Nous avons simulé l'occurrence d'un EI donné selon un modèle de régression logistique $y_i \sim \text{Bernoulli}(\alpha_i)$ avec

$$\alpha_i = \frac{1}{1 + \exp(-\beta_0 - \sum_{p=1}^{P} \beta_p x_{ip})}.$$

En ce qui concerne la matrice d'exposition aux médicaments, nous avons utilisé la BNPV extraite sur la période allant du 1er janvier 2000 au 29 décembre 2017 selon les mêmes filtres que ceux détaillés en section 2.5.3. Pour chaque réplication de chaque scénario, nous avons sélectionné aléatoirement 100 000 notifications. Pour chacun de ces jeux de données, nous avons sélectionné au hasard un sous-ensemble de 500 médicaments parmi ceux qui ont été notifiés plus de 10 fois parmi les 100 000 notifications considérées.

Nous avons étudié 27 scénarios de simulations qui différaient selon :

- la valeur de l'intercept β_0, afin de simuler des réponses d'intérêt plus ou moins rares, β_0 = -2, -4, -6 ;
- le nombre de variables médicament associées à la réponse parmi les 500 variables considérées. Ces variables sont appelées dans la suite vrais prédicteurs : $n_{TP} = 0$, 5, 20 ;
- la valeur des coefficients de régression pour les n_{TP} vrais prédicteurs, identique pour tous les vrais prédicteurs : $\beta_{TP} = 1$, 2 ;
- la fréquence de notification des vrais prédicteurs (s'il y en a) : fréquente (les vrais prédicteurs sont notifiés au moins 100 fois) ou rare (les vrais prédicteurs sont notifiés entre 20 et 100 fois).

Nous avons effectué 500 réplications de chacun de ces scénarios, et pour chacune de ces réplications, les n_{TP} vrais prédicteurs étaient choisis au hasard.

3.5.2 Méthodes mises en œuvre

Nous considérons quatre méthodes de détection basées sur le lasso adaptatif avec des poids définis à partir : d'un lasso avec validation croisée, de coefficients univariés, d'un lasso avec le BIC et de CISL. Afin de sélectionner la valeur de λ dans le lasso adaptatif, la validation croisée est utilisée pour les deux premiers poids. Pour nos deux propositions de poids, le BIC est utilisé. Afin de mesurer la pertinence de l'utilisation du BIC avec le lasso adaptatif, nous avons inclus dans la comparaison une méthode basée sur les mêmes poids adaptatifs que ceux définis pour adapt-univ, à savoir des poids basés sur des coefficients univariés, mais au lieu d'une validation croisée, nous avons utilisé le BIC pour effectuer la sélection de variables avec le lasso adaptatif. Dans ce qui suit, nous appelons cette approche adapt-univ-bic. Le tableau 3.1 résume toutes les approches de détection de signaux mises en œuvre basées sur le lasso adaptatif. Pour chaque approche, il est précisé comment les poids adaptatifs sont obtenus et quelle méthode de sélection de variables est utilisée.

Pour toutes les méthodes reposant sur la validation croisée, nous avons utilisé les mêmes *folds*. Ces *folds* étaient différents pour chaque réplication. Pour CISL, nous avons

Méthode	Construction des poids adaptatifs	Critère pour la sélection de variables
adapt-cv	lasso-cv	validation croisée
adapt-univ	coefficients univariés	validation croisée
adapt-univ-bic	coefficients univariés	BIC
adapt-bic	lasso-bic	BIC
adapt-cisl	CISL	BIC

Tableau 3.1 – Caractéristiques des méthodes de détection de signaux basées sur le lasso adaptatif.

utilisé le quantile à 10% de la distribution de la quantité (2.4) comme seuil de détection. Pour les approches basées sur le PS, nous avons appliqué un filtre supplémentaire en ne considérant que les médicaments qui avaient plus de trois occurrences communes avec la réponse d'intérêt, comme c'était le cas dans notre étude précédente (voir 2.5.4). Ici, nous avons fixé à un les p-valeurs des variables exclues par ce filtre, les méthodes étaient implémentées sur les variables restantes.

Dans un souci d'exhaustivité, nous avons également inclus une méthode de disproportionnalité dans la comparaison : le *Reporting Fisher Exact Test* (RFET) (Ahmed *et al.*, 2010). Contrairement au ROR et au PRR (voir section 1.2), l'approche RFET ne repose pas sur des hypothèses asymptotiques qui sont rarement vérifiées étant donné les faibles occurrences de bon nombre de médicaments et EI. Ainsi cette approche de disproportionnalité repose sur le test exact de Fisher. Comme pour les approches basées sur le PS, le RFET a été implémenté uniquement pour les médicaments qui ont plus de trois notifications en commun avec la réponse. Des tests unilatéraux ont été réalisés avec cette approche, pour ne considérer que des associations délétères. Nous avons appliqué la même procédure de correction des tests multiples à RFET que celle présentée en section 3.4.3, avec la même stratégie détaillée ci-dessus pour les p-valeurs des variables filtrées.

Au total, nous avons comparé 14 approches de détection de signaux : une méthode de disproportionnalité, quatre approches basées sur le lasso, cinq approches basées sur le lasso adaptatif et quatre approches basées sur le PS.

3.5.3 Résultats

Le tableau 3.2 montre le nombre moyen de variables et le nombre moyen de vrais prédicteurs conservés après avoir écarté les variables ayant moins de trois notifications en

commun avec la réponse simulée. Ce filtre concerne les méthodes basées sur le score de propension en grande dimension et RFET. Le tableau 3.2 indique également le nombre moyen de cas par scénario. Les effectifs moyens sont donnés avec leurs écarts interquartiles. À mesure que le nombre de cas diminue, le nombre de variables retenues après filtrage diminue également, ce qui inclut les vrais prédicteurs. Lorsque les vrais prédicteurs sont rarement notifiés et que la réponse est particulièrement rare, aucun prédicteur n'est retenu dans un grand nombre de réplications (scénarios 24, 25, 26 où les effectifs sont arrondis à zéro).

Scénario	β_0	n_{TP}	β_{TP}	Nombre moyen et IQR de covariables retenues après filtrage	Nombre moyen et IQR de vrais prédicteurs retenus après filtrage	Nombre moyen et IQR de cas
Aucun vrais prédicteurs						
1	-2	0	0	389 (12)	NA	11923 (122)
2	-4	0	0	165 (13)	NA	1798 (63)
3	-6	0	0	30 (8)	NA	247 (21)
Les vrais prédicteurs sont notifiés plus de 100 fois						
4	-2	5	1	394 (12)	5 (0)	12327 (291)
5	-2	5	2	400 (12)	5 (0)	12937 (606)
6	-2	20	1	407 (13)	20 (0)	13589 (723)
7	-2	20	2	424 (14)	20 (0)	15956 (1664)
8	-4	5	1	172 (14)	5 (0)	1879 (90)
9	-4	5	2	184 (17)	5 (0)	2083 (180)
10	-4	20	1	193 (16)	20 (0)	2155 (160)
11	-4	20	2	237 (28)	20 (0)	3121 (618)
12	-6	5	1	34 (9)	2 (1)	259 (24)
13	-6	5	2	44 (13)	4 (1)	294 (36)
14	-6	20	1	49 (12)	11 (3)	300 (31)
15	-6	20	2	95 (26)	18 (1)	507 (134)
Les vrais prédicteurs sont notifiés entre 20 et 100 fois						
16	-2	5	1	391 (11)	5 (0)	11958 (121)
17	-2	5	2	392 (12)	5 (0)	12013 (122)
18	-2	20	1	396 (11)	20 (0)	12064 (130)
19	-2	20	2	400 (12)	20 (0)	12284 (136)
20	-4	5	1	167 (12)	2 (2)	1805 (62)
21	-4	5	2	171 (12)	4 (1)	1822 (62)
22	-4	20	1	175 (13)	8 (4)	1827 (63)
23	-4	20	2	191 (13)	17 (2)	1897 (65)
24	-6	5	1	30 (8)	0 (0)	248 (21)
25	-6	5	2	31 (8)	0 (1)	251 (22)
26	-6	20	1	31 (9)	0 (0)	251 (21)
27	-6	20	2	35 (9)	2 (2)	263 (23)

TABLEAU 3.2 – Nombre moyen de variables et de vrais prédicteurs retenus après filtrage des variables ayant moins de trois occurrences communes avec la réponse. Nombre moyen de cas selon les scénarios de simulations.
IQR : écart interquartile (*interquartile range*)

Dans un premier temps, nous avons comparé nos propositions, adapt-bic et adapt-cisl, aux autres approches reposant sur le lasso adaptatif. La figure 3.1 représente le FDR et

la sensibilité moyenne des approches présentées dans le tableau 3.1 pour les scénarios 1 à 15, c'est-à-dire les scénarios dans lesquels il n'y a pas de vrais prédicteurs (scénario 1-3) et les scénarios avec des vrais prédicteurs fréquemment notifiés (scénario 4-15).

Adapt-cv et adapt-univ ont montré une grande sensibilité au prix d'un FDR très élevé, en particulier adapt-univ. Hormis pour le scénario 5, où adapt-cv a montré un FDR inférieur à 0,10, cette approche a toujours montré un FDR supérieur à 0,30 allant même jusqu'à afficher un FDR à 0,70 pour le scénario 12. Adapt-univ a majoritairement affiché un FDR aux alentours de 0,60, allant même jusqu'à des valeurs proches de 0,80 pour les scénarios 4, 8 et 13. Pour les scénarios 1 à 3 (sans vrais prédicteurs), on constate que le FDR d'adapt-univ diminue fortement avec la valeur de l'intercept, avec un FDR proche de 0,10 quand $\beta_0 = -6$.

En comparant adapt-univ et adapt-univ-bic, on constate que l'utilisation du BIC pour effectuer la sélection de variables dans le lasso adaptatif a permis de nettement réduire la valeur du FDR dans tous les scénarios. Adapt-univ-bic a montré dans la majorité des scénarios un FDR proche de 0,10.

En comparant les approches qui s'appuient sur le BIC dans la phase du lasso adaptatif (à savoir adapt-univ-bic, adapt-bic et adapt-cisl) nous constatons que les approches basées sur nos propositions de poids adaptatifs conduisent à de meilleurs résultats dans l'ensemble. Elles ont montré à la fois un FDR plus faible et une sensibilité plus élevée qu'adapt-univ-bic. Ces différences de performance étaient particulièrement remarquables dans les scénarios 4, 8, 10, et 13 à 15. Néanmoins, adapt-univ-bic avait un FDR légèrement inférieur à celui d'adapt-bic et d'adapt-cisl lorsqu'il n'y avait pas de vrais prédicteurs (scénarios 1 à 3). Adapt-bic et adapt-cisl ont montré dans ces scénarios un FDR d'environ 0,10, tandis qu'adapt-univ-bic affichait un FDR d'environ 0,05.

Parmi nos propositions, adapt-cisl a généralement donné de meilleurs résultats qu'adapt-bic avec un FDR plus faible et une sensibilité légèrement plus élevée, en particulier dans les scénarios 12 à 15, dans lesquels la réponse d'intérêt était particulièrement déséquilibrée. Dans ces scénarios, le FDR d'adapt-cisl variait entre 0,01 et 0,10 et sa sensibilité variait entre 0,06 et 0,72, tandis que le FDR d'adapt-bic variait entre 0,03 et 0,12 et sa sensibilité variait entre 0,03 et 0,68.

Les résultats de simulations pour les scénarios 16 à 27, c'est-à-dire pour les scénarios

où les vrais prédicteurs sont notifiés entre 20 et 100 fois, sont présentés dans la figure 3.2 pour toutes ces approches. Les mêmes comportements ont été observés en termes de fausses découvertes lorsque les vrais prédicteurs étaient peu notifiés. Comme on pouvait s'y attendre, toutes les approches ont montré une sensibilité bien moindre dans ces scénarios par rapport aux scénarios 4 à 15. Adapt-cv et adapt-univ ont montré un FDR plus élevé que les autres approches dans ces scénarios également. Elles ont également été les approches les plus sensibles. Adapt-univ-bic a montré un FDR plus faible dans ces scénarios par rapport aux scénarios 1 à 15, avec un FDR toujours inférieur à 0,10. Dans l'ensemble, nos propositions adapt-cisl et adapt-bic ont donné les meilleurs résultats, avec un FDR qui est resté équivalent ou plus faible que celui d'adapt-univ-bic, et une sensibilité équivalente ou meilleure. Ces résultats ont été particulièrement visibles dans les scénarios 16 à 19 et 21.

Comme les approches adapt-cisl et adapt-bic ont montré les meilleures performances parmi les approches de détection basées sur le lasso adaptatif, nous n'avons retenu que ces deux méthodes dans la comparaison avec les autres méthodes. La figure 3.3 montre le FDR et la sensibilité moyenne de nos deux approches, de RFET, des approches basées sur le lasso et de celles basées sur le PS en grande dimension pour les scénarios où les vrais prédicteurs étaient notifiés plus de 100 fois (scénarios 4 à 15). La figure 3.4 montre les mêmes résultats lorsque les vrais prédicteurs étaient notifiés entre 20 et 100 fois (scénarios 16 à 27). Les résultats pour les scénarios où il n'y avait pas de vrais prédicteurs (1 à 3) sont indiqués dans le tableau 3.3.

Lasso-cv, ps-iptwT ont eu les meilleures performances en termes de sensibilité mais ils ont tous deux montré un FDR élevé dans tous les scénarios. Dans une moindre mesure, RFET et ps-adjust ont montré le même comportement dans les scénarios 7,11 et 15 avec un FDR allant jusqu'à 0,62 pour RFET et jusqu'à 0,34 pour ps-adjust. RFET a également montré ce comportement dans les scénarios 4 à 6. Dans tous les autres scénarios, ils ont tous deux affiché un FDR plutôt faible. Dans l'ensemble, ps-adjust s'est mieux comporté que RFET en termes de sensibilité et de FDR dans les scénarios 4 à 15, néanmoins RFET a montré un FDR plus faible dans les scénarios 16 à 27. A l'opposé, la méthode ps-mw s'est montrée très conservatrice : son FDR est resté très faible dans les différents scénarios, parfois même beaucoup plus bas que le seuil fixé de 5 %. Sa sensibilité a baissé et est

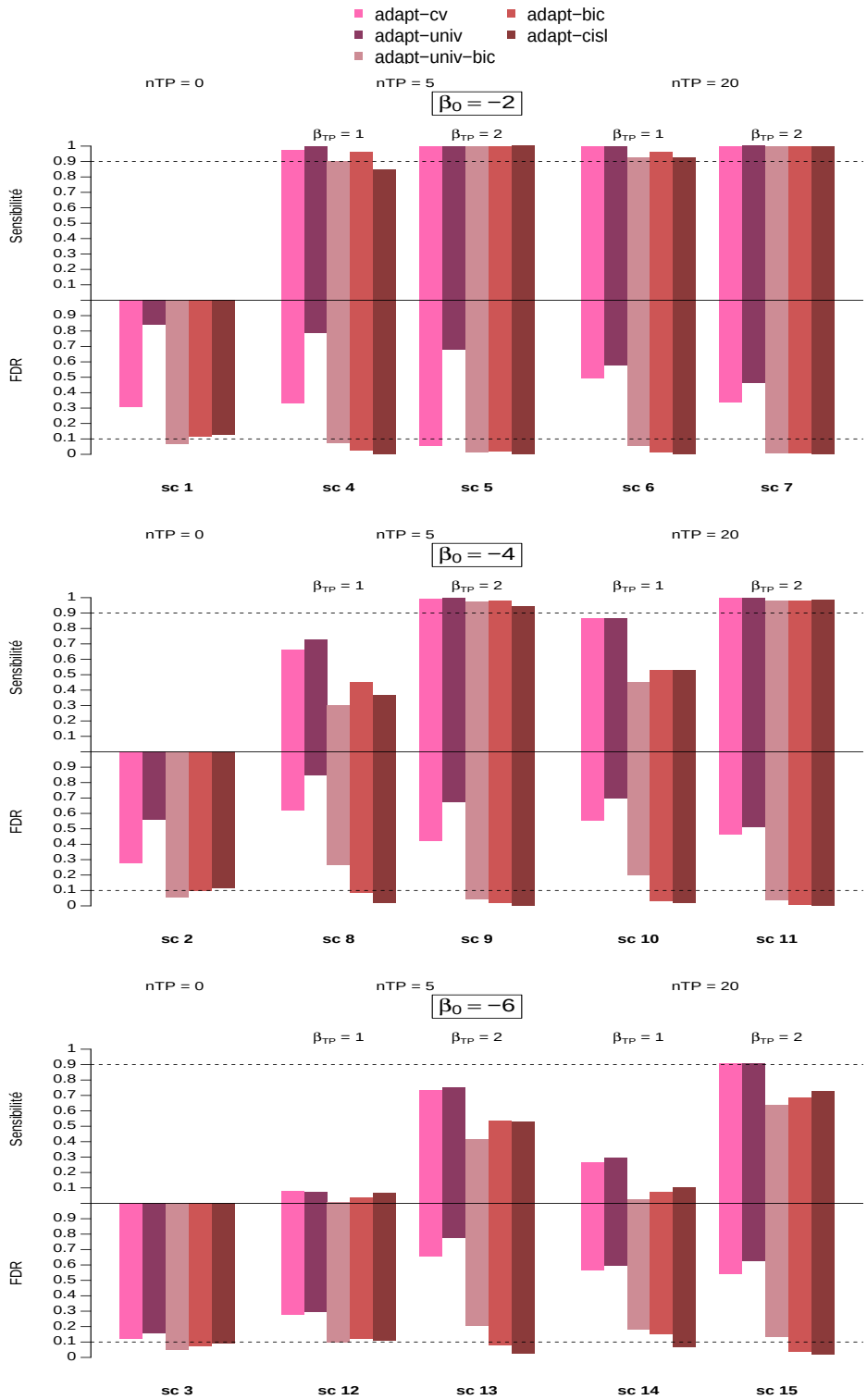


FIGURE 3.1 – Sensibilité et FDR des méthodes basées sur le lasso adaptatif pour les scénarios 1 à 15.

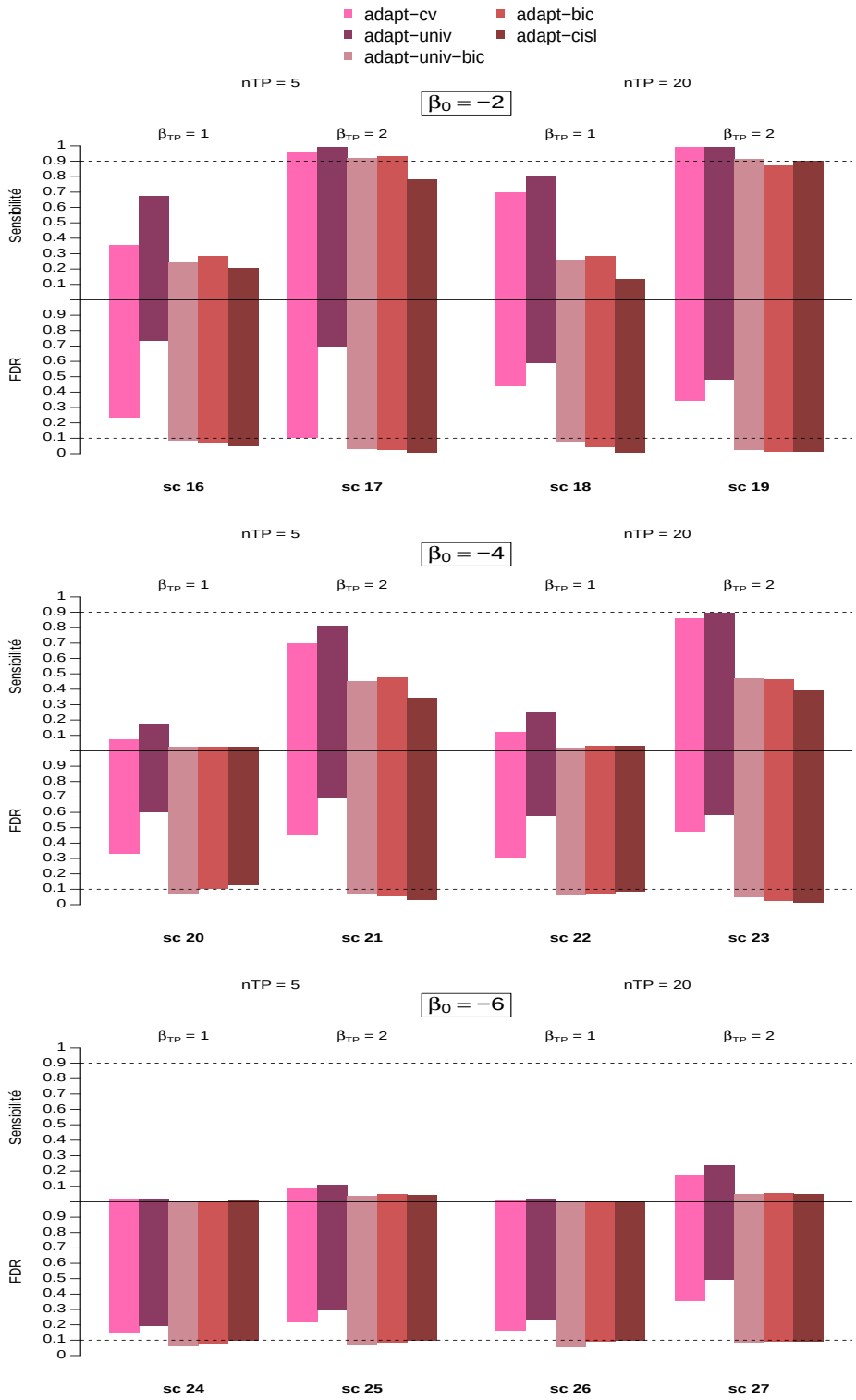


FIGURE 3.2 – Sensibilité et FDR des méthodes basées sur le lasso adaptatif pour les scénarios 16 à 27.

même devenue nulle à mesure que la réponse était plus rare et que les vrais prédicteurs étaient moins notifiés (scénarios 12 à 15, 16, 18 et 20 à 27). L'approche ps-iptw a donné de très mauvais résultats dans tous les scénarios avec une sensibilité très faible et un FDR extrêmement élevé. Parmi les approches basées sur le lasso autres que lasso-cv, les performances de lasso-perm ont été mauvaises en termes de fausses découvertes dans les scénarios où il n'y avait pas de vrais prédicteurs (scénarios 1 à 3) avec un FDR autour de 0,35, et dans les scénarios où la réponse était déséquilibrée, que ce soit avec des vrais prédicteurs fréquents (scénarios 12 à 15) ou rares (scénarios et 24 à 27). CISL et lasso-bic ont montré de très bonnes performances avec une sensibilité acceptable et un FDR faible dans la majorité des scénarios. Lorsque la réponse était rare, c'est-à-dire $\beta_0 = -6$, CISL a montré une augmentation de son FDR. Cette augmentation était notable dans les scénarios 3, 12 et surtout dans les scénarios 24 à 26, avec un FDR entre 0,15 et 0,20. Dans l'ensemble, le lasso-bic avait une sensibilité et un FDR légèrement supérieurs à ceux de CISL. Bien que le lasso-bic ait eu un comportement assez stable, il a montré une augmentation surprenante de son FDR dans les scénarios 14 et 20, avec un FDR à 0,15 et 0,10, respectivement.

Par rapport au lasso-bic, nos propositions ont montré une sensibilité équivalente ou légèrement inférieure et un FDR plus faible dans tous les scénarios. En particulier, cette différence de FDR était visible dans les scénarios 12 à 15. Comme lasso-bic, adapt-bic et adapt-cisl ont montré une augmentation en termes de FDR pour le scénario 20 avec un FDR de 0,12 pour adapt-cisl et de 0,10 pour adapt-bic.

Méthodes	FDR		
	scénario 1 : $\beta_0 = -2$	scénario 2 : $\beta_0 = -4$	scénario 3 : $\beta_0 = -6$
RFET	0.00	0.01	0.00
lasso-cv	0.31	0.28	0.12
lasso-bic	0.12	0.10	0.07
lasso-perm	0.37	0.41	0.35
CISL	0.06	0.04	0.18
adapt-bic	0.11	0.10	0.07
adapt-cisl	0.13	0.11	0.09
ps-adjust	0.01	0.15	0.14
ps-mw	0.00	0.00	0.00
ps-iptw	0.99	1.00	1.00
ps-iptwT	0.23	0.35	0.29

Tableau 3.3 – FDR d'adapt-bic et adapt-cisl, de RFET, des approches basées sur le lasso et de celles basées sur le score de propension en grande dimension pour les scénarios 1 à 3.

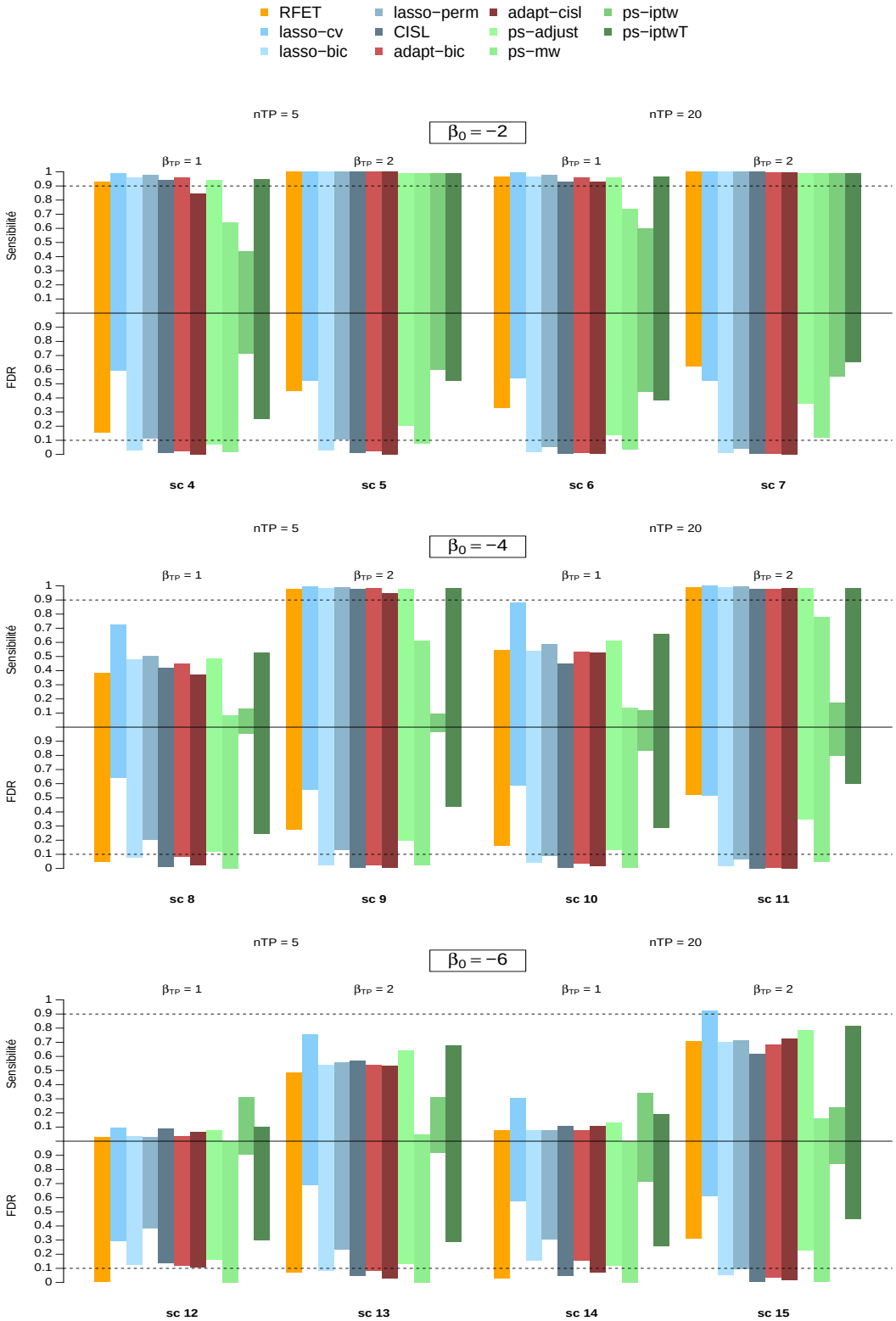


FIGURE 3.3 – Sensibilité et FDR d'adapt-bic et adapt-cisl, de RFET, des approches basées sur le lasso et de celles basées sur le score de propension en grande dimension pour les scénarios 4 à 15.

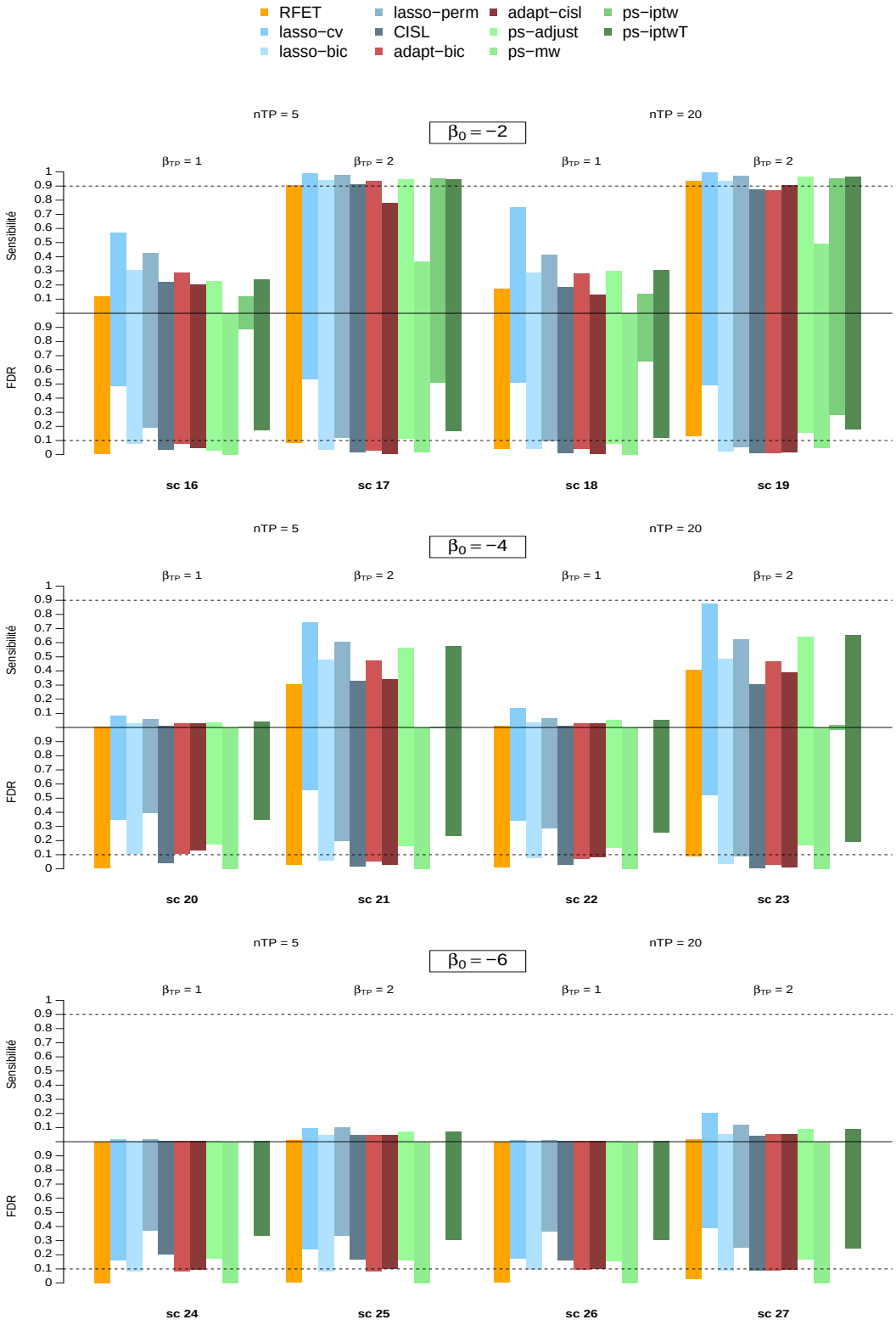


FIGURE 3.4 – Sensibilité et FDR d'adapt-bic et adapt-cisl, de RFET, des approches basées sur le lasso et de celles basées sur le score de propension en grande dimension pour les scénarios 16 à 27.

La figure 3.5 représente la distribution du FDR obtenu sur les 27 scénarios de simulation pour chaque approche. En plus de fournir des faibles valeurs de FDR, les méthodes lasso-bic, CISL, adapt-bic, adapt-cisl et ps-mw ont montré une faible variabilité de leur FDR sur l'ensemble des scénarios de simulation. A l'inverse les méthodes RFET, lasso-cv, lasso-perm et ps-iptw ont montré une grande variabilité de leur FDR sur les différents scénarios. Ps-adjust et ps-iptwT ont montré une variabilité de leur FDR moindre, mais dans l'ensemble plus élevé que 10%. Les méthodes ps-mw et ps-iptw ont montré des valeurs de sensibilités plus faibles.

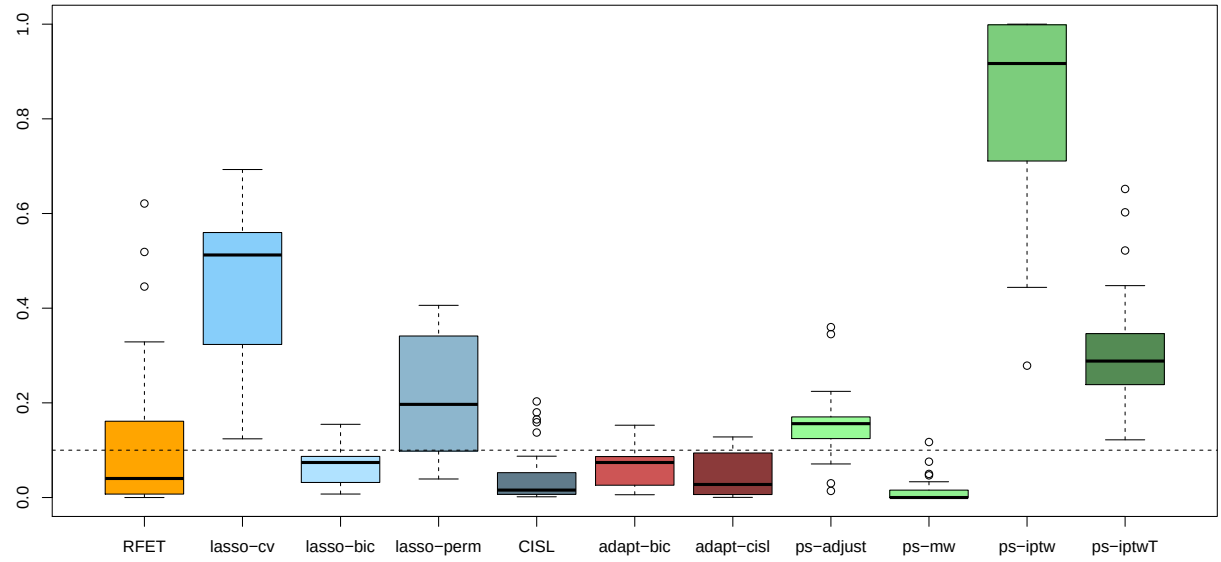


FIGURE 3.5 – Distribution du *False Discovery Rate* obtenu pour les 27 scénarios de simulations.

3.6 Application aux données réelles

3.6.1 Paramétrage de la comparaison

Nous avons mis en œuvre l'ensemble des méthodes de détection de signaux présentées ci-dessus sur l'extraction de la BNPV utilisée pour déterminer les matrices d'expositions médicamenteuses dans l'étude de simulations. Cette extraction de la BNPV comprenait 452 914 notifications avec 6 617 d'EI différents et 2 378 médicaments différents dont 1 692 médicaments notifiés plus de 10 fois.

Nous avons utilisé le même ensemble de signaux de référence que précédemment à savoir l'ensemble DILIrank présenté en section 2.5.4. Sur cette nouvelle extraction de la BNPV, cela a conduit à considérer 25 187 notifications de l'évènement DILI. Parmi les médicaments notifiés plus de 10 fois dans la BNPV, 1 136 avaient plus de trois notifications en commun avec un évènement DILI. Comme dans l'étude de simulations, les approches basées sur le PS et RFET ont été mises en œuvre pour ces 1 136 médicaments et les 556 autres ont eu leur p-valeur fixée à un. Au final, l'ensemble de signaux de référence DILIrank contenait 203 témoins négatifs et 133 témoins positifs parmi les 1 692 médicaments. Sur les 1 136 médicaments retenus après filtrage, DILIrank contenait 123 témoins négatifs et 119 témoins positifs. Les performances de toutes les approches sont mesurées en termes de FDP, de sensibilité et de spécificité.

3.6.2 Résultats

Le tableau 3.4 présente les performances des méthodes calculées à partir de l'ensemble des signaux de référence DILIrank. Malgré la grande variabilité en termes de nombre de signaux générés, on observe que huit méthodes (à savoir lasso-cv, lasso-bic, lasso-perm, adapt-cv, adapt-univ, adapt-unv-bic, adapt-bic, ps-adjust) parmi 14 ont montré un FDP et une sensibilité dans les mêmes ordres de grandeur : un FDP aux alentours de 7% et une sensibilité avoisinant les 44%. RFET, ps-iptw et ps-iptwT ont eu les valeurs de FDP les plus élevées : 15,9 %, 20 % et 14,3 % respectivement. Ps-iptw a affiché la plus faible valeur de sensibilité : 9 %. Comme dans les simulations, les méthodes adapt-cisl et adap-bic ont montré de bonnes performances en termes de fausses découvertes et adapt-cisl a généré le moins de faux positifs avec deux fausses découvertes. Elles ont montré une FDP à 3,3 % et

Méthodes	Nombre de signaux	Nombre de signaux avec un statut connu (positif ou négatif)	Nombre de faux positifs	FDP (%)	Spécificité (%)	Sensibilité (%)
RFET	249	82	13	15,9	93,6	51,9
lasso-cv	220	79	5	6,3	97,5	55,6
lasso-bic	170	65	5	7,7	97,5	45,1
lasso-perm	158	60	4	6,7	98,0	42,1
CISL	109	48	2	4,2	99,0	34,6
adapt-cv	188	70	5	7,1	97,5	48,9
adapt-univ	179	65	4	6,2	98,0	45,9
adapt-univ-bic	163	61	4	6,6	98,0	42,9
adapt-bic	151	60	4	6,7	98,0	42,1
adapt-cisl	153	60	2	3,3	99,0	43,6
ps-adjust	209	71	6	8,5	97,0	48,9
ps-mw	86	40	0	0,0	100,0	30,1
ps-iptw	49	15	3	20,0	98,5	9,0
ps-iptwT	260	84	12	14,3	94,1	54,1

Tableau 3.4 – Performances de chaque méthode en termes de nombre de signaux générés, de proportion de fausses découvertes (FDP), de spécificité et de sensibilité. Les statistiques résumées sont calculées sur la base des médicaments dont le statut est connu selon l'ensemble de référence DILIrank.

6,7% respectivement. Certaines méthodes telles que lasso-cv et adapt-univ ont montré de meilleures performances que dans les simulations. La méthode ps-mw n'a détecté aucun faux positif.

La figure 3.6-(A) montre la répartition des signaux générés par adapt-cisl, adapt-bic et lasso-bic et la figure 3.6-(B) montre cette répartition pour adapt-cisl, adapt-bic et CISL. Parmi les signaux générés, les vrais positifs et les faux positifs selon DILIrank sont également représentés. La figure 3.6-(A) montre que tous les signaux générés par nos deux propositions ont également été générés par le lasso-bic : 140 signaux étaient communs aux trois méthodes, 13 ont été générés par l'adapt-cisl et le lasso-bic, et 11 ont été générés par l'adapt-bic et le lasso-bic. Il y avait 54 vrais positifs et deux faux positifs parmi les 140 signaux générés par les trois méthodes. Parmi les signaux générés uniquement par adapt-cisl ou adapt-bic et communs avec ceux du lasso-bic, adapt-cisl a généré quatre vrais positifs et aucun faux positif, alors qu'adapt-bic était un peu moins performant avec deux vrais positifs et deux faux positifs.

La figure 3.6-(B) montre que tous les signaux générés par CISL ont également été

générés par adapt-cisl. Adapt-bic n'a généré aucun signal en commun avec CISL seulement. Les 11 signaux générés par adapt-bic seul en figure 3.6-(B) comprenaient deux vrais positifs et deux faux positifs, c'est-à-dire les mêmes que ceux générés par lasso-bic. Dans l'ensemble, adapt-cisl a obtenu de bons résultats puisque les deux seuls faux positifs détectés par cette approche l'ont été par toutes les méthodes considérées ici. Elle a généré plus de signaux que CISL en détectant 12 vrais positifs supplémentaires mais sans détecter de faux positifs. Ce n'était pas le cas du lasso-bic et d'adapt-bic.

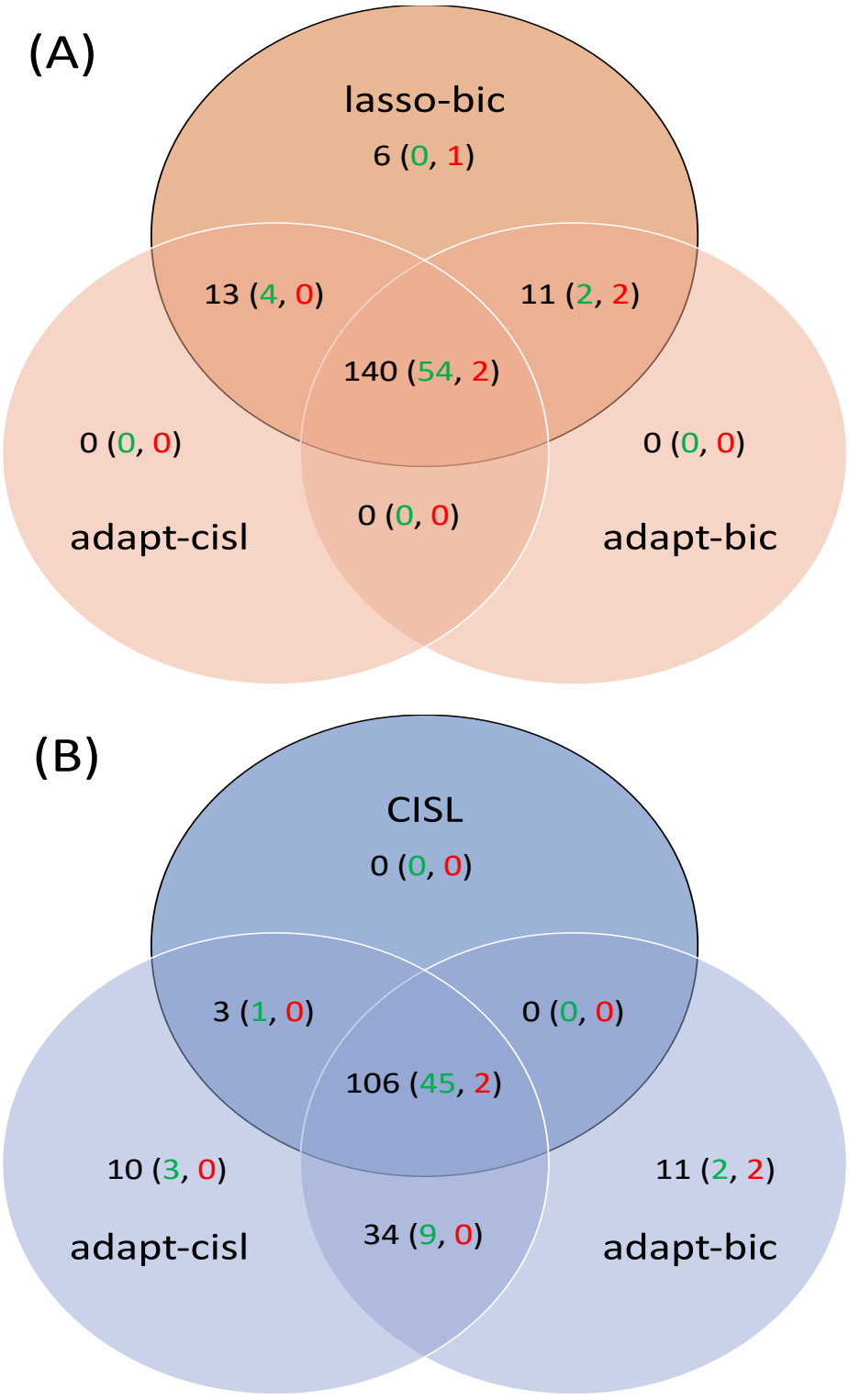


FIGURE 3.6 – Répartition des signaux générés par adapt-cisl, adapt-bic et **(A)** lasso-bic ; **(B)** CISL. Parmi les signaux générés, les vrais positifs sont en vert et les faux positifs en rouge.

3.7 Discussion

Cette étude nous a permis d'explorer de nouvelles approches pour la détection des signaux basées sur une méthodologie appropriée pour la sélection des variables : le lasso adaptatif. En plus de définir de nouveaux poids adaptatifs dans le contexte de la grande dimension, nous avons utilisé le BIC pour effectuer la sélection de variables. Pour évaluer les performances de nos stratégies, nous avons réalisé une vaste étude de simulations et une application sur des données réelles. Nous y avons comparé nos approches à d'autres versions du lasso adaptatif en grande dimension proposées dans la littérature, ainsi qu'à des approches de détection basées sur la régression lasso. Nous avons également inclus dans cette comparaison les approches basées sur le PS en grande dimension présentées au chapitre précédent.

En comparant dans un premier temps toutes les approches basées sur le lasso adaptatif, nous constatons que les poids adaptatifs proposés et l'utilisation du BIC pour la sélection des variables sont pertinents. Ensuite, une comparaison étendue aux autres méthodes de détection montre que nos propositions sont particulièrement compétitives. Nos approches montrent des performances très satisfaisantes en termes de fausses découvertes, notre principal critère d'intérêt. En plus d'afficher un FDR faible, nos approches montrent un comportement stable sur tous les scénarios de simulation. Par rapport à la régression lasso où le BIC est utilisé pour effectuer la sélection des variables, lasso-bic, nos propositions tendent à montrer un FDR plus faible au prix d'une sensibilité légèrement plus faible. Pour l'adapt-bic, ce dernier résultat n'est pas surprenant puisque par construction, les variables sélectionnées par cette approche sont un sous-ensemble des variables sélectionnées par l'approche lasso-bic.

Nos travaux confirment également que CISL est une approche pertinente pour la détection de signaux. Le choix du quantile, que nous avons fixé à 10%, semble approprié pour un grand nombre de situations, sauf lorsque la réponse et les vrais prédicteurs sont rares. Comme attendu, la validation croisée n'est pas appropriée pour la détection de signaux, compte tenu des faibles performances en termes de FDR de toutes les approches basées sur ce critère : lasso-cv, adapt-cv et adapt-univ. La méthode de permutation avec la régression lasso, lasso-perm, n'a pas donné de résultats pleinement satisfaisants avec de mauvaises performances dans certains scénarios.

Parmi les approches basées sur PS pour la détection de signaux, nous avons pu approfondir et compléter notre travail présenté au chapitre précédent. Certains résultats viennent étayer nos précédentes conclusions. La pondération sur le score de propension avec les poids MW a montré de très bons résultats lorsque la réponse était fréquente, mais est devenue très conservatrice à mesure que la réponse se faisait plus rare. Ce comportement n'est pas à l'avantage de cette approche, car il est fréquent dans les données de pharmacovigilance d'avoir des EI très peu notifiés. L'ajustement sur le PS n'a pas montré de meilleures performances pour la détection car, bien que cette approche ait montré une sensibilité élevée, elle a aussi affiché un FDR assez élevé pour plusieurs scénarios de simulations. L'approche ps-iptw a montré de très mauvaises performances dans tous les scénarios, comme cela a été constaté au chapitre précédent. La troncature de ces poids améliore les résultats, mais cette approche reste une méthode de détection de signaux insatisfaisante car elle conduit à un nombre important de fausses découvertes. Bien que l'on effectue une procédure de correction de tests multiples censée assurer un contrôle du FDR au seuil de 5%, on constate que les approches par ajustement ou pondération IPTW (avec et sans troncature) sur le PS ont un FDR supérieur à ce seuil en simulation. C'est le cas également pour la procédure RFET, soumise à la même correction de tests multiples.

Dans l'ensemble, les résultats obtenus dans l'application aux données réelles viennent confirmer ceux obtenus dans les simulations. Parmi toutes les approches testées, notre approche adapt-cisl a montré le meilleur compromis entre un FDR très faible, une sensibilité acceptable et un nombre raisonnable de signaux générés.

Les deux poids adaptatifs définis en section 3.3 ont été construits dans le but de tirer parti de méthodes de détection de signaux que l'on savait pertinentes à savoir lasso-bic et CISL. Ainsi les premiers poids adaptatifs ont été construits en suivant le schéma proposé par Bühlmann et van de Geer (2011) ou encore Huang *et al.* (2008a) : avoir recours à une première régression lasso pour déterminer les poids adaptatifs à utiliser dans un second temps pour le lasso adaptatif. Plutôt que d'utiliser le critère de validation croisée dans la première régression lasso, nous avons choisi d'utiliser un critère a priori plus adapté pour la sélection de variables : le BIC. Les deuxièmes poids adaptatifs ont été construits à partir de la quantité (3.2) estimée par CISL. La construction de ces poids permet de s'affranchir du choix d'une valeur de quantile, et exploiter l'intégralité de l'information portée par

CISL. Nous avons choisi de baser nos poids sur le nombre de fois que la quantité estimée $\widehat{\tau}$ était non nulle, mais on pourrait imaginer utiliser d'autres informations relatives à la distribution de $\widehat{\tau}$ pour construire les poids adaptatifs dans adapt-cisl, comme la médiane ou le maximum.

Dans ce travail, nous avons aussi utilisé le BIC comme critère de sélection de variables avec le lasso adaptatif. Au vu des résultats du chapitre précédent, ce dernier nous a semblé pertinent. Les résultats des simulations viennent nous conforter dans ce sens, en particulier en comparant les deux approches adapt-univ (qui repose sur la validation croisée) et adapt-univ-bic. Hormis le critère de sélection de la pénalité dans l'étape du lasso adaptatif, rien de diffère entre ces deux approches de détection. On constate dans l'étude de simulations que ce choix influence grandement les résultats, avec de bien meilleures performances en termes de fausses découvertes pour l'approche qui repose sur le BIC. Cet écart de performance est cependant moins évident dans l'application aux données réelles.

L'utilisation du BIC dans le cadre spécifique du lasso adaptatif a été étudiée par Hui *et al.* (2015) qui ont proposé le critère ERIC (*Extended Regularized Information Criterion*). Dans leur travail, ils ont considéré le BIC obtenu à partir de la vraisemblance pénalisée, et ont fait valoir que ce BIC ne permet pas de prendre en compte la diminution de la variance des estimations due à l'information portée par les poids adaptatifs. En utilisant le BIC qui est basé sur la vraisemblance non pénalisée, nous évitons certains des problèmes soulevés par Hui *et al.* (2015). Il serait cependant intéressant de comparer la sélection des variables pour le lasso adaptatif opérée avec le BIC non pénalisé et celle obtenue avec ERIC dans notre contexte.

Pour préserver les spécificités des données de pharmacovigilance, à savoir une dimension importante et un grand déséquilibre, nous avons basé notre étude de simulations sur des données réelles. Nous avons fait varier le nombre et la fréquence des vrais prédicteurs, leur force d'association avec la réponse, ainsi que la rareté de cette dernière. Cependant, avec cette stratégie, il n'a pas été possible de faire varier la structure de corrélation entre les vrais prédicteurs et les autres variables. Une extension possible de ce travail serait de développer une stratégie plus complexe de génération de données simulées pour permettre de faire varier ce paramètre.

3.8 Conclusion

Les résultats des simulations et de l'application suggèrent que le lasso adaptatif est une méthodologie prometteuse pour la pharmacovigilance lorsque les poids adaptatifs et le critère de sélection de variables sont choisis de manière appropriée. Du point de vue du temps de calcul, à l'instar des approches lasso, ces approches adaptatives s'avèrent nettement moins coûteuses que les approches reposant sur les scores de propension en grande dimension.

Ce travail nous a aussi permis d'étudier plus avant les performances des approches basées sur le score de propension présentées au chapitre précédent. Grâce à l'étude de simulations, nous avons pu constater que l'approche par pondération MW qui avait notre préférence au vu des résultats de l'étude empirique, n'est pas appropriée dans les situations où l'EI est peu notifié, et lorsque les médicaments associés à la réponse sont peu notifiés également. Cette approche fait toujours preuve d'un FDR très faible, mais sa sensibilité est quasi nulle dès que la réponse d'intérêt est très déséquilibrée.

Ce travail nous a aussi permis de faire la part des choses entre plusieurs critères de sélection du paramètre λ pour la régression lasso dans le cadre de la pharmacovigilance. Il nous a ainsi confortée sur la pertinence de l'utilisation du critère lié au BIC. La validation croisée est quant à elle définitivement peu adaptée à notre contexte. La méthode de sélection de la pénalité basée sur des permutations ne s'est pas avérée pleinement satisfaisante ici.

Nos propositions de poids adaptatifs sont ici mises en œuvre dans le cadre de la détection en pharmacovigilance, mais pourraient être utilisées dans un autre contexte.

3.9 Valorisation de l'étude

Ce travail a fait l'objet d'une soumission auprès de la revue *Statistical Methods in Medical Research*. Le manuscrit soumis est disponible en annexe D.2.

Ce travail a été présenté lors d'un congrès international sous forme de poster avec courte présentation orale lors de la 41$^{\text{ème}}$ conférence de l'*International Society for Clinical Biostatistics* (août 2020, conférence virtuelle).

Toutes les méthodes de détection présentées dans ce travail ont été implémentées dans un package R disponible sur le CRAN : `adapt4pv`.

Chapitre 4

Apport des bases médico-administratives pour la détection de signaux en pharmacovigilance

4.1 Introduction

Les bases de notifications spontanées restent la pierre angulaire pour la surveillance de la sécurité des médicaments une fois qu'ils sont introduits sur la marché. Cependant, de part la nature de leur recueil, les notifications spontanées souffrent de certaines limites, notamment une importante sous déclaration des évènements indésirables (Bégaud *et al.*, 2002). Cette sous notification est multi-factorielle. Elle peut être liée, entre autres, à la faible gravité de l'EI impliqué, ou au fait que cet EI soit bien connu. A l'inverse, la médiatisation d'un risque médicamenteux peut entraîner une notification plus importante des cas observés ou des cas du même évènement indésirable liés à des médicaments de la même classe que le médicament visé par les médias (Weber, 1984; Pariente *et al.*, 2007). En outre, les bases de notifications ne comprenant que des individus ayant expérimenté au moins un évènement indésirable, la population couverte par ces bases de données n'est pas représentative de la population générale.

Les bases médico-administratives (BMA) sont alors apparues comme des sources de

données alternatives prometteuses pour la pharmacovigilance. Bien qu'elles n'aient pas été créées pour la recherche médicale, ces bases de données contiennent des informations relatives à l'ensemble des actes de soins remboursés, afin de pouvoir procéder à l'enregistrement et au remboursement de ces actes pour les bénéficiaires. En fonction des pays, les informations enregistrées dans ces bases peuvent varier car elles sont conditionnées par le système de remboursement des prestations en place.

Acté par la loi de modernisation de notre système de santé de janvier 2016 [1], le SNDS regroupe les données du Système National d'Information Inter Régimes de l'Assurance Maladie (SNIIRAM), du Programme de Médicalisation des Systèmes d'Information (PMSI) et celles relatives aux causes médicales de décès. Il est prévu que le SNDS comprenne également les données relatives au handicap et certaines données d'organismes complémentaires, comme les mutuelles privées. Créé en 1999, le SNIIRAM contient l'ensemble des consommations de soins qui donnent lieu à un remboursement, comme les consultations de praticiens libéraux médicaux et paramédicaux, les médicaments délivrés en officine de ville, les prélèvements biologiques, etc (Tuppin *et al.*, 2017). Le PMSI regroupe les données des hôpitaux du secteur public et du secteur privé. La base des causes médicales de décès est gérée par le centre d'épidémiologie sur les causes médicales de décès et regroupe les informations issues des certificats médicaux de décès. Le SNDS couvre à l'heure actuelle la quasi totalité de la population française. L'arrêté du 20 juin 2005 acte la création au sein du SNIIRAM et du PMSI d'un « échantillon généraliste de ces données représentatif des personnes protégées des régimes [..] afin d'assurer le suivi de la consommation de soins et des taux de recours aux soins [2] ». Cet échantillon généraliste des bénéficiaires (EGB), échantillon représentatif au $1/97^{\text{ème}}$ de la population française, contient les données du SNIIRAM depuis 2003 et celle du PMSI depuis 2007. Il a été construit pour permettre un suivi sur 20 ans bénéficie d'un accès facilité pour la communauté des chercheurs.

L'exploitation des BMA pour répondre à des questions de recherche en santé est rendue difficile par le fait que ces bases n'ont pas été conçues avec cet objectif. En particulier, les informations cliniques ou paracliniques qui motivent la prescription de certains médicaments n'y sont généralement pas renseignées (Schneeweiss et Avorn, 2005).

1. Article 193 de la loi n°2016-41 disponible sur le site Légifrance.gouv.
2. Article 3 alinéa 4 de l'arrêté, disponible sur le site Légifrance.gouv.

L'absence de ces nombreuses variables d'ajustement qui sont potentiellement des facteurs de confusion, peut entraîner des biais importants dans la conduite d'étude pharmacoépidémiologiques. La gestion de cette confusion non mesurée a donc suscité des efforts de recherche importants (Uddin *et al.*, 2016).

Dans le domaine de la pharmacovigilance, des méthodes pour exploiter les BMA pour la détection de signaux ont été développées. Bon nombre de ces méthodes ont été testées dans le cadre du projet mené par l'OMOP (voir section 1.2). On peut classer ces différentes méthodes en plusieurs catégories :

- les méthodes de disproportionnalité (Curtis *et al.*, 2009; Choi *et al.*, 2010, 2011; Kim *et al.*, 2011; Zorych *et al.*, 2013), ou des méthodes dérivées adaptées aux données longitudinales comme le *Longitudinal Gamma Poisson Shrinkage* (Schuemie, 2011)
- les méthodes basées sur des designs classiques en épidémiologie comme les études appariées cas-témoins, les études de cohorte ou les designs auto-contrôlés (Maclure, 1991; Farrington, 1995),
- la *sequence symmetry analysis* (SSA) (Hallas, 1996)
- les approches basées sur des règles d'associations temporelles (Jin *et al.*, 2006, 2010; Norén *et al.*, 2010).

Un des résultats du projet OMOP a été de démontrer la pertinence d'une méthode autocontrôlée : la série de cas (Farrington, 1995). Avec ce type de méthodes, seuls les individus qui ont vécu l'EI d'intérêt sont considérés. Les cas sont leur propre témoin ce qui permet de prendre en compte les facteurs de confusion individuels non dépendants du temps.

Dans ce chapitre, nous avons dans un premier temps cherché à mettre en œuvre des approches de détection de signaux dans l'EGB. Nous avons choisi les approches parmi celles présentant les meilleurs résultats dans la littérature au moment où ces analyses ont été menées. Dans un second temps, nous avons tenté d'exploiter à la fois la BNPV et l'EGB en testant des approches de détection basées sur une exploitation conjointe des données. Les performances de ces différents travaux sont évaluées avec le même ensemble de référence que dans les chapitres précédents : DILIrank.

4.2 Détection de signaux dans l'Échantillon Généraliste des Bénéficiaires

L'OMOP, qui s'est développée et a donné naissance au consortium *Observational Health Data Sciences and Informatics* (OHDSI), a inclus par la suite une autre méthode auto-contrôlée dans son catalogue de méthodes de détection : le *case-crossover* (CCO) (Maclure, 1991). Cette méthode n'a pas été testée dans le travail publié en 2013. Tout comme la série de cas, le CCO a été développé afin de déterminer l'effet d'une exposition courte sur un EI aigu. Néanmoins il a été montré que son utilisation restait pertinente dans l'étude de l'effet d'une exposition prolongée sur des EI dont l'apparition est retardée (Wang *et al.*, 2004). Avec cette approche, seule la survenue du premier évènement est considérée.

Une revue de la littérature a repris le constat fait par l'OMOP concernant les méthodes auto-contrôlées et a également souligné les bonnes performances de la SSA pour la détection de signaux sur les BMA (Arnaud *et al.*, 2017). Tout comme les méthodes auto-contrôlées, la SSA permet de considérer uniquement la population des individus qui ont expérimenté l'EI d'intérêt. En revanche, la SSA utilise l'information d'autres cas comme référence et non l'information des cas eux-mêmes, contrairement à la série de cas ou au CCO. Ainsi, on ne peut pas parler formellement de la SSA comme d'une méthode auto-contrôlée (Hallas et Pottegård, 2014).

Nous avons mis en œuvre deux méthodes de détection de signaux dans l'EGB qui nous ont semblé les plus pertinentes au moment où ces analyses ont été réalisées. Nous avons considéré des approches de détection basées sur la SSA et le CCO, avec une variation de cette dernière méthode qui permet de prendre en compte les coprescriptions et d'avoir recours à des régressions pénalisées de type lasso (Avalos *et al.*, 2012; Suchard *et al.*, 2013; Volatier *et al.*) . Les performances de toutes les approches de détection considérées sont mesurées en termes de sensibilité et de spécificité, de PPV et de FDP.

4.2.1 *Case-Crossover* : définition et mise en œuvre

Méthodes

Avec le *case-crossover*, la période d'observation considérée comme « période cas » correspond à l'intervalle de temps juste avant la survenue de l'EI, et la ou les « périodes témoins » sont définies sur une ou plusieurs périodes de temps antérieure(s) à la date de survenue de l'EI. La figure 4.1 représente schématiquement le design de cette approche.

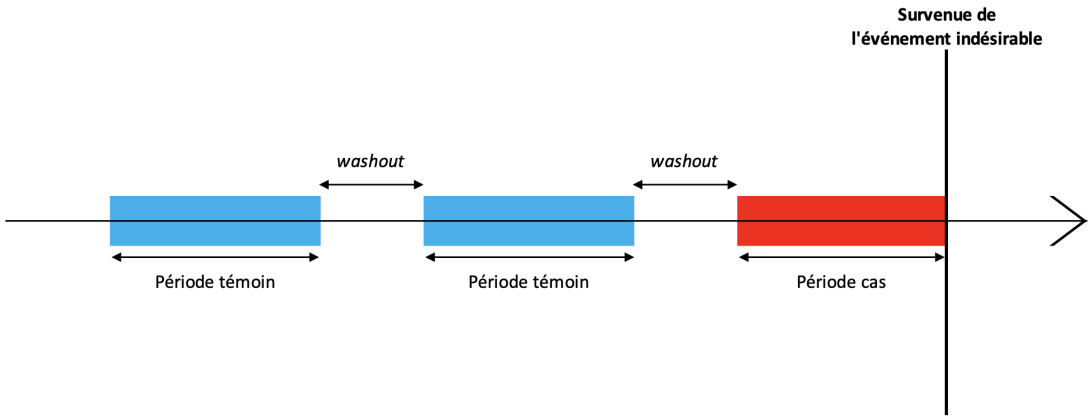


FIGURE 4.1 – Schéma du design du *case-crossover*

Dans ce découpage de la période d'étude, plusieurs contraintes sont à prendre en compte (DELANEY et SUISSA, 2009). Afin d'assurer l'indépendance entre la période cas et les différentes périodes témoins, une ou plusieurs périodes dites de *washout* sont nécessaires. Pour respecter l'hypothèse d'echangeabilité selon laquelle les différentes périodes d'observation sont comparables en termes de facteurs de confusion variant dans le temps, les périodes témoins doivent être proches dans le temps de la période cas. Ainsi formalisé, le CCO peut être vu comme une version du design cas-témoins. L'outil classiquement utilisé pour l'analyse des données appariées est la régression logistique conditionnelle, qui permet d'introduire un terme d'intercept différent par strate d'individus appariés. En plus des notations définies en section 2.2, on note T le nombre de périodes considérées par individu, qui consistent donc en une période cas et $T-1$ périodes témoins. Pour cette section uniquement, la matrice de données $\mathbf{X}$ sera de dimension $(N \times T) \times P$: on compte par individu autant de lignes que de périodes considérées. De même, la réponse d'intérêt est ici définie par $\mathbf{y} \in \mathbb{R}^{(N \times T)}$. On note $x_{it} = (x_{it1}, .., x_{itP})$ le vecteur de covariables observées

pour l'individu $i \in \{1, ..., N\}$ sur la période $t \in \{1, ..., T\}$. On considère le modèle :

$$P_{it} = P(y_{it} = 1|x_{it}) = \frac{\exp(\alpha_i + \sum_{j=1}^{P} \beta_j \ x_{itj})}{1 + \exp(\alpha_i + \sum_{j=1}^{P} \beta_j \ x_{itj})} = \frac{\exp(\alpha_i + x_{it}\beta)}{1 + \exp(\alpha_i + x_{it}\beta)}, \tag{4.1}$$

avec $\beta \in \mathbb{R}^P$ le vecteur de paramètres d'intérêt. On peut réécrire le modèle (4.1) de la manière suivante :

$$\text{logit}(P_{it}) = \log\left(\frac{P_{it}}{1 - P_{it}}\right) = \alpha_i + x_{it}\beta.$$

Pour une strate i, il n'y a qu'une seule valeur de $t \in \{1, .., T\}$ telle que $y_{it} = 1$. Dans le cas du *case-crossover*, c'est la dernière période d'observation qui remplit cette condition, on a alors $y_{iT} = 1$. Ainsi, conditionnellement au fait qu'il n'y a qu'un seul cas par strate, la vraisemblance conditionnelle pour un individu donné s'écrit :

$$\begin{aligned}
\frac{P_{iT} \times \prod_{t=1}^{T-1}(1 - P_{it})}{\sum_{l=1}^{T} P_{il} \prod_{t=1, t\neq l}^{T}(1 - P_{it})} &= \frac{\frac{P_{iT}}{1-P_{iT}} \prod_{t=1}^{T}(1 - P_{it})}{\sum_{l=1}^{T} \frac{P_{il}}{1-P_{il}} \prod_{t=1}^{T}(1 - P_{it})} \\
&= \frac{\frac{P_{iT}}{1-P_{iT}}}{\sum_{l=1}^{T} \frac{P_{il}}{1-P_{il}}} \\
&= \frac{\exp(\alpha_i) \exp(x_{iT}\beta)}{\exp(\alpha_i) \sum_{l=1}^{T} \exp(x_{il}\beta)} \\
&= \frac{\exp(x_{iT}\beta)}{\sum_{l=1}^{T} \exp(x_{il}\beta)}.
\end{aligned}$$

Le terme d'intercept α_i disparaît. Ainsi la log-vraisemblance conditionnelle de ce modèle est donnée par l'expression :

$$l(\beta) = \sum_{i=1}^{N} \left[\sum_{t=1}^{T} y_{it} x_{it}\beta - \log\left(\sum_{l=1}^{T} \exp(x_{il}\beta) \right) \right].$$

Dans le cadre de la grande dimension, il a été développé une version pénalisée du modèle (4.1) avec une pénalisation de type lasso (Avalos *et al.*, 2012; Simpson *et al.*, 2013) :

$$l_\lambda(\beta) = \sum_{i=1}^{N} \left[\sum_{t=1}^{T} y_{it} x_{it}\beta - \log\left(\sum_{l=1}^{T} \exp(x_{il}\beta) \right) \right] - \lambda|\beta|_1. \tag{4.2}$$

On note $\widehat{\beta}_\lambda = \arg\max_\beta(l_\lambda(\beta))$, avec $\widehat{\beta}_\lambda \in \mathbb{R}^P$. Ainsi, tout comme pour le lasso logistique présenté en section 2.3.1, la pénalité de norme L_1 sur les coefficients de régression permet

de réduire à exactement zéro certains coefficients du vecteur $\widehat{\beta}_\lambda$ et ainsi, de procéder à une sélection de variables. Le problème lié à la sélection du paramètre de pénalité λ se pose également dans le cadre de la régression logistique conditionnelle pénalisée.

Ici, nous avons considéré deux critères différents pour le choix de λ dans (4.2) pour mettre en œuvre des méthodes de détection de signaux : le BIC et l'AIC. La procédure utilisée est la même que celle présentée en section 2.3.4 : des régressions logistiques conditionnelles non pénalisées sont implémentées à partir des variables sélectionnées par le lasso pour différentes valeurs de λ, le BIC et l'AIC étant ensuite calculés pour chacun de ces modèles. Les signaux générés sont les variables qui ont un coefficient strictement positif dans les régressions logistiques conditionnelles multiples non pénalisées minimisant le BIC et l'AIC respectivement. Dans la suite, on fera référence à ces approches de détection sous les appellations cco-bic et cco-aic.

Nous avons également mis en œuvre une approche de détection univariée, qui consiste à implémenter des régressions logistiques conditionnelles univariées où la réponse d'intérêt est régressée par rapport à chaque exposition séparément. Cette approche naïve nous servira dans la suite comme d'un niveau de référence en termes de performances quand aucun moyen de prise en compte de la confusion n'est mis en œuvre. On appellera cette approche cco-univ. On considère comme signaux les variables qui sont associées de manière délétère avec la réponse d'intérêt et qui ont une p-valeur corrigée pour la multiplicité des tests inférieure à 5% (Benjamini et Yekuteli, 2001).

Les régressions logistiques conditionnelles pénalisées ont été implémentées à l'aide de la version 3.0.0 du package R `Cyclops` (Suchard *et al.*, 2013), et les régressions logistiques conditionnelles non pénalisées ont été implémentées à l'aide de la version 3.1-8 package R `survival`.

Paramétrages

Nous avons testé au total huit paramétrages différents dans l'étape de mise en forme des données pour appliquer les approches autour du CCO. Nous avons fait varier :

- le nombre de périodes témoins : 1 ou 4 périodes
- la durée de la période de *washout* entre les périodes d'étude : 15 ou 30 jours

- la durée des périodes d'étude : 30, 60 ou 90 jours.

Toutes les configurations liées à ces différences de paramétrage ne sont pas testées. Les paramétrages que nous avons considérés sont présentés dans le tableau 4.1.

Paramétrage	Nombre de périodes témoins	Durée des périodes de *washout* (jours)	Durée des périodes à risque (jours)
1	1	30	30
2	4	30	30
3	1	15	30
4	4	15	30
5	1	30	60
6	4	30	60
7	1	30	90
8	4	30	90

TABLEAU 4.1 – Paramétrages testés dans la mise en œuvre du *case-crossover* pour la détection de signaux dans l'EGB.

Pour chaque paramétrage, on ne considère que les médicaments qui ont été prescrits chez au moins dix individus sur l'ensemble des périodes d'étude. En fonction du nombre de périodes d'étude considérées et de leur longueur, le nombre de variables médicaments varie.

4.2.2 *Sequence Symmetry Analysis* : définition et mise en œuvre

Méthodes

La SSA a été introduite par PETRI *et al.* (1988) puis formalisée par HALLAS (1996). Cette méthode d'analyse de données longitudinales repose sur la comparaison du nombre de séquences « évènement A puis évènement B » versus « évènement B puis évènement A ». Historiquement, A et B étaient des consommation de médicaments, avec A : médicament suspect et B : médicament pour soigner l'événement indésirable que l'on suspecte être induit par A. Cette approche a été étendue au cas où B est un évènement indésirable par la suite (TSIROPOULOS *et al.*, 2009). Dans tous les cas, les évènements A et B sont des évènement incidents.

Si on considère que A est la prise du médicament suspect j pour $j \in \{1, ..., P\}$ et B la survenue de notre EI d'intérêt Y, la SSA consiste à dénombrer le nombre de séquences « médicament puis EI (noté $X_j \to Y$) » et « EI puis médicament (noté $Y \to X_j$) » sur

une fenêtre d'observation donnée. Pour chaque individu, on considère sur cette fenêtre d'observation uniquement la première occurrence des évènements « prise de médicament » et « survenue de l'EI ». Soit U le nombre de jours de la fenêtre d'observation considérée. Pour cette section uniquement, la matrice $\mathbf{X}$ sera de dimension $(N \times U) \times P$: on compte par individu autant de lignes que de jours dans la période d'observation. De même, la réponse d'intérêt est ici définie par $\mathbf{y} \in \mathbb{R}^{N \times U}$. Ainsi, on a $x_{iuj} = 1$ si l'individu i a pris le médicament j le jour u, et $y_{iv} = 1$ si ce même individu a vécu l'EI d'intérêt le jour v, pour $u, v \in \{1, ..., U\}$. Dans un premier temps, la SSA consiste à calculer le *crude sequence ratio* défini par :

$$r_c = \frac{|X_j \to Y|}{|Y \to X_j|} = \frac{\sum_{i=1}^{N} \left(\sum_{u=1}^{U-1} x_{iuj} \sum_{v=u+1}^{U} y_{iv} \right)}{\sum_{i=1}^{N} \left(\sum_{u=1}^{U-1} y_{iu} \sum_{v=u+1}^{U} x_{ivj} \right)}.$$

Afin de prendre en compte les tendances de prescriptions et d'hospitalisations dont les prévalences varient dans le temps, on calcule le *null sequence ratio*, qui peut être interprété comme une valeur de référence du *crude sequence ratio*. En notant $\mathbf{x}_{uj}$ le vecteur de taille N qui est l'indicatrice de présence du médicament j le jour u, et $\mathbf{y}_u$ le vecteur de taille N qui est l'indicatrice de survenu de la réponse le jour u, on définit :

$$a = \frac{\sum_{u=1}^{U-1} \mathbf{x}_{uj} \sum_{v=u+1}^{U} \mathbf{y}_u}{\sum_{u=1}^{U} \mathbf{x}_{uj} \sum_{u=1}^{U} \mathbf{y}_u}.$$

Le *null sequence ratio* est défini par :

$$r_n = \frac{a}{1-a}.$$

La mesure d'intérêt avec l'approche SSA est alors l'*adjusted sequence ratio* défini comme le rapport

$$r_a = \frac{r_c}{r_n}.$$

La figure 4.2 présente un exemple schématique de la SSA avec le calcul de r_c, r_n et r_a. Enfin, l'intervalle de confiance bilatéral de l'*adjusted sequence ratio* est déterminé, et si la borne inférieure de cet intervalle est supérieur à 1, alors on considère le couple (médicament

j, EI) comme un signal. La SSA est ensuite mise en œuvre pour tous les médicaments enregistrés et sur plusieurs fenêtres d'observation.

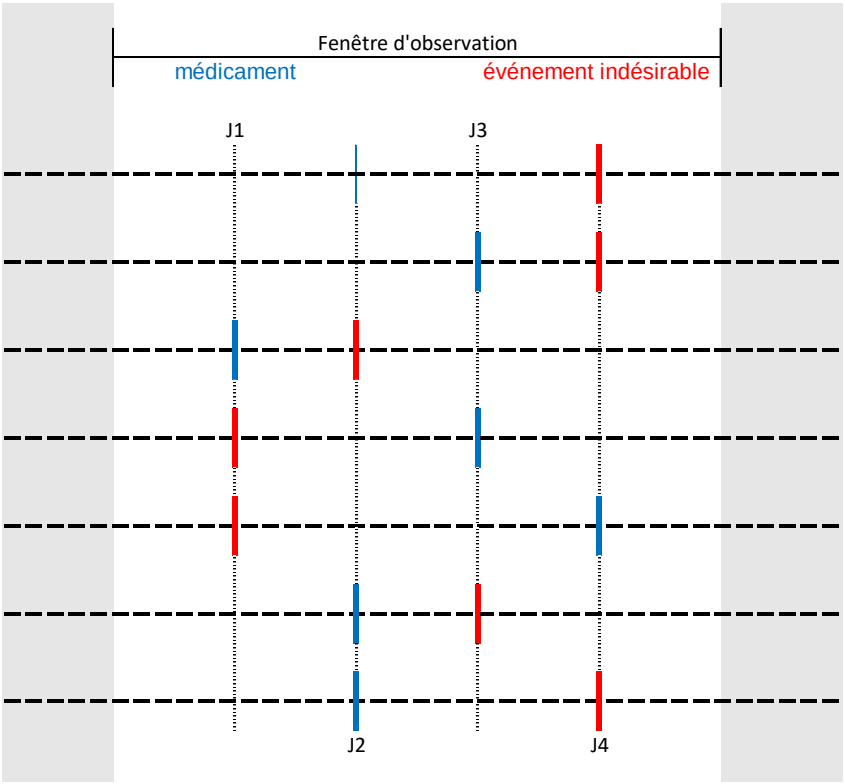


FIGURE 4.2 – Présentation schématique de la *sequence symmetry analysis*. Dans cette exemple, on a $r_c = 5/2 = 2{,}50$, $a = [(1 \times 5) + (3 \times 4) + (2 \times 3)]/(7 \times 7)$ =0,47, r_n =0,89 et r_a =2,81.

Pour déterminer l'intervalle de confiance de l'*adjusted sequence ratio*, nous avons implémenté une méthode dite exacte, qui est celle évoquée dans le travail de HALLAS (1996) sur la base des travaux de MORRIS et GARDNER (1988). Dans son travail, HALLAS (1996) affirme que la variance du *null sequence ratio* est faible devant celle du *crude sequence ratio*, et donc que l'intervalle de confiance de l'*adjusted sequence ratio* est quasiment entièrement déterminé par l'intervalle de confiance du *crude sequence ratio*. Afin de déterminer l'intervalle de confiance du *crude sequence ratio*, on calcule l'intervalle de confiance à $(1-\alpha)\%$ de la probabilité de succès d'une loi binomiale, où le succès est l'évènement « $X_j \to Y$ », et le nombre d'expériences est le nombre total de séquences. L'intervalle de confiance de cette probabilité est calculé à partir de la loi binomiale. On en déduit l'intervalle de confiance à $(1-\alpha)\%$ du *crude sequence ratio*, puis celui de l'*adjusted sequence ratio* en divisant les deux bornes de l'intervalle de confiance du *crude sequence ratio* par la valeur du *null sequence ratio*.

Nous avons également déterminé l'intervalle de confiance de l'*adjusted sequence ratio* par *bootstrap*, comme cela est fait dans la littérature (Wahab *et al.*, 2013; Arnaud *et al.*, 2018). Nous avons obtenu par *bootstrap* une distribution de $B = 200$ valeurs de l'*adjusted sequence ratio*, ce qui nous permet de déterminer l'intervalle de confiance à $(1 - \alpha)\%$ de cette quantité en prenant les percentiles d'ordre $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$ de la distribution.

Pour ces deux intervalles de confiance, nous avons fixé $\alpha = 0{,}05$. Un signal est généré avec la méthode de l'intervalle de confiance exact si, sur au moins une fenêtre d'observation, la borne inférieure de l'intervalle de confiance de l'*adjusted sequence ratio* obtenu avec cette méthode est strictement supérieure à un. La même règle de décision est appliquée pour la méthode de l'intervalle de confiance obtenu par *bootstrap*.

Paramétrages

Wahab *et al.* (2013) ont observé des différences en termes de sensibilité de la SSA pour la détection de signaux en fonction de la durée de la fenêtre d'observation. Comme dans leur travail, nous avons implémenté la SSA en testant plusieurs durées pour observer les séquences : 3, 6 et 12 mois. De plus nous avons considéré deux façons différentes de définir deux fenêtres d'observation consécutives :

- soit de manière totalement disjointe : la seconde fenêtre d'observation commence quand la première se termine,
- soit à la façon de fenêtres glissantes sur un mois : la seconde fenêtre d'observation commence un mois après le début de la première.

Tous ces paramétrages sont résumés dans le tableau 4.2.

Paramétrage	Durée de la fenêtre d'observation (mois)	Superposition des fenêtres d'observation
m3-no	3	non
m3-o	3	oui
m6-no	6	non
m6-o	6	oui
m12-no	12	non
m12-o	12	oui

Tableau 4.2 – Paramétrages testés dans la mise en œuvre de la *sequence symmetry analysis* pour la détection de signaux dans l'EGB

Tout comme dans le travail d'Arnaud *et al.* (2018), nous avons considéré une exposition médicamenteuse dans une fenêtre d'observation donnée s'il y a au moins (i) trois séquences dans la fenêtre qui impliquent ce médicament et l'EI d'intérêt, (ii) une séquence « médicament puis EI » dans la fenêtre. S'il n'y a pas de séquence « EI puis médicament » dans la fenêtre étudiée, le *crude sequence ratio* est fixé à 0.49.

4.2.3 Extraction et mise en forme des données de l'Échantillon Généraliste des Bénéficiaires

Période d'étude et définition des évènements indésirables dans l'EGB

Nous avons travaillé sur les données de l'EGB du 1er janvier 2009 au 31 décembre 2016. La survenue d'un évènement indésirable a été établie à partir des données d'hospitalisations du PMSI, en regardant les diagnostics principaux et reliés [3]. On considère un individu comme exposé à un médicament à partir de la date de délivrance du-dit médicament. Les médicaments sont codés selon la classification ATC et les diagnostics d'hospitalisation selon la Classification Internationale des Maladies, 10ème révision (CIM-10).

Nous avons déterminé les cas de DILI à partir des codes PT du dictionnaire MedDRA qui nous ont été fournis par l'équipe Médicament et santé des populations de l'unité Inserm 1219. Afin de traduire ces codes PT en codes CIM-10, nous avons eu recours à l'*Unified Medical Language System Terminology Service* : un thésaurus qui permet de traduire en différentes classifications un même concept médical. Ainsi, nous avons considéré qu'un individu atteint de DILI est un individu hospitalisé avec un diagnostic principal ou un diagnostic relié indiquant au moins l'un des codes CIM-10 suivants : K71 à K76, Z94.4, G93.7, R16.0, R16.2, R17.

Évènements indésirables et consommations incidentes

Pour les deux familles de méthodes présentées, nous avons considéré des évènements indésirables incidents. La survenue d'un EI est considérée comme incidente s'il n'y a eu aucune hospitalisation pour un cas de DILI dans les 12 mois précédents, après avoir vérifié

3. Le diagnostic principal est « le motif qui a mobilisé l'essentiel de l'effort médical et soignant au cours de l'hospitalisation dans l'unité médicale » et le diagnostic relié est « renseigné lorsque le diagnostic principal est insuffisant [...], il répond à la question : pour quelle maladie, ou état, la prise en charge mentionnée en diagnostic principal a-t-elle été faite ? ». Site du Département de l'Information Médicale

que les données relatives à l'individu étaient bien présentes dans l'EGB l'année précédente. Les cas de DILI que nous avons considérés sont donc survenus entre 2010 et 2016.

Pour la méthode SSA, nous avons considéré uniquement des expositions incidentes. Nous avons ainsi appliqué les mêmes critères que pour les hospitalisations : nous nous sommes assurée de la présence des individus dans l'EGB et avons vérifié qu'aucune délivrance pour le même médicament n'a été effectuée dans les 12 mois précédents.

Filtre des indications grâce à la base de données SIDER

Afin de limiter le biais d'indication dans notre analyse, nous avons eu recours à la base de données SIDER (Kuhn *et al.*, 2016). Le biais d'indication se produit lorsqu'un médicament est prescrit pour une indication associée à l'EI d'intérêt. La base de données SIDER contient les informations relatives aux indications et aux effets indésirables des médicaments. Toutes ces informations sont obtenues à partir d'un travail intensif de *text mining* des RCP des médicaments fournies par la FDA. Les évènements indésirables sont renseignés à l'aide de la terminologie MedDRA et les médicaments sont codés avec la classification ATC. Dans certaines études, SIDER a été utilisée comme ensemble de référence (Wang *et al.*, 2018). Ici, nous avons utilisé cette base de données pour filtrer les médicaments indiqués dans le traitement de DILI avant toute analyse.

4.2.4 Description des données

Nous avons extrait 1 637 hospitalisations incidentes pour des cas de DILI sur la période 2010-2016. Ces 1 637 hospitalisations concernent 1 484 individus : 153 individus ont été hospitalisés deux fois sur la période, mais leurs séjours à l'hôpital étaient suffisamment espacés dans le temps pour que leur deuxième hospitalisation soit considérée comme incidente selon notre définition. Pour ces individus, nous n'avons considéré que la première hospitalisation pour un cas DILI.

Sur la période 2009-2016, les 1 484 cas de DILI ont été à l'origine de 606 302 délivrances de médicaments pour un total de 1 207 médicaments différents. Parmi ces médicaments, 94 étaient des témoins négatifs selon DILIrank et 89 étaient des témoins positifs. Au total, 3,77% et 7,18% des délivrances concernaient des délivrances de témoins négatifs et de témoins positifs respectivement.

4.2.5 Résultats de la détection avec les approches basées sur le *case-crossover*

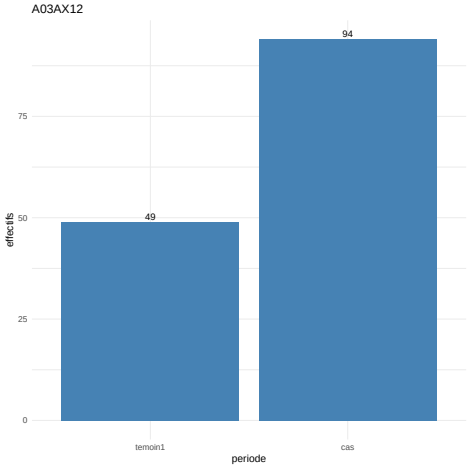
On présente uniquement les résultats obtenus avec le paramétrage 3, défini par une seule période témoin, une durée des périodes d'observation de 30 jours et une période de *whashout* de 15 jours. Les résultats obtenus pour les autres paramétrages sont présentés en annexe C. Le nombre d'expositions médicamenteuses considérées avec le paramétrage 3 était de 211, une fois le filtre des médicaments indiqués pour une DILI (perte de 68 médicaments) et celui des médicaments peu consommés (perte de 454 médicaments) appliqués. 12 de ces 211 médicaments étaient classés comme témoins positifs et 20 comme témoins négatifs selon DILIrank. Le tableau 4.3 présente les résultats de la détection de signaux avec les approches présentées en section 4.2.1.

Approches	Nombre de signaux	Nombre de signaux à statut connu	Vrais positifs			Faux positifs		
			n	PPV (%)	Sensibilité (%)	n	FDP (%)	Spécificité (%)
cco-univ	1	0	0	NA	0,00	0	NA	100,00
cco-aic	30	5	1	20,00	8,33	4	80,00	80,00
cco-bic	4	1	1	100,00	8,33	0	0,00	100,00

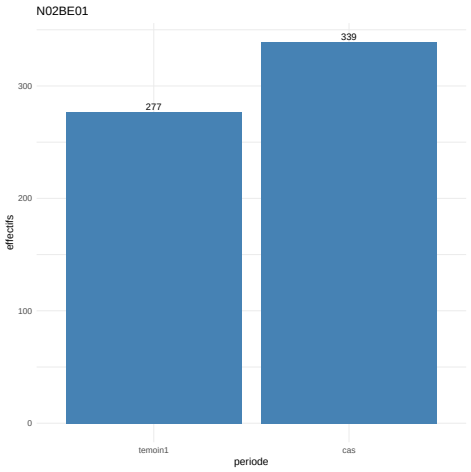
Tableau 4.3 – Résultats de la détection de signaux pour les approches basées sur le *case-crossover* pour le paramétrage 3. 211 expositions médicamenteuses différentes étaient considérées pour ce paramétrage, dont 12 étaient des témoins positifs et 20 des témoins négatifs selon l'ensemble de référence DILIrank.
PPV : Valeur Prédictive Positive
FDP : Proportion de Fausses Découvertes

Les approches cco-univ et cco-bic ont généré respectivement un et quatre signaux. Le signal généré par cco-univ n'avait pas de statut connu dans DILIrank. Le seul signal généré par cco-bic qui avait un statut connu était un témoin positif. L'approche cco-aic a généré un nombre beaucoup plus important de signaux : 30, dont quatre faux positifs et un vrai positif (FDP = 80 %). Le signal généré par cco-univ l'a été par cco-bic également, et l'ensemble des signaux générés par cco-bic est inclus dans celui des signaux générés par cco-aic. Le graphique 4.3 représente les effectifs de délivrances sur la période cas et la période témoin (a) du phloroglucinol qui a été généré par cco-univ, cco-bic et cco-aic et qui a un statut inconnu vis-à-vis de DILI ; (b) du paracétamol, généré par cco-bic et cco-aic et qui est un vrai positif. On constate que l'écart le plus important entre le nombre de consommations en période cas et le nombre de consommations en période témoin est

atteint pour le phloroglucinol.



(a) Statut inconnu : Phloroglucinol, généré par cco-univ, cco-bic et cco-aic.



(b) Vrai positif : Paracétamol, généré par cco-bic et cco-aic.

FIGURE 4.3 – Nombre de délivrances du phlroglucinol (A03AX12) et du paracétamol (N02BE01) sur les différentes périodes d'observation du paramétrage 3 du *case-crossover*. Selon DILIrank, ces médicaments ont respectivement un statut inconnu et de témoin positif.

Les résultats obtenus en termes de détection pour les paramétrages 1, 2 et 4 à 8, sont présentés dans le tableau C.1 en annexe. Pour chaque paramétrage il est précisé le nombre de médicaments considérés induit par la mise en forme des données ainsi que le nombre de témoins négatifs et de témoins positifs présents parmi ces médicaments. Plus la durée des périodes d'étude et le nombre de périodes témoins augmentent, plus le nombre de signaux générés par les méthodes augmente. Le nombre de témoins négatifs selon DILIrank parmi les différents médicaments considérés étaient toujours supérieur au nombre de témoins positifs.

Dans tous les paramétrages, cco-aic est l'approche qui a généré le plus de signaux : entre 31 pour le paramétrage 1 et 78 pour le paramétrage 8. Le nombre de périodes

témoins considérées a influé fortement sur le nombre de signaux généré par cco-univ. Pour les paramétrages avec une seule période témoin, cco-univ a généré entre zéro et trois signaux, alors qu'elle a généré entre 17 et 60 signaux quand il y avait quatre périodes témoins. L'approche cco-bic a généré un nombre de signaux faible pour les paramétrages 1 et 3, 22 signaux pour les paramétrages 6 et 8, et environ 10 signaux pour les autres paramétrages.

Dans l'ensemble, les performances en termes de fausses découvertes pour les trois méthodes et pour les paramétrages 1, 2 et 4 à 8 étaient décevantes. Dès que plusieurs signaux à statut connu étaient générés, au moins un faux positif était détecté. De manière générale, les signaux contenaient plus de faux positifs que de vrais positifs. Ainsi, les valeurs de FDP associées à cco-univ, cco-aic et cco-bic étaient supérieures à 60% dans quasiment tous les paramétrages, et leur sensibilité était généralement entre 15% et 25%.

4.2.6 Résultats de la détection avec les approches basées sur la *sequence symmetry analysis*

On présente uniquement les résultats obtenus avec le paramétrage appelé m3-o dans le tableau 4.2 : les fenêtres d'observations ont une durée de trois mois et sont définies comme des fenêtres glissantes sur un mois. Les résultats des autres paramétrages sont présentés en annexe dans les tableaux C.2 et C.3. Pour le paramétrage m3-o, le nombre de fenêtres d'observations considérées était de 82. En moyenne, il y avait 14 expositions médicamenteuses différentes considérées par fenêtre d'observation. Sur toute la période d'étude, toutes fenêtres d'observation confondues, il y avaient en moyenne 4 séquences à considérer par exposition médicamenteuse. Au total, 117 médicaments différents furent considérés sur toutes ces périodes, dont deux étaient des témoins positifs et 10 des témoins négatifs selon DILIrank.

L'approche par intervalle de confiance exact n'a généré aucun signal. Seule l'approche basée sur des intervalles de confiance construits par *bootstrap* a détecté un vrai positif sur les deux identifiables parmi ses 17 signaux générés.

Le tableau C.2 en annexe présente pour les autres paramétrages que m3-o le nombre de fenêtres d'observation considérées, le nombre moyen de médicaments par fenêtre et le

Approches	Nombre de signaux	Nombre de signaux à statut connu	Vrais positifs			Faux positifs		
			n	PPV (%)	Sensibilité (%)	n	FDP (%)	Spécificité (%)
IC exact	0	0	0	NA	0,00	0	NA	100,00
IC *bootstrap*	17	1	1	100,00	50,00	0	0,00	100,00

TABLEAU 4.4 – Résultats de la détection de signaux pour les approches basées sur la *sequence symmetry analysis* pour le paramétrage m3-o. 117 expositions médicamenteuses différentes étaient considérées pour ce paramétrage, dont 10 étaient des témoins négatifs et 2 des témoins positifs selon l'ensemble de référence DILIrank.
PPV : Valeur Prédictive Positive
FDP : Proportion de Fausses Découvertes

nombre moyen de séquences par médicament toutes fenêtres d'observations confondues. Il présente aussi le nombre de médicaments différents considérés au total par paramétrage, ainsi que les nombres de vrais et faux positifs parmi ces médicaments. Plus la durée des fenêtres d'observation augmente, plus le nombre de médicaments différents augmente, ainsi que le nombre de médicaments considérés par fenêtre et le nombre de séquences par médicament. Pour tous les paramétrages, le nombre de témoins positifs selon DILIrank parmi les médicaments considérés était inférieur au nombre de témoins négatifs.

Les résultats obtenus en termes de détection pour les paramétrages présentés dans le tableau 4.2 sont présentés dans le tableau C.3. L'approche basée sur l'intervalle de confiance exact de l'*adjusted sequence ratio* n'a jamais généré de signal. L'approche basée sur l'intervalle de confiance calculé par *bootstrap* a généré un plus grand nombre de signaux quand les fenêtres d'observation étaient définies à la façon de fenêtres glissantes. Dès que cette approche a généré des signaux à statut connu selon DILIrank, le nombre de vrais et le nombre de faux positifs détectés étaient sensiblement équivalents. Ainsi, cette approche a montré des FDP supérieures à 60% et des valeurs de sensibilité autour de 30%.

4.2.7 Discussion

Nous avons mis en œuvre des méthodes de détection de signaux basées respectivement sur le *case-crossover* et la *sequence symmetry analysis*. Nous avons fait varier un certain nombre de paramètres dans la mise en œuvre de nos approches. Un effort important de mise en forme des données de l'EGB a été fait, avec en particulier le filtrage des indications liées à l'évènement d'intérêt à partir de la base experte SIDER.

Pour la détection basée sur le *case-crossover*, on constate que le nombre de périodes

témoins influence le nombre de signaux générés. L'approche basée sur l'AIC a généré plus de signaux que celle basée sur le BIC. Ce résultat n'est pas surprenant. L'AIC est un critère adapté à des fins de prédiction, et a ainsi tendance à sélectionner des modèles moins parcimonieux que le BIC. L'approche univariée a généré un nombre de signaux intermédiaire. Les trois méthodes de détection ont montré dans l'ensemble des performances assez décevantes en termes de fausses découvertes. Dans la majorité des paramétrages, toutes les approches mises en œuvre ont détecté plus de faux positifs que de vrais positifs, ce qui s'est particulièrement vérifié pour les paramétrages où les périodes d'observation étaient de 90 jours (paramétrage 7 et 8).

En regardant plus en détail les résultats obtenus avec le paramétrage 3 du *case-crossover*, on peut tenter d'émettre quelques hypothèses sur les signaux générés. Le signal généré par cco-univ, cco-aic et cco-bic correspond au phloroglucinol, un antispasmodique utilisé, entre autres, dans la prise en charge des coliques hépatiques. Son statut d'association avec DILI n'est pas renseigné dans DILIrank, mais au vu de son indication on peut raisonnablement penser qu'il s'agit d'un biais protopathique : le médicament est initié à cause de symptômes précurseurs de la maladie. Un des faux positifs, détecté par cco-aic est l'hydroxyzine. Il s'agit d'un anxiolytique qui peut être utilisé dans la prémédication aux anesthésies générales. Ainsi, en supposant qu'une partie des séjours hospitaliers à partir desquels nous avons défini une DILI aient donné lieu à un acte de chirurgie, on peut comprendre la forte consommation de ce médicament avant la survenue de l'EI.

Pour la détection basée sur la *sequence symmetry analysis*, on constate également que le paramétrage de cette méthode influence fortement les résultats. Comme le nombre de fenêtres d'observation est plus important quand on décide de les définir de manière glissante, on observe mécaniquement un plus grand nombre de signaux générés par les approches dans ces situations. Des fenêtres d'observation plus étendues permettent d'augmenter le nombre de séquences observées et donc d'augmenter aussi potentiellement le nombre de signaux générés. La méthode de détection basée sur les intervalles de confiance obtenus par *bootstrap* est celle qui a généré le plus de signaux. L'approche basée sur les intervalles de confiance exacts a généré un seul signal pour le paramétrage m12-o. Dans l'ensemble, les méthodes basées sur la SSA ont montré des performances décevantes pour la détection, en détectant autant de vrais positifs que de faux positifs.

Les performances de toutes ces approches ont été évaluées à partir des signaux à statut connus de l'ensemble de référence DILIrank. Très peu des signaux générés par ces méthodes avaient un statut connu. Ces faibles effectifs étaient dus aux restrictions importantes de l'ensemble de référence DILIrank induites par la mise en forme des données. En particulier, le nombre de témoins positifs présents parmi les médicaments considérés était bien inférieur au nombre de témoins négatifs.

Nous avons choisi ici d'implémenter des méthodes de détection basées sur le *case-crossover* et non sur la série de cas comme cela a été initialement fait dans l'étude de l'OMOP publiée en 2013. Une des hypothèses de validité de la série de cas consiste à considérer que la probabilité d'être exposé à un médicament ne doit pas être modifiée par la survenue d'un évènement indésirable (Whitaker *et al.*, 2009). Cette hypothèse nous a semblé particulièrement forte, ce qui nous a motivée à envisager le *case-crossover*, une autre méthode auto-contrôlée qui ne s'intéresse qu'à la période de temps antérieure à la survenue de la première occurrence d'un EI. Plusieurs études nous ont confortée dans ce choix. Lors de la mise en œuvre de la série de cas, elles considéraient uniquement la première occurrence de l'évènement d'intérêt (Zhou *et al.*, 2018) ou ont trouvé que le meilleur paramétrage de cette méthode était celui qui ne considère que la première survenue de l'EI d'intérêt (Thurin *et al.*, 2020). Tout comme la série de cas, le *case-crossover* permet de gérer la confusion qui ne varie pas au cours du temps. En revanche, bien que certaines précautions puissent être prises dans le paramétrage de cette méthode, le *case-crossover* est par conception sensible à la confusion dépendante du temps comme par exemple les tendances de prescription. Comme la période cas est toujours postérieure à la période contrôle, la méthode sera positivement biaisée si la fréquence globale de prescription augmente avec le temps. Afin de remédier à cela, Suissa (1995) a développé le design du *case-time-control* qui considère des témoins appariés au cas du *case-crossover* pour mesurer l'effet de tendance dans ce groupe.

Les résultats obtenus à l'issue des différentes détections sont difficiles à interpréter. Bien que l'évènement DILI soit fréquemment notifié dans les données de notifications, et malgré la taille de l'EGB, les effectifs des cas incidents de DILI sont faibles. Cela s'explique en partie par le fait que l'on considère ici la survenue d'EI graves puisque requérant une hospitalisation. Ces faibles effectifs de cas implique un manque de puissance qui pourrait

expliquer le nombre relativement faible de signaux générés par certaines approches. Ces faibles effectifs impliquent également une restriction de l'ensemble de signaux de référence considéré. Ainsi, peu de signaux générés par une méthode sont pris en compte dans l'évaluation de ses performances. De plus, on observe un déséquilibre entre le nombre de témoins positifs et négatifs présents, en faveur de ces derniers. Il est alors difficile de statuer sur les performances en termes de sensibilité des méthodes testées. En plus de ces considérations liées à un manque de puissance, on constate que certaines fausses découvertes peuvent s'expliquer par leur indication, et ce malgré un filtre des médicaments indiqués dans le traitement de l'évènement d'intérêt opéré à l'aide de la base de données SIDER.

Initialement, nous avions songé à mettre en œuvre une méthode de détection de signaux basée sur l'exploitation conjointe des détections dans l'EGB d'une part, et dans la BNPV d'autre part. Notre idée était de combiner les p-valeurs obtenues avec l'approche cco-univ dans l'EGB avec celle obtenues avec les approches basées sur le score de propension dans la BNPV, d'après la méthode proposée par Li *et al.* (2015). Au vu des résultats obtenus dans ce travail, cette stratégie ne nous a pas semblé appropriée. Nous avons alors envisagé une autre stratégie dans l'exploitation de l'EGB pour la détection de signaux en pharmacovigilance : exploiter l'information de l'exposition médicamenteuse d'une population témoin issue de l'EGB complétée par celle des cas de la base des notifications française.

4.3 Exploitation conjointe des données de l'Échantillon Généraliste des Bénéficiaires et des notifications spontanées

Malgré une taille importante, l'EGB ne fournit pas une population de cas de notre évènement d'intérêt suffisante. En revanche, l'EGB dispose de nombreux individus témoins qui n'ont pas vécu l'évènement. D'un autre côté, on ne dispose pas dans la BNPV, par construction de la base de données, d'individus témoins représentatifs de la population générale. Pour toutes les approches de détection présentées dans les chapitres 2 et 3, les

expositions médicamenteuses des cas d'un EI donné sont comparées à celles d'un groupe témoin constitué par l'ensemble des individus notifiés pour un autre EI que celui d'intérêt. Dans cette optique, nous avons cherché à créer une base dite hybride, composée des cas de la BNPV pour un EI d'intérêt et d'individus témoins de l'EGB.

Dans un premier temps, nous avons comparé les performances de deux méthodes de détection de signaux sur la base hybride ainsi créée et sur la base des notifications spontanées. Dans un second temps, nous avons cherché à intégrer l'information sur l'exposition médicamenteuse des témoins portée par cette base hybride pour améliorer la détection de signaux dans la BNPV. Cette intégration est rendue possible par l'utilisation du lasso pondéré (Bergersen *et al.*, 2011), qui est l'appellation donnée au lasso adaptatif présenté en section 3.2 quand les poids propres à chaque variable sont dérivés d'une source de données annexe. Cette méthode d'intégration de données est comparée à une détection basée sur le lasso dans la base des notifications spontanées seule.

4.3.1 Méthodes

Dans un premier temps, on effectue dans la base hybride une détection de signaux pour DILI à l'aide de la méthode de disproportionnalité RFET (Ahmed *et al.*, 2010). Cette approche est mise en œuvre de la même manière que précédemment (voir section 3.5.2) : RFET est implémenté pour les médicaments qui ont au moins trois notifications en commun avec DILI. Les p-valeurs des variables filtrées à cette étape sont fixées à un et nous appliquons la correction de tests multiples proposée par Benjamini et Yekuteli (2001). Le seuil de détection est fixé à 5%. En plus de RFET, nous avons mis en œuvre la méthode de détection lasso-bic (voir section 2.3.4). Ces deux méthodes sont implémentées dans la base hybride d'une part, et dans la BNPV d'autre part.

Dans un second temps, nous définissions à partir des détections mises en œuvre dans la base hybride, des pénalités propres à chaque variable à intégrer dans un lasso pondéré. On utilisera le mot poids dans la suite pour désigner ces coefficients multiplicateurs de λ.

A partir de la détection opérée avec RFET sur la base hybride, nous définissons deux poids dit « naïfs » qui seront par la suite intégrés dans un lasso pondéré implémenté dans la BNPV. Pour ces deux pondérations, les variables non significativement associées à la réponse dans la base hybride sont exclues des variables candidates à la sélection dans le

lasso pondéré. Dans la suite on note $\mathcal{D} \subset \{1, ..., P\}$ l'ensemble des variables qui ont une p-valeur corrigée significative.

Poids 1 : Coefficients univariés Les premiers poids définis reposent sur les coefficients univariés déduits de la valeur de l'odds ratio entre la réponse d'intérêt et chaque variable médicament. Pour une variable X_j, son poids associé est :

$$w_j^{(1)} = \begin{cases} \dfrac{1}{|\log(\mathrm{OR}_j)|} & \text{si} \quad j \in \mathcal{D}, \\ \infty & \text{sinon.} \end{cases}$$

Ainsi, plus une variable est associée fortement à la réponse, moins elle est pénalisée.

Poids 2 : P-valeurs En considérant les p-valeurs corrigées des variables de $\mathcal{D}$, nous proposons d'attribuer à ces variables un poids dans le lasso pondéré déterminé à partir de la fonction de répartition empirique de ces p-valeurs. Pour une variable X_j on définit ainsi :

$$w_j^{(2)} = \begin{cases} \dfrac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \mathbb{1}_{\text{p-valeur}_d \leq \text{p-valeur}_j} & \text{si} \quad j \in \mathcal{D}, \\ \infty & \text{sinon.} \end{cases}$$

Avec cette transformation monotone, les variables avec les p-valeurs les plus faibles ont des poids plus faibles et sont donc favorisées dans la sélection de variables opérée par le lasso pondéré.

Au vu des résultats du chapitre 3, nous avons également mis en œuvre les poids définis dans la section 3.3. A partir du lasso-bic et de CISL implémentés dans la base hybride, nous définissons les poids w^{lb} (voir équation (3.1)) et w^{cisl} (voir équation (3.3)). Tous ces poids sont par la suite intégrés dans un lasso pondéré implémenté sur la BNPV. Pour toutes ces approches, on considère les signaux générés pour différentes valeurs de pénalité dans le lasso pondéré. Sont considérés comme signaux les variables qui ont un coefficient estimé strictement positif. Plus la valeur de la pénalité diminue, plus le nombre de variables sélectionnées par les lasso pondérés, et donc potentiellement le nombre de signaux générés par ces approches, augmente.

4.3.2 Constitution de la base hybride

Comme nous cherchons à extraire de l'EGB uniquement les données de consommations de médicaments, nous avons pu travailler sur la période 2006-2016. Dans un premier temps, nous avons réalisé une extraction de la BNPV sur cette période selon les mêmes filtres que ceux présentés en section 2.5.3. Nous avons considéré les notifications pour lesquelles l'âge et le sexe étaient renseignées. Nous avons ensuite déterminé les cas de DILI présents dans la base de données selon les critères présentés en section 2.5.4. Au total, cette extraction de la BNPV comprenait 14 509 cas de DILI sur 252 118 notifications.

Dans un second temps, nous avons apparié les cas de la BNPV à des témoins de l'EGB sur le sexe et par classe d'âge de 10 ans. Les cas et les témoins étaient appariés au moment de la survenue de l'EI des cas. Sont considérés comme témoins les individus qui n'ont pas été hospitalisés pour un quelconque problème hépatique l'année de leur appariement et les années antérieures. Ainsi nous avons retiré tous les individus qui ont été hospitalisés avec un diagnostic principal, relié ou associé [4] dont le code CIM-10 commençait par K7 [5]. Nous avons cherché à apparier jusqu'à 100 témoins par cas, tout en interdisant le fait qu'un témoin soit apparié avec plusieurs cas. Au final, nous obtenons 525 506 témoins uniques de l'EGB appariés aux cas de DILI.

Les consommations médicamenteuses des témoins ont été regardées dans le mois précédent la date de survenue de l'EI du cas auquel ils sont appariés. Les témoins étaient considérés comme exposés à un médicament s'ils avaient au moins une délivrance de ce médicament dans le mois. En se restreignant aux expositions médicamenteuses communes aux cas de la BNPV et aux témoins de l'EGB et qui sont présents plus de 10 fois dans la base hybride, nous avons considéré 1 023 variables médicaments. Sur ces 1 023 médicaments, 101 étaient des témoins positifs et 96 étaient des témoins négatifs selon DILIrank. Au total, 714 médicaments étaient notifiés au moins trois fois avec un cas de DILI, dont 88 étaient des témoins positifs et 68 étaient des témoins négatifs.

4. Le diagnostic associé concerne l'ensemble des « morbidités (maladies, symptômes et autres motifs de recours aux soins) associées au diagnostic principal et ayant donné lieu à une prise en charge effective diagnostic ou thérapeutique au cours du séjour dans l'unité médicale ». Site du Département de l'Information Médicale

5. Les codes CIM-10 K70 à K77 correspondent aux maladies du foie

4.3.3 Résultats

Dans un premier temps, nous comparons les détections obtenues sur la base hybride et sur l'extraction de la BNPV considérée. Le tableau 4.5 présente le nombre de signaux générés, le nombre de vrais et faux positifs détectés et les performances en termes de FDP, PPV, sensibilité et spécificité des méthodes RFET et lasso-bic, mises en œuvre sur les deux bases de données. Sur la base hybride, la procédure RFET a généré 336 signaux dont 97 avaient un statut connu selon DILIrank. 74 de ces signaux étaient des vrais positifs, menant à une sensibilité de la méthode à 73%, et 23 de ces signaux étaient des faux positifs ce qui a conduit à une FDP de 23%. La méthode lasso-bic a généré 217 signaux dont 74 avaient un statut connu dans DILIrank : 64 étaient des vrais positifs et 10 des faux négatifs, conduisant respectivement à une sensibilité de 63% et une FDP de 13%.

RFET et lasso-bic ont généré significativement moins de signaux lorsqu'elles ont été implémentées sur la BNPV avec respectivement 145 et 99 signaux. Parmi les 54 signaux à statut connu générés par RFET, 50 étaient des vrais positifs et 4 des faux positifs. Ainsi cette approche a montré une sensibilité de 49% et une FDP de 7%. Lasso-bic n'a détecté aucun faux positif sur cette extraction de la BNPV. Elle a détecté 45 vrais positifs, et a ainsi montré une sensibilité à 44%.

Approches	Nombre de signaux	Nombre de signaux à statut connu	Vrais positifs			Faux positifs		
			n	PPV (%)	Sensibilité (%)	n	FDP (%)	Spécificité (%)
Base Hybride								
RFET	336	97	74	76,29	73,27	23	23,71	76,04
lasso-bic	217	74	64	86,49	63,37	10	13,51	89,58
BNPV								
RFET	145	54	50	92,59	49,50	4	7,41	95,83
lasso-bic	99	45	45	100,00	44,55	0	0,00	100,00

Tableau 4.5 – Résultats de la détection de signaux avec les méthodes RFET et lasso-bic, dans la base hybride et dans la Base Nationale de Pharmacovigilance (BNPV).
PPV : Valeur Prédictive Positive
FDP : Proportion de Fausses Découvertes

Dans un second temps, nous comparons, à nombre de signaux générés égaux, la répartition des signaux générés par les méthodes de détection basée sur le lasso pondéré présentées en section 4.3.1 avec un lasso classique implémenté dans la BNPV. La figure 4.4 représente la répartition des signaux, des vrais positifs, et des faux positifs (en colonne)

générés par les approches basées sur la lasso pondéré avec les poids $w^{(1)}, w^{(2)}, w^{lb}, w^{cisl}$ (en ligne) et un lasso classique.

En regardant la répartition des signaux générés par le lasso d'une part, et les différents lasso pondérés d'autre part (première colonne de la figure 4.4), on constate que les poids w^{lb} ont conduit à des signaux générés très différents (graphique (c.1)). Cela était également le cas pour l'approche basée sur les poids w^{cisl} pour un nombre de signaux générés supérieur à 70 (graphique (d.1)). Les approches basées sur les poids naïfs $w^{(1)}$ et $w^{(2)}$ ont généré un plus grand nombre de signaux en commun avec le lasso classique (graphique (a.1) et (b.1)).

Parmi les signaux générés par les deux approches basées sur les poids naïfs et non générés par le lasso classique, peu étaient de vrais positifs. La méthode qui repose sur les poids $w^{(2)}$ a montré de moins bonnes performances avec un nombre de vrais positifs détectés toujours plus faibles que le lasso classique, et un faux positif détecté à partir de 38 signaux générés (figure (b.3)). La méthode de détection qui repose sur les poids w^{lb} a détecté jusqu'à huit vrais positifs de plus que le lasso classique pour un même nombre de signaux générés (figure (c.2)). Elle a également détecté jusqu'à trois faux positifs de plus que le lasso classique (figure (c.3)). A partir de 127 signaux générés, la méthode basée sur les poids w^{cisl} a détecté plus de vrais positifs que le lasso classique (figure (d.2)). En revanche, dès 70 signaux générés cette approche a détecté deux faux positifs (figure (d.3)). A partir de 150 signaux générés, cette approche a détecté autant de faux positifs que le lasso classique.

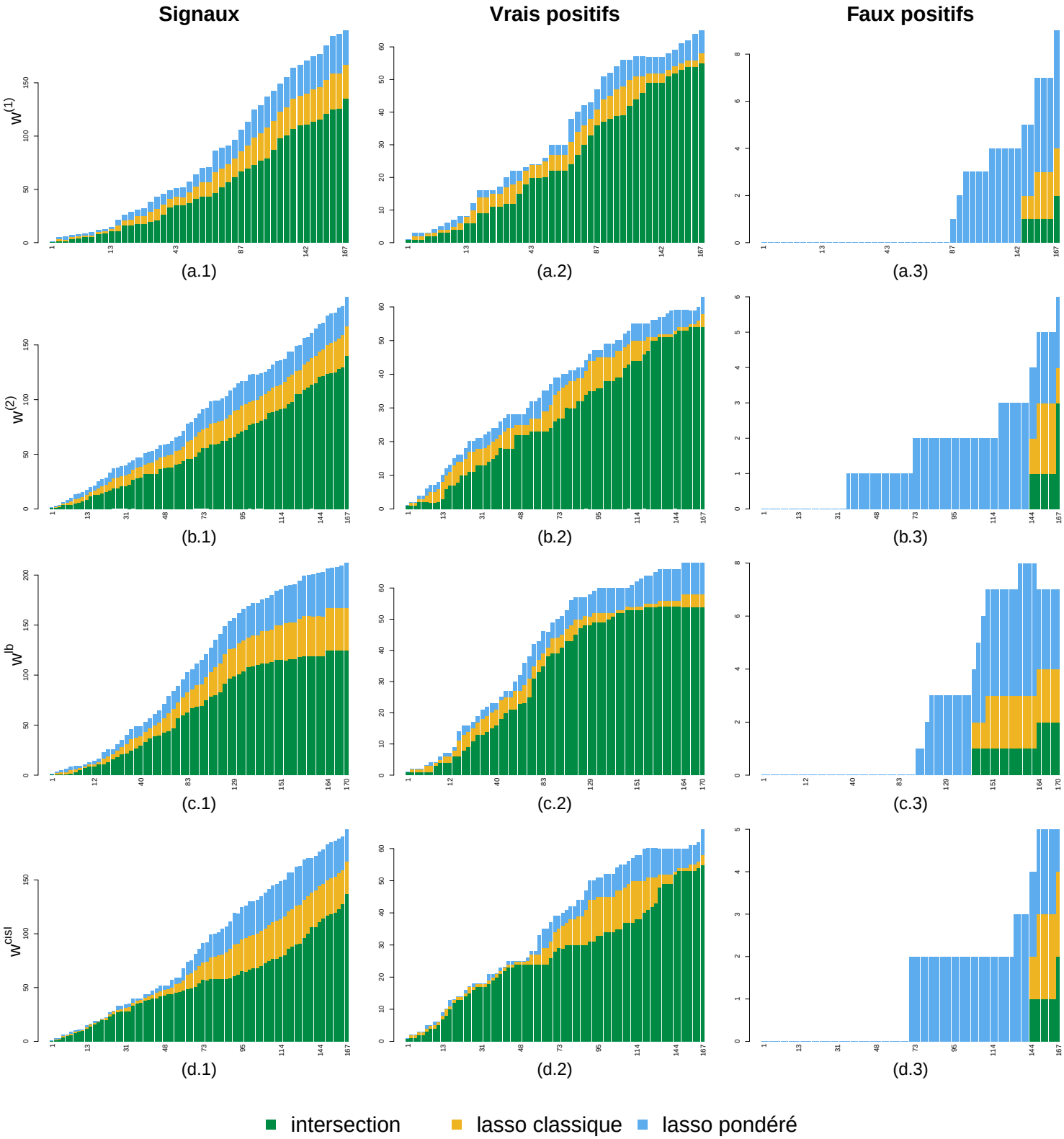


FIGURE 4.4 – Comparaison des signaux générés et des vrais/faux positifs détectés (en colonnes) entre le lasso classique et les lasso pondérés avec poids obtenus dans la base hybride (en ligne). Les effectifs en vert, jaune et bleu sont les effectifs des signaux générés par le lasso classique et le lasso pondéré, le lasso classique uniquement et le lasso pondéré uniquement, respectivement. Par exemple (graphe (d.2)), pour 114 signaux générés, 38 vrais positifs étaient détectés par le lasso classique et le lasso pondéré avec les poids w^{cisl}, 12 étaient détectés par le lasso classique seul et 8 par la lasso pondéré seul.

4.3.4 Discussion

Le travail présenté ici est une première tentative pour exploiter conjointement les données de la BNPV et de l'EGB. Dans la BNPV, des cas d'un EI donné sont comparés à des cas d'autres évènements indésirables. Partant de ce constat, nous avons cherché à construire une population de témoins ayant des caractéristiques proches de celles des cas et avons apparié des individus de la BNPV avec des individus de l'EGB. Cette démarche reste assez empirique.

En comparant les performances de méthodes mises en œvre sur cette base hybride et la BNPV, on constate que les méthodes génèrent le double de signaux sur la base hybride. Les performances des méthodes sur cette base se sont avérées mauvaises en termes de fausses découvertes. De part la grande différence dans le recueil des données entre la BNPV et l'EGB, la base hybride ainsi créée nous est apparue comme peu fiable en elle même pour la détection.

Nous avons alors cherché à intégrer l'information que pouvait recéler cette base via un lasso pondéré mis en œuvre dans la BNPV « originale ». Ainsi nous avons défini deux poids dits naïfs, qui reposent sur une détection dans la base hybride avec une méthode de disproportionnalité. Par la suite, nous avons décidé d'inclure dans la comparaison les poids qui présentaient de bons résultats dans la chapitre 3 dérivés d'un lasso-bic et de CISL. Chacun des lasso pondérés ont été comparés à une régression lasso classique. Afin de regarder plus globalement les détections des différentes approches, nous nous sommes affranchie du choix de λ dans toutes les régressions pénalisées.

Les méthodes de détection développées ici reposant sur le lasso pondéré ont, dans l'ensemble, des performances moins bonnes en termes de fausses découvertes par rapport à une méthode basée sur une régression lasso classique. Néanmoins elles ont permis de détecter certains vrais positifs qui n'étaient pas détectés avec le lasso classique. En particulier, l'approche qui repose sur des poids obtenus à partir d'un lasso-bic dans la base hybride a même détecté un nombre plus important de vrais positifs que le lasso classique à nombre de signaux générés équivalent. Pour un nombre de signaux générés élevé, la méthode de détection basée sur les poids dérivés de CISL a détecté plus de vrais positifs et autant de faux positifs qu'un lasso classique. Néanmoins pour un nombre de signaux générés plus faible, ses performances étaient moins satisfaisantes. Les deux approches basées sur les

poids naïfs ont amené les résultats les moins satisfaisants avec peu de vraies découvertes comparées au lasso classique. En particulier, le lasso pondéré avec les poids basés sur les p-valeurs corrigées a détecté des faux positifs pour un nombre faible de signaux générés (38 signaux).

4.4 Conclusion

La taille relativement faible de l'EGB ne permet pas de disposer d'un effectif suffisant de cas pour la détection de signaux concernant l'évènement indésirable DILI. La taille réduite de l'ensemble de référence couplée au déséquilibre du nombre de témoins négatifs et positifs rendent difficile l'évaluation des performances de nos méthodes de détection. Ainsi les approches qui nous avaient semblé pertinentes à mettre en œuvre au moment de cette étude ont donné des résultats décevants. Nos ambitions autour du développement de nouvelles méthodes de détection basées sur l'exploitation conjointe des détections réalisées dans la BNPV d'une part, et l'EGB d'autre part se sont vues contrariées.

Nous avons alors tenté d'exploiter conjointement les données de ces deux bases de données. Au prix de quelques fausses découvertes, l'information apportée par la constitution d'un groupe de témoins issu de l'EGB dans la détection sur la BNPV a permis de détecter des vrais positifs qui n'étaient pas détectés par un lasso classique. Cela a été particulièrement le cas avec la méthode basée sur les poids w^{lb}, qui a fourni des résultats satisfaisants avec des signaux générés très différents d'un lasso classique. Ces résultats ouvrent des perspectives intéressantes et mériteraient d'être approfondis avec d'autres études empiriques.

4.5 Valorisation

Le travail présenté en section 4.3 a été valorisé au travers d'un congrès national : les 51$^{\text{èmes}}$ Journées de Statistique (juin 2019) de la Société Française de Statistique. Le résumé long de cette communication est en annexe D.3. Il a aussi donné lieu à une communication affichée à un congrès international : la 40$^{\text{ème}}$ conférence annuelle de l'*International Society for Clinical Biostatistics* (juillet 2019).

Chapitre 5

Conclusion générale et perspectives

5.1 Conclusion générale

Au cours de ce travail de thèse, nous avons développé des méthodes de détection de signaux en pharmacovigilance. Les chapitres 2 et 3 de ce manuscrit explorent deux champs méthodologiques différents pour la détection sur des données de notifications spontanées. Tous deux proposent de considérer la forme individuelle des données afin de prendre en compte les coprescriptions. Le chapitre 4 explore des stratégies de détection qui exploite des données médico-administratives, celles de l'Échantillon Généraliste des Bénéficiaires.

Dans le chapitre 2, nous avons considéré la méthode du score de propension en grande dimension qui est une extension du score de propension, un outil issu du domaine de la causalité et utilisée largement en épidémiologie et pharmacoépidémiologie. Notre travail a permis de mettre en avant l'intérêt d'une stratégie de pondération sur le PS en grande dimension pour la pharmacovigilance : la pondération avec poids *matching weights*. Bien que conputationnellement intensive, cette stratégie permet de se reposer sur la théorie des tests classiques pour le seuil de détection.

Dans le chapitre 3, nous avons considéré une variation de la régression pénalisée lasso, le lasso adaptatif, qui présente de meilleures propriétés en termes de sélection de variables que le lasso. Nous avons construit de nouveaux poids adaptatifs dans le cadre de la grande dimension et proposé de se reposer sur le critère BIC pour le choix de λ. Nous avons mis en œuvre une vaste étude de simulations afin d'évaluer les performances de nos propositions et de celles présentées au chapitre précédent dans différentes situations. La méthode ba-

sée sur le score de propension en grande dimension qui avait notre préférence s'est avérée très conservatrice dès que la réponse d'intérêt était rare. Les méthodes proposées basées sur la lasso adaptatif montrent globalement de meilleurs résultats en termes de fausses découvertes et de sensibilité que les méthodes concurrentes. Une étude empirique analogue à celle du chapitre 2 vient compléter l'évaluation et confirmer ces résultats. Toutes les méthodes présentées dans les chapitres 2 et 3 sont implémentées dans le package R `adapt4pv` disponible sur le CRAN. Bien que mises en œuvre dans le contexte de la pharmacovigilance, nos propositions autour du lasso adaptatif en grande dimension peuvent être utilisées pour d'autres applications.

Dans le chapitre 4 de ce manuscrit nous avons tenté de tirer parti de la base de l'EGB et des notifications spontanées. Dans un premier temps, une détection de signaux basée sur l'EGB seule ne s'est pas avérée concluante. Un manque de puissance à imputer à la taille relativement faible de l'EGB ne nous a pas permis d'évaluer correctement les performances des méthodes testées. Dans un second temps nous avons considéré une détection conduite sur les notifications spontanées basée sur un lasso adaptatif intégrant, au travers de ses poids, l'information relative à l'exposition médicamenteuse d'individus contrôles mesurée dans l'EGB. Une des méthodes basée sur des poids analogues à ceux définis au chapitre 3 a montré dans l'ensemble des performances comparables, voire meilleure en termes de vraies découvertes, à une méthode basée sur une régression lasso classique. Les signaux générés par cette méthode sont différents, ce qui laisse à penser que l'information introduite dans la détection a permis de fournir une détection complémentaire à une détection plus reposant uniquement sur les notifications spontanées.

5.2 Perspectives

Plusieurs développements aux travaux présentés ici sont envisageables. L'étude de simulations réalisée dans le troisième chapitre a mis en évidence que pour certaines des méthodes proposées reposant sur le score de propension en grande dimension, une correction de tests multiples ne permettait pas d'assurer un contrôle sur le *False Discovery Rate*. Ces résultats surprenants constituent une première piste d'investigation envisagée pour la suite de nos travaux.

Une autre perspective autour des méthodes de détection basées sur le PS en grande dimension présentée en section 2.7 est de compléter l'évaluation des différentes méthodes d'estimation via le calcul de mesures d'équilibre. D'autres travaux peuvent être menés autour de la meilleure prise en compte de cette notion d'équilibre induit par les scores de propension. Imai et Ratkovic (2014) ont proposé une méthode d'estimation du score qui repose sur l'équilibre qu'il peut induire chez les variables observées : le *covariate balancing propensity score*. A notre connaissance cette approche n'a pas été mise en œuvre dans le cadre de la grande dimension, ni dans celui de la pharmacovigilance. Cette méthode serait un compétiteur intéressant pour les autres approches d'estimation du score que nous avons testées ici. Récemment, Li *et al.* (2018) ont proposé une nouvelle pondération sur le score de propension, les *overlap weights*, et ont montré que ces poids possèdent de bonnes propriétés en termes d'équilibre induits sur les variables observées dans la pseudo population qui découlent de cette pondération. Dans un autre travail, Li *et al.* (2019) ont également mis en évidence les bonnes performances de cette pondération comparées à celles de l'IPTW pour l'estimation de l'effet du traitement sur la réponse d'intérêt. Ils définissent dans ce même travail une forme explicite de l'estimateur de la variance de l'effet traitement avec cette pondération.

Tout au long de cette thèse, nous avons eu recours au critère BIC pour choisir la pénalité λ pour le lasso. Nous avons choisi de calculer le BIC à partir des modèles de régression classiques, constitués des variables sélectionnées par le lasso, pour différentes valeurs de pénalité. Les méthodes de détection de signaux basées sur ce critère nous ont donné satisfaction. Une extension de ce travail pourrait être de comparer les performances de cette méthode à une approche qui repose sur le calcul du BIC obtenu à partir de la vraisemblance pénalisée. Une possibilité serait même de considérer une forme de combinaison de ces deux approches. Hastie *et al.* (2009) ont proposé une version simplifiée du *relaxed lasso* initialement introduit par Meinshausen (2007) en définissant l'estimateur $\widehat{\beta}^{\text{relax}}$ comme une combinaison linéaire de l'estimateur obtenu avec le lasso pour un λ_k donné et celui obtenu par maximisation de la vraisemblance du modèle non pénalisé comprenant les variables actives associées à λ_k. Ainsi, en faisant varier le paramètre de cette combinaison linéaire on pourrait imaginer comparer les différences dans les sélections de variables obtenues avec ces BIC « différents ». Enfin, on pourrait étendre ce travail par

une comparaison avec d'autres variations autour du BIC qui ont été proposées dans le cadre de la régression lasso comme l'*extended BIC* (Chen et Chen, 2008), le *modified BIC* (Wang *et al.*, 2009) ou encore le *high-dimensional BIC* (Wang et Zhu, 2011). Pour ce qui est du cadre spécifique de l'adaptive lasso, on pourra également considérer le critère de l'*Extended Regularized Information Criterion* proposé par Hui *et al.* (2015).

Une publication que nous avons trouvée tardivement au cours de ce travail de thèse a également attiré notre attention. Il s'agit des travaux de Shortreed et Ertefaie (2017) autour de l'*outcome-adaptive lasso*. Cette méthode permet de « fusionner » les champs méthodologiques abordés dans les chapitres 2 et 3 de cette thèse. Ainsi, Shortreed et Ertefaie (2017) proposent de construire un score de propension à partir d'un lasso adaptatif. Les poids adaptatifs sont déterminés à partir des coefficients de régression obtenus par maximum de vraisemblance dans la régression de la réponse d'intérêt par rapport à l'ensemble des variables observées et du traitement considéré. L'objectif de cette méthode est de favoriser dans la sélection des variables à inclure dans le modèle d'estimation du PS les variables qui sont associées à la réponse d'intérêt, comme cela est recommandé dans la littérature (Brookhart *et al.*, 2006). Le paramètre λ est choisi selon une mesure d'équilibre, la *weighted Average Standard Mean Difference*, faisant intervenir la pondération IPTW. Mettre en œuvre cette méthode avec les enseignements tirés de notre travail serait intéressant et potentiellement prometteur. Ainsi, la pondération IPTW résultant en des poids instables dans notre situation, pourrait être remplacée par les *matching weights* ou les *overlap weights*. Enfin, plutôt que des poids adaptatifs déterminés à partir des estimations obtenues par maximum de vraisemblance, on pourra utiliser les poids proposés dans le chapitre 3.

Le travail présenté dans le chapitre 4 de ce manuscrit à mis en évidence la difficulté de l'utilisation des bases médico-administratives pour la pharmacovigilance dans certaines situations. Pour l'évènement indésirable que nous avons considéré, on peut incriminer la relative faible taille des données de l'EGB. Plus généralement, exploiter ces bases de données reste une tâche ardue, à cause des biais auxquelles elles sont soumises. En particulier, une analyse préliminaire des signaux générés sur l'EGB laisse à penser que ces signaux sont dus à des bais d'indications ou des biais protopathiques. Ni les méthodes mises en

œuvre, pourtant pertinentes d'après la littérature, ni le filtre des indications avec la base de données SIDER n'ont *a priori* permis de pallier entièrement ces biais. Pour se prononcer sur les réelles performances des méthodes implémentées, il faudrait évaluer les signaux qui n'étaient pas présents dans l'ensemble de référence considéré. Ainsi, que ce soit pour identifier au mieux les biais à l'œuvre ou pour évaluer pleinement les performances des méthodes, il est de notre point de vue primordial d'impliquer une certaine expertise dans la détection sur les bases médico-administratives. Par exemple, on pourrait envisager de solliciter des experts pharmacologues pour évaluer les signaux générés, ou effectuer une recherche dans la littérature, éventuellement de manière automatisée à l'aide d'algorithme de *text mining* comme cela a été proposé par Avillach *et al.* (2013).

A l'heure actuelle, aucun système de surveillance ne repose sur des bases médico-administratives pour la pharmacovigilance en routine. Au delà des difficultés liées aux aspects méthodologiques dans l'exploitation des données, on peut également pointer le manque de réactivité dans la mise à jour et dans la mise à disposition de ces bases. Bien que les notifications spontanées soient également soumises à certains biais (Weber, 1984; Bégaud *et al.*, 2002), elles restent la pierre angulaire pour la détection de signaux en pharmacovigilance. La réactivité des systèmes de surveillance ainsi que l'expertise médicale impliquée dans la construction des bases de données sont autant d'arguments en faveur des notifications spontanées. Ainsi, c'est plus la complémentarité des notifications spontanées et des BMA qui, à notre avis, est à viser (du moins pour le moment). En effet, les bases médico-administratives permettent d'étudier des évènements indésirables ou des populations qui ne peuvent être étudiés dans les notifications spontanées (Demailly *et al.*, 2020). Là aussi, pour définir précisément ces évènements ou ces populations ciblées, il est nécessaire d'avoir recours à une certaine expertise médicale. Pour le moment, fournir une manière pertinente d'exploiter conjointement les données de notifications spontanées et les bases médico-administratives dans toutes les situations est difficile. Des propositions ont été faites pour interconnecter ces deux sources de données (Harpaz *et al.*, 2013; Li *et al.*, 2015; Pacurariu *et al.*, 2015) et il est certain que cette question va continuer à susciter de l'intérêt et donner à lieu à de nombreuses recherches.

Bibliographie

Abadie, A. et G. W. Imbens. 2016, «Matching on the Estimated Propensity Score», *Econometrica*, vol. 84, n° 2, p. 781–807. 22, 44

Ahmed, I., B. Bégaud et P. Tubert-Bitter. 2015, «chapter 13 : evaluation of post-marketing safety using spontaneous reporting databases», dans *Statistical Methods for EvaluatingSafety in Medecine Product Development*, édité par A. L. Goud, Wiley, p. 332–344. 4

Ahmed, I., C. Dalmasso, F. Haramburu, F. Thiessard, P. Broët et P. Tubert-Bitter. 2010, «False discovery rate estimation for frequentist pharmacovigilance signal detection methods», *Biometrics*, vol. 66, n° 1, p. 301–309. 5, 27, 59, 97

Ahmed, I., F. Haramburu, A. Fourrier-Réglat, F. Thiessard, C. Kreft-Jais, G. Miremont-Salamé, B. Bégaud et P. Tubert-Bitter. 2009, «Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting», *Statistics in Medicine*, vol. 28, n° 13, p. 1774–1792. 5

Ahmed, I., A. Pariente et P. Tubert-Bitter. 2018, «Class-imbalanced subsampling lasso algorithm for discovering adverse drug reactions», *Statistical Methods in Medical Research*, vol. 27, n° 3, p. 785–797. 5, 7, 14, 18, 19

Akaike, H. 1973, «Information theory and an extension of maximum likelihood principle», dans *Proc. 2nd Int. Symp. on Information Theory*, p. 267–281. 6

Ali, M. S., R. H. H. Groenwold, W. R. Pestman, S. V. Belitser, K. C. B. Roes, A. W. Hoes, A. de Boer et O. H. Klungel. 2014, «Propensity score balance measures in pharmacoepidemiology : a simulation study», *Pharmacoepidemiology and Drug Safety*, vol. 23, n° 8, p. 802–811. 44

Almenoff, J., J. M. Tonning, A. L. Gould, A. Szarfman, M. Hauben, R. Ouellet-Hellstrom, R. Ball, K. Hornbuckle, L. Walsh, C. Yee, S. T. Sacks, N. Yuen, V. Patadia, M. Blum, M. Johnston, C. Gerrits, H. Seifert et K. LaCroix. 2005, «Perspectives on the use of data mining in pharmacovigilance», *Drug Safety*, vol. 28, n° 11, p. 981–1007. 5

Arnaud, M., B. Bégaud, F. Thiessard, Q. Jarrion, J. Bezin, A. Pariente et F. Salvo. 2018, «An Automated System Combining Safety Signal Detection and Prioritization from Healthcare Databases : A Pilot Study», *Drug Safety*, vol. 41, n° 4, p. 377–387. 87, 88

Arnaud, M., B. Bégaud, N. Thurin, N. Moore, A. Pariente et F. Salvo. 2017, «Methods for safety signal detection in healthcare databases : a literature review», *Expert opinion on drug safety*, vol. 16, n° 6, p. 721–732. 80

Arnaud, M., F. Salvo, I. Ahmed, P. Robinson, N. Moore, B. Bégaud, P. Tubert-Bitter et A. Pariente. 2016, «A Method for the Minimization of Competition Bias in Signal Detection from Spontaneous Reporting Databases», *Drug Safety*, vol. 39, n° 3, p. 251–260. 5

Austin, P. C. 2009, «Some methods of propensity-score matching had superior performance to others : results of an empirical investigation and monte carlo simulations», *Biometrical Journal : Journal of Mathematical Methods in Biosciences*, vol. 51, n° 1, p. 171–184. 22

Austin, P. C. 2011, «An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies», *Multivariate Beahavioral Research*, vol. 46, n° 3, p. 399–424. 21, 44, 56

Austin, P. C. et E. A. Stuart. 2015, «Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies», *Statistics in Medicine*, vol. 34, n° 28, p. 3661–3679. 22

Avalos, M., Y. Grandvalet, D. Adroher, L. Orriols et E. Lagarde. 2012,

«Analysis of multiple exposures in the case-crossover design via sparse conditional likelihood», *Statistics in Medicine*, vol. 31, n° 21, p. 2290–2302. 80, 82

Avillach, P., J. C. Dufour, G. Diallo, F. Salvo, M. Joubert, F. Thiessard, F. Mougin, G. Trifirò, A. Fourrier-Réglat, A. Pariente et M. Fieschi. 2013, «Design and validation of an automated method to detect known adverse drug reactions in MEDLINE : A contribution from the EU-ADR project», *Journal of the American Medical Informatics Association*, vol. 20, n° 3, p. 446–452. 109

Ayers, K. L. et H. J. Cordell. 2010, «SNP Selection in Genome-Wide and Candidate Gene Studies via Penalized Logistic Regression», *Genetic Epidemiology*, vol. 34, n° 8, p. 879–891. 50, 55

Bate, A., M. Lindquist, I. R. Edwards, S. Olsson, R. Orre, A. Lansner et R. M. De Freitas. 1998, «A Bayesian neural network method for adverse drug reaction signal generation», *European Journal of Clinical Pharmacology*, vol. 54, n° 4, p. 315–321. 5

Bégaud, B., K. Martin, F. Haramburu et N. Moore. 2002, «Rates of spontaneous reporting of adverse drug reactions in France», *Journal of the American Medical Association*, vol. 288, n° 13, p. 1588. 8, 77, 109

Benjamini, Y. et Y. Hochberg. 1995, «Controlling the false discovery rate a practical and powerful approach to multiple testing», *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 57, n° 1, p. 289–300. 5

Benjamini, Y. et D. Yekuteli. 2001, «The Control of the False Discovery Rate in Multiple Testing under Dependency», *The Annals of Statistics*, vol. 29, n° 4, p. 1165–1188. 56, 83, 97

Bergersen, L. C., I. K. Glad et H. Lyng. 2011, «Weighted lasso with data integration», *Statistical Applications in Genetics and Molecular Biology*, vol. 10, n° 1. 97

Brookhart, M. A., S. Schneeweiss, K. J. Rothman, R. J. Glynn, J. Avorn et T. Stürmer. 2006, «Variable selection for propensity score models», *American Journal of Epidemiology*, vol. 163, n° 12, p. 1149–1156. 21, 43, 108

Bross, I. D. 1966, «Spurious effects from an extraneous variable», *Journal of Chronic Diseases*, vol. 19, n° 6, p. 637–647. 23

Bühlmann, P. et S. van de Geer. 2011, «Lasso for linear models», dans *Statistics for High-Dimensional Data*, Springer Science & Business Media, p. 7–42. 50, 52, 54, 74

Caruana, E., S. Chevret, M. Resche-Rigon et R. Pirracchio. 2015, «A new weighted balance measure helped to select the variables to be included in a propensity score model», *Journal of Clinical Epidemiology*, vol. 68, n° 2, p. 1415–1422. 44

Caster, O., K. Juhlin, S. Watson et G. N. Norén. 2014, «Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank : Retrospective evaluation against emerging safety signals», *Drug Safety*, vol. 37, n° 8, p. 617–628. 5

Caster, O., G. N. Norén, D. Madigan et A. Bate. 2010, «Large-scale regression-based pattern discovery : The example of screening the WHO global drug safety database», *Statistical Analysis and Data Mining*, vol. 3, n° 4, p. 197–208. 5, 6, 7, 14, 17, 19

Caster, O., L. Sandberg, T. Bergvall, S. Watson et G. N. Norén. 2017, «vigiRank for statistical signal detection in pharmacovigilance : First results from prospective real-world use», *Pharmacoepidemiology and Drug Safety*, vol. 26, n° 8, p. 1006–1010. 5

Chan, C. L., S. Rudrappa, P. S. Ang et S. C. Li. 2017, «Detecting signals of disproportionate reporting from Singapore ' s spontaneous adverse event reporting system – an application of the Sequential Probability Ratio Test Detecting signals of disproportionate reporting from Singapore ' s spontaneous adverse», *Drug Safety*, vol. 40, n° 8, p. 703–713. 5

Chen, J. et Z. Chen. 2008, «Extended Bayesian information criteria for model selection with large model spaces», *Biometrika*, vol. 95, n° 3, p. 759–771. 45, 108

Chen, M., A. Suzuki, S. Thakkar, K. Yu, C. Hu et W. Tong. 2016, «DILIrank :

The largest reference drug list ranked by the risk for developing drug-induced liver injury in humans», *Drug Discovery Today*, vol. 21, n° 4, p. 648–653. 11, 15, 31, II

Chen, M., V. Vijay, Q. Shi, Z. Liu, H. Fang et W. Tong. 2011, «FDA-approved drug labeling for the study of drug-induced liver injury», *Drug Discovery Today*, vol. 16, n° 15-16, p. 697–703. 11, 15, 31, II, XV

Choi, N.-K., Y. Chang, Y. K. Choi, S. Hahn et B.-J. Park. 2010, «Determinants of cholesterol and triglycerides recording in patients treated with lipid lowering therapy in UK primary care», *Pharmacoepidemiology and drug safety*, vol. 19, p. 138–146. 79

Choi, N.-K., Y. Chang, J.-Y. Kim, Y.-K. Choi et B.-J. Park. 2011, «Determinants of cholesterol and triglycerides recording in patients treated with lipid lowering therapy in UK primary care», *Pharmacoepidemiology and drug safety*, vol. 20, p. 1278–1286. 79

Cole, S. R. et M. A. Hernan. 2008, «Constructing Inverse Probability Weights for Marginal Structural Models», *American Journal of Epidemiology*, vol. 168, n° 6, p. 656–664. 20, 21

Curtis, J. R., H. Cheng, E. Delzell, D. Fram, M. Kilgore, K. Saag, H. Yun et W. DuMouchel. 2009, «Adaptation of Bayesian Data Mining Algorithms to Longitudinal Claims Data : Coxib Safety as an Example», *Medical Care*, vol. 46, n° 9, p. 969–975. 79

Dalmasso, C., P. Broët et T. Moreau. 2005, «A simple procedure for estimating the false discovery rate», *Bioinformatics*, vol. 21, n° 5, p. 660–668. 27

Delaney, J. A. et S. Suissa. 2009, «The case-crossover study design in pharmacoepidemiology», *Statistical Methods in Medical Research*, vol. 18, n° 1, p. 53–65. 81

Demailly, R., S. Escolano, F. Haramburu, P. Tubert-Bitter et I. Ahmed. 2020, «Identifying Drugs Inducing Prematurity by Mining Claims Data with High-Dimensional Confounder Score Strategies», *Drug Safety*, p. 1–11. 9, 15, 23, 109

Ding, Y., M. Markatou et R. Ball. 2020, «An evaluation of statistical approaches to postmarketing surveillance», *Statistics in Medicine*, vol. 39, n° 7, p. 845–874. 5

DUMOUCHEL, W. 1999, «Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system», *The American Statistician*, vol. 53, n° 3, p. 177–190. 4

EVANS, S. J. W., P. C. WALLER et S. DAVIS. 2001, «Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports», *Pharmacoepidemiology and Drug Safety*, vol. 10, n° 6, p. 483–486. 4

FAN, J. et R. LI. 2001, «Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties», *Journal of the American Statistical Association*, vol. 96, n° 456, p. 1348–1360. 7, 49, 50

FARRINGTON, C. P. 1995, «Relative Incidence Estimation from Case Series for Vaccine Safety Evaluation», *Biometrics*, vol. 51, n° 1, p. 228–235. 8, 79

FRANKLIN, J. M., W. EDDINGS, P. C. AUSTIN, E. A. STUART et S. SCHNEEWEISS. 2017, «Comparing the performance of propensity score methods in healthcare database studies with rare outcomes», *Statistics in Medicine*, vol. 36, n° 12, p. 1946–1963. 15, 23, 27, 56

FRANKLIN, J. M., W. EDDINGS, R. J. GLYNN et S. SCHNEEWEISS. 2015, «Regularized regression versus the high-dimensional propensity score for confounding adjustment in secondary database analyses», *American Journal of Epidemiology*, vol. 182, n° 7, p. 651–659. 15, 23

GENKIN, A., D. D. LEWIS et D. MADIGAN. 2007, «Large-scale bayesian logistic regression for text categorization», *Technometrics*, vol. 49, n° 3, p. 291–304. 6

GROENWOLD, R. H. H., F. DE VRIES, A. DE BOER, W. R. PESTMAN, F. H. RUTTEN, A. W. HOES et O. H. KLUNGEL. 2011, «Balance measures for propensity score methods : a clinical example on beta-agonist use and the risk of myocardial infarction», *Pharmacoepidemiology and Drug Safety*, vol. 20, n° 11, p. 1130–1137. 44

HALLAS, J. 1996, «Evidence of depression provoked by cardiovascular medication : a prescription sequence symmetry analysis», *Epidemiology*, p. 478–484. 79, 84, 86

Hallas, J. et A. Pottegård. 2014, «Use of self-controlled designs in pharmacoepidemiology», *Journal of Internal Medicine*, vol. 275, n° 6, p. 581–589. 80

Harpaz, R., W. Dumouchel, N. H. Shah, D. Madigan, P. Ryan et C. Friedman. 2012, «Novel data-mining methodologies for adverse drug event discovery and analysis», *Clinical Pharmacology and Therapeutics*, vol. 91, n° 6, p. 1010–1021. 5

Harpaz, R., S. Vilar, W. DuMouchel, H. Salmasian, K. Haerian, N. H. Shah, H. S. Chase et C. Friedman. 2013, «Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions», *Journal of the American Medical Informatics Association*, vol. 20, n° 3, p. 413–419. 5, 6, 9, 109

Hastie, T., R. Tibshirani et J. Friedman. 2009, «Boosting and additive trees», dans *The Elements of Statistical Learning*, New York : Springer series in statistics, p. 337–387. 26, 107

Hoerl, A. E. et R. W. Kennard. 2000, «Ridge regression : Biased estimation for nonorthogonal problems», *Technometrics*, vol. 42, n° 1, p. 80–86. 52

Huang, J., S. Ma et C.-H. Zhang. 2008a, «The iterated lasso for high-dimensional logistic regression», *The University of Iowa, Department of Statistics and Actuarial Sciences*. 50, 52, 74

Huang, J., S. Ma et C.-H. Zhang. 2008b, «Adaptive Lasso for Sparse High-Dimensional Regression Models», *Statistica Sinica*, p. 1603–1618. 50, 52, 54, 55

Huang, L., J. Zalkikar et R. C. Tiwari. 2011, «A likelihood ratio test based method for signal detection with application to FDA's drug safety data», *Journal of the American Statistical Association*, vol. 106, n° 496, p. 1230–1241. 5

Hui, F. K., D. I. Warton et S. D. Foster. 2015, «Tuning Parameter Selection for the Adaptive Lasso Using ERIC», *Journal of the American Statistical Association*, vol. 110, n° 509, p. 262–269. 75, 108

Imai, K. et M. Ratkovic. 2014, «Covariate balancing propensity score», *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 76, p. 243–263. 107

Jin, H., J. Chen, H. He, C. Kelman, D. McAullay et C. M. O'Keefe. 2010, «Signaling potential adverse drug reactions from administrative health databases», *IEEE Transactions on knowledge and data engineering*, vol. 22, n° 6, p. 839–853. 79

Jin, H., J. Chen, C. Kelman, H. He, D. McAullay et C. M. O'Keefe. 2006, «Mining unexpected associations for signalling potential adverse drug reactions from administrative health databases», dans *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, p. 867–876. 79

Ju, C., M. Combs, S. D. Lendle, J. M. Franklin, R. Wyss, S. Schneeweiss et M. van der Laan. 2019, «Propensity score prediction for electronic healthcare databases using super learner and high-dimensional propensity score methods», *Journal of Applied Statistics*, vol. 46, n° 12, p. 2216–2236. 15, 24

Kim, J., M. Kim, J.-H. Ha, J. Jang, M. Hwang, B. K. Lee, M. W. Chung, T. M. Yoo et M. J. Kim. 2011, «Signal detection of methylphenidate by comparing a spontaneous reporting database with a claims database», *Regulatory Toxicology and Pharmacology*, vol. 61, n° 2, p. 154–160. 79

Knight, K. et W. Fu. 2000, «Asymptotics for Lasso-type estimators», *Annals of Statistics*, vol. 28, n° 5, p. 1356–1378. 49

Kuhn, M., I. Letunic, L. J. Jensen et P. Bork. 2016, «The SIDER database of drugs and side effects», *Nucleic Acids Research*, vol. 44, n° D1, p. D1075–D1079. 89

Lee, B. K., J. Lessler et E. A. Stuart. 2010, «Improving propensity score weighting using machine learning», *Statistics in Medicine*, vol. 29, n° 3, p. 337–346. 24

Li, F., K. L. Morgan et A. M. Zaslavsky. 2018, «Balancing Covariates via Propensity Score Weighting», *Journal of the American Statistical Association*, vol. 113, n° 521, p. 390–400. 107

Li, F., L. E. Thomas et F. Li. 2019, «Addressing Extreme Propensity Scores via the Overlap Weights», *Practice of Epidemiology*, vol. 188, n° 1, p. 250–257. 43, 107

Li, L. et T. Greene. 2013, «A weighting analogue to pair matching in propensity score analysis», *International Journal of Biostatistics*, vol. 9, n° 2, p. 215–234. 27, 44, 56

Li, Y., P. B. Ryan, Y. Wei et C. Friedman. 2015, «A Method to Combine Signals from Spontaneous Reporting Systems and Observational Healthcare Data to Detect Adverse Drug Reactions», *Drug Safety*, vol. 38, n° 10, p. 895–908. 10, 96, 109

Lunceford, J. K. et M. Davidian. 2004, «Stratification and weighting via the propensity score in estimation of causal treatment effects : A comparative study», *Statistics in Medicine*, vol. 23, n° 19, p. 2937–2960. 22

Maclure, M. 1991, «The Case-Crossover Design : A Method for Studying Transient Effects on the Risk of Acute Events», *American Journal of Epidemiology*, vol. 133, n° 2, p. 144–153. 79, 80

Mallows, C. 1973, «Some comments on cp», *Technometrics*, vol. 15, p. 661–675. 6

Mccaffrey, D. F., G. Ridgeway et A. R. Morral. 2004, «Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies», *Psychological Methods*, vol. 9, n° 4, p. 403. 15, 24

Meinshausen, N. 2007, «Relaxed Lasso», *Computational statistics & data analysis*, vol. 52, n° 1, p. 374–393. 107

Meinshausen, N. et P. Bühlmann. 2006, «High-dimensional graphs and variable selection with the Lasso», *Annals of Statistics*, vol. 34, n° 3, p. 1436–1462. 49

Meinshausen, N. et P. Bühlmann. 2010, «Stability selection», *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 72, n° 4, p. 417–473. 7, 14, 18

Morris, J. A. et M. J. Gardner. 1988, «Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates», *British Medical Journal*, vol. 296, p. 1313–1316. 86

Norén, G. N., A. Bate, R. Orre et I. R. Edwards. 2006, «Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events», *Statistics in Medicine*, vol. 25, n° 21, p. 3740–3757. 5

Norén, G. N., J. Hopstadius, A. Bate, K. Star et I. R. Edwards. 2010, «Temporal pattern discovery in longitudinal electronic patient records», *Data Mining and Knowledge Discovery*, vol. 20, n° 3, p. 361–387. 79

Pacurariu, A. C., S. M. Straus, G. Trifirò, M. J. Schuemie, R. Gini, R. Herings, G. Mazzaglia, G. Picelli, L. Scotti, L. Pedersen, P. Arlett, J. van der Lei, M. C. Sturkenboom et P. M. Coloma. 2015, «Useful Interplay Between Spontaneous ADR Reports and Electronic Healthcare Records in Signal Detection», *Drug Safety*, vol. 38, n° 12, p. 1201–1210. 9, 109

Pariente, A., P. Avillach, F. Salvo, F. Thiessard, G. Miremont-Salamé, A. Fourrier-Reglat et A. F. d. C. R. d. P. (CRPV). 2012, «Effect of Competition Bias in Safety Signal Generation», *Drug Safety*, vol. 35, n° 10, p. 855–864. 5

Pariente, A., F. Gregoire, A. Fourrier-Reglat, F. Haramburu et N. Moore. 2007, «Impact of safety alerts on measures of disproportionality in spontaneous reporting databases the notoriety bias», *Drug Safety*, vol. 30, n° 10, p. 891–898. 77

Petri, H., H. C. W. De Vet, J. Naus et J. Urquhart. 1988, «Prescription sequence analysis : a new and fast method for assessing certain adverse reactions of prescription drugs in large populations», *Statistics in Medicine*, vol. 7, n° 11, p. 1171–1175. 84

Pfeiffer, R. M. et R. Riedl. 2015, «On the use and misuse of scalar scores of confounders in design and analysis of observational studies», *Statistics in Medicine*, vol. 34, n° 18, p. 2618–2635. 22

van Puijenbroek, E. P., A. Bate, H. G. M. Leufkens, M. Lindquist, R. Orre et A. C. G. Egberts. 2002, «A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions.», *Pharmacoepidemiology and drug safety*, vol. 11, n° 1, p. 3–10. 4

Rosenbaum, P. R. et D. B. Rubin. 1983, «The central role of the propensity score in observational studies for causal effects», *Biometrilca*, vol. 70, n° 1, p. 41–55. 20

Rosenbaum, P. R. et D. B. Rubin. 1984, «Reducing bias in observational studies using subclassification on the propensity score», *Journal of the American Statistical Association*, vol. 79, n° 387, p. 516–524. 22

Ryan, P. B. et M. J. Schuemie. 2013, «Evaluating performance of risk identification methods through a large-scale simulation of observational data», *Drug Safety*, vol. 36, n° S1, p. 171–180. 8

Ryan, P. B., M. J. Schuemie, S. Gruber, I. Zorych et D. Madigan. 2013a, «Empirical performance of a new user cohort method : Lessons for developing a risk identification and analysis system», *Drug Safety*, vol. 36, n° S1, p. 59–72. 9

Ryan, P. B., M. J. Schuemie, E. Welebob, J. Duke, S. Valentine et A. G. Hartzema. 2013b, «Defining a Reference Set to Support Methodological Research in Drug Safety», *Drug Safety*, vol. 36, n° S1, p. 33–47. 11

Ryan, P. B., P. E. Stang, J. M. Overhage, M. A. Suchard, A. G. Hartzema, W. DuMouchel, C. G. Reich, M. J. Schuemie et D. Madigan. 2013c, «A comparison of the empirical performance of methods for a risk identification system», *Drug Safety*, vol. 36, n° S1, p. 143–158. 8

Sabourin, J. A., W. Valdar et A. B. Nobel. 2015, «A permutation approach for selecting the penalty parameter in penalized model selection», *Biometrics*, vol. 71, n° 4, p. 1185–1194. 50, 55

Schapire, R. E. 1990, «The Strength of Weak Learnability», *Machine Learning*, vol. 5, n° 2, p. 197–227. 24

Schneeweiss, S. et J. Avorn. 2005, «A review of uses of health care utilization databases for epidemiologic research on therapeutics», *Journal of Clinical Epidemiology*, vol. 58, n° 4, p. 323–337. 78

Schneeweiss, S., W. Eddings, R. J. Glynn, E. Patorno, J. Rassen et J. M. Franklin. 2017, «Variable Selection for Confounding Adjustment in High-dimensional Covariate Spaces When Analyzing Healthcare Databases», *Epidemiology*, vol. 28, n° 2, p. 237–248. 23

Schneeweiss, S., J. A. Rassen, R. J. Glynn, J. Avorn, H. Mogun et M. A. Brookhart. 2009, «High-dimensional propensity score adjustment in studies of treatment effects using health care claims data», *Epidemiology*, vol. 20, n° 4, p. 512. 9, 15, 22, 23

Schuemie, M. J. 2011, «Methods for drug safety signal detection in longitudinal observational databases : LGPS and LEOPARD», *Pharmacoepidemiology and drug safety*, vol. 20, p. 292–299. 79

Schwarz, G. 1978, «Estimating the dimension of a model», *The annals of statistics*, vol. 6, n° 2, p. 461–464. 6

Seeger, J. D., K. Bykov, D. B. Bartels, K. Huybrechts et S. Schneeweiss. 2017, «Propensity score weighting compared to matching in a study of dabigatran and warfarin», *Drug Safety*, vol. 40, n° 2, p. 169–181. 43

Setoguchi, S., S. Schneeweiss, M. A. Brookhart, R. J. Glynn et E. F. Cook. 2010, «Evaluating uses of data mining techniques in propensity score estimation : a simulation study», *Pharmacoepidemiology and Drug Safety*, vol. 17, n° 6, p. 546–555. 24

Shortreed, S. M. et A. Ertefaie. 2017, «Outcome-adaptive lasso : variable selection for causal inference», *Biometrics*, vol. 73, n° 4, p. 1111–1122. 108

Simpson, S. E., D. Madigan, I. Zorych, M. J. Schuemie, P. B. Ryan et M. A. Suchard. 2013, «Multiple self-controlled case series for large-scale longitudinal observational databases», *Biometrics*, vol. 69, n° 4, p. 893–902. 9, 82

Stone, M. 1974, «Cross-Validatory Choice and Assessment of Statistical Predictions», *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 36, n° 2, p. 111–133. 6

Suchard, M. A., S. E. Simpson, I. Zorych, P. Ryan et D. Madigan. 2013, «Massive parallelization of serial inference algorithms for a complex generalized linear model», *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 23, n° 1, p. 1–17. 80, 83

Suissa, S. 1995, «The case-time-control design», *Epidemiology*, vol. 6, n° 3, p. 248–253. 95

Szarfman, A., J. M. Tonning et P. M. Doraiswamy. 2004, «Pharmacovigilance in the 21st century : New systematic tools for an old problem», *Pharmacotherapy*, vol. 24, n° 9, p. 1099–1104. 4

Tatonetti, N. P., P. P. Ye, R. Daneshjou et R. B. Altman. 2012, «Data-Driven Prediction of Drug Effects and Interactions», *Science Translational Medicine*, vol. 4, n° 125. 15, 25

Thiessard, F., E. Roux, G. Miremont-Salamé, A. Fourrier-Réglat, F. Haramburu, P. Tubert-Bitter et B. Bégaud. 2005, «Trends in spontaneous adverse drug reaction reports to the French pharmacovigilance system (1986-2001)», *Drug Safety*, vol. 28, n° 8, p. 731–740. 3

Thurin, N., R. Lassalle, M. Schuemie, M. Pénichon, J. J. Gagne, J. A. Rassen, J. Benichou, A. Weil, P. Blin, N. Moore et C. Droz-Perroteau. 2020, «Empirical assessment of case-based methods for drug safety alert identification in the French National Healthcare System database (SNDS) : Methodology of the ALCAPONE project», *Pharmacoepidemiology and drug safety*. 95

Tibshirani, R. 1996, «Regression Shrinkage and Selection via the Lasso», *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 58, n° 1, p. 267–288. 6

Tsiropoulos, I., M. Andersen et J. Hallas. 2009, «Adverse events with use of antiepileptic drugs : a prescription and event symmetry analysis», *Pharmacoepidemiology and Drug Safety*, vol. 18, n° 6, p. 483–491. 84

Tuppin, P., J. Rudant, P. Constantinou, C. Gastaldi-Ménager, A. Rachas, L. de Roquefeuil, G. Maura, H. Caillol, A. Tajahmady, J. Coste, C. Gissot, A. Weill et A. Fagot-Campagna. 2017, «L'utilité d'une base médico-administrative nationale pour guider la décision publique : du système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) vers le système national

des données de santé (SNDS) en France», *Revue d'Epidemiologie et de Sante Publique*, vol. 65, p. S149–S167. 78

Uddin, M. J., R. H. Groenwold, M. S. Ali, A. de Boer, K. C. Roes, M. A. Chowdhury et O. H. Klungel. 2016, «Methods to control for unmeasured confounding in pharmacoepidemiology : an overview», *International Journal of Clinical Pharmacy*, vol. 38, n° 3, p. 714–723. 79

Vansteelandt, S. et R. M. Daniel. 2014, «On regression adjustment for the propensity score», *Statistics in Medicine*, vol. 33, n° 23, p. 4053–4072. 21

Volatier, É., F. Salvo, A. Pariente, É. Courtois, S. Escolano, P. Tubert-Bitter et I. Ahmed. «High-dimensional propensity score-adjusted case-crossover for discovering adverse drug reactions from computerized administrative healthcare databases», *soumis à Statistics in Medicine*. 23, 80

Wahab, I. A., N. L. Pratt, M. D. Wiese, L. M. Kalisch et E. E. Roughead. 2013, «The validity of sequence symmetry analysis (SSA) for adverse drug reaction signal detection», *Pharmacoepidemiology and Drug Safety*, vol. 22, n° 5, p. 496–502. 87

Wang, H., B. Li et C. Leng. 2009, «Shrinkage tuning parameter selection with a diverging number of parameters», *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 71, n° 3, p. 671–683. 45, 108

Wang, L., M. Rastegar-Mojarad, Z. Ji, S. Liu, K. Liu, S. Moon, F. Shen, Y. Wang, L. Yao, J. M. Davis III et H. Liu. 2018, «Detecting Pharmacovigilance Signals Combining Electronic Medical Records With Spontaneous Reports : A Case Study of Conventional Disease-Modifying Antirheumatic Drugs for Rheumatoid Arthritis», *Frontiers in Pharmacology*, vol. 9, n° August, p. 1–11. 89

Wang, P. S., S. Schneeweiss, R. J. Glynn, H. Mogun et J. Avorn. 2004, «Use of the case-crossover design to study prolonged drug exposures and insidious outcomes», *Annals of Epidemiology*, vol. 14, n° 4, p. 296–303. 80

Wang, T. et L. Zhu. 2011, «Consistent tuning parameter selection in high dimensional

sparse linear regression», *Journal of Multivariate Analysis*, vol. 102, n° 7, p. 1141–1151. 45, 108

Weber, J. C. P. 1984, «Epidemiology of adverse reactions to nonsteroidal antiinflammatory drugs», dans *Advances in Inflammation Research*, édité par K. Rainsford et G. Velo, Raven Press. 8, 77, 109

Westreich, D., J. Lessler et M. J. Funk. 2010, «Propensity score estimation : neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression», *Journal of Clinical Epidemiology*, vol. 63, n° 8, p. 826–833. 24

Whitaker, H. J., M. N. Hocine et C. P. Farrington. 2009, «The methodology of self-controlled case series studies», *Statistical Methods in Medical Research*, vol. 18, n° 1, p. 7–26. 95

Williamson, E. J., R. Morley, A. Lucas et J. R. Carpenter. 2012, «Variance estimation for stratified propensity score estimators», *Statistics in Medicine*, vol. 31, n° 15, p. 1617–1632. 44

Yoshida, K., S. Hernández-Díaz, D. H. Solomon, J. W. Jackson, J. J. Gagne, R. J. Glynn et J. M. Franklin. 2017, «Matching weights to simultaneously compare three treatment groups : Comparison to three-way matching», *Epidemiology*, vol. 28, n° 3, p. 387–395. 42

Zhou, X., I. J. Douglas, R. Shen et A. Bate. 2018, «Signal detection for recently approved products : Adapting and Evaluating Self-Controlled Case Series Method using a US Claims and UK EMR Database», *Drug Safety*, vol. 41, n° 5, p. 523–536. 95

Zorych, I., D. Madigan, P. Ryan et A. Bate. 2013, «Disproportionality methods for pharmacovigilance in longitudinal observational databases», *Statistical Methods in Medical Research*, vol. 22, n° 1, p. 39–56. 79

Zou, B., F. Zou, J. J. Shuster, P. J. Tighe, G. G. Koch et H. Zhou. 2016, «On variance estimate for covariate adjustment by propensity score analysis», *Statistics in Medicine*, vol. 35, n° 20, p. 3537–3548. 44

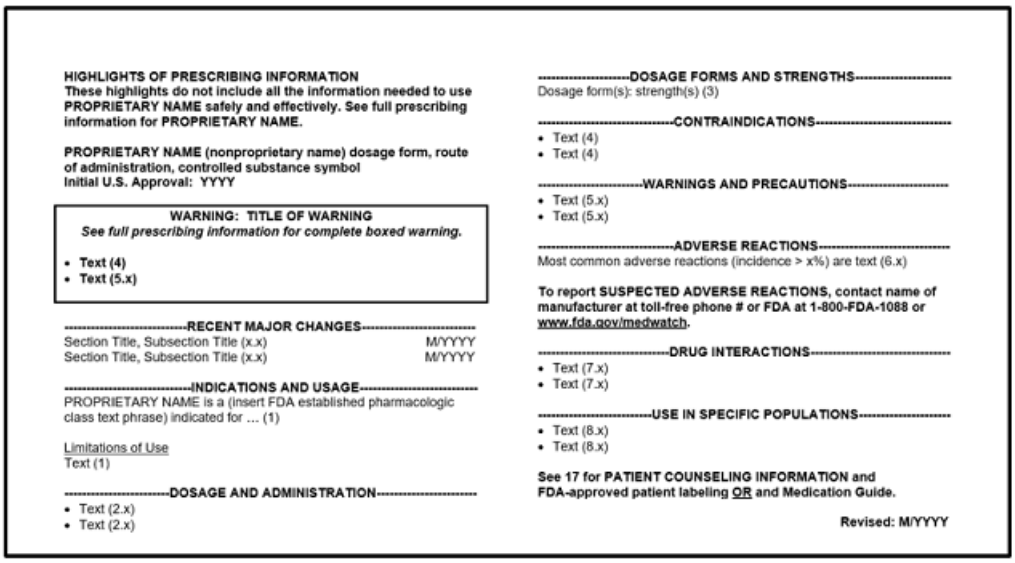
Zou, H. 2006, «The adaptive lasso and its oracle properties», *Journal of the American Statistical Association*, vol. 101, n° 476, p. 1418–1429. 7, 50, 51

Zou, H., T. Hastie, R. Tibshirani *et al.*. 2007, «On the "degrees of freedom" of the lasso», *The Annals of Statistics*, vol. 35, n° 5, p. 2173–2192. 45

Annexe A

Documents complémentaires pour la création de l'ensemble de signaux de référence qui concerne l'évènement indésirable *drug-induced liver injury*

FIGURE A.1 – Format des résumés des caractéristiques des produits approuvés par la *Food and Drug Administration*



HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use PROPRIETARY NAME safely and effectively. See full prescribing information for PROPRIETARY NAME.

PROPRIETARY NAME (nonproprietary name) dosage form, route of administration, controlled substance symbol
Initial U.S. Approval: YYYY

WARNING: TITLE OF WARNING
See full prescribing information for complete boxed warning.

- **Text (4)**
- **Text (5.x)**

----------RECENT MAJOR CHANGES----------
Section Title, Subsection Title (x.x) M/YYYY
Section Title, Subsection Title (x.x) M/YYYY

----------INDICATIONS AND USAGE----------
PROPRIETARY NAME is a (insert FDA established pharmacologic class text phrase) indicated for ... (1)

Limitations of Use
Text (1)

----------DOSAGE AND ADMINISTRATION----------
- Text (2.x)
- Text (2.x)

----------DOSAGE FORMS AND STRENGTHS----------
Dosage form(s): strength(s) (3)

----------CONTRAINDICATIONS----------
- Text (4)
- Text (4)

----------WARNINGS AND PRECAUTIONS----------
- Text (5.x)
- Text (5.x)

----------ADVERSE REACTIONS----------
Most common adverse reactions (incidence > x%) are text (6.x)

To report SUSPECTED ADVERSE REACTIONS, contact name of manufacturer at toll-free phone # or FDA at 1-800-FDA-1088 or www.fda.gov/medwatch.

----------DRUG INTERACTIONS----------
- Text (7.x)
- Text (7.x)

----------USE IN SPECIFIC POPULATIONS----------
- Text (8.x)
- Text (8.x)

See 17 for PATIENT COUNSELING INFORMATION and FDA-approved patient labeling OR and Medication Guide.

Revised: M/YYYY

Tableau A.1 – Nom et codes ATC des médicaments listés par Chen *et al.* (2011, 2016) dans la création de l'ensemble de référence DILIrank. Pour chaque médicament est précisé : la section du résumé des caractéristiques de ce médicament où apparaît un mot-clé relatif à DILI, le degré d'association de ce médicament avec DILI, et la sévérité du DILI associé à ce mot-clé.

Nom du médicament	Section de l'étiquette	Niveau d'association avec DILI	Score de sévérité	Code ATC
ceftriaxone	Adverse reactions	Less-DILI-Concern	4	J01DD04
trimethoprim	Adverse reactions	Less-DILI-Concern	4	J01EA01
tetracycline	Warnings and precautions	Less-DILI-Concern	2	S03AA02 S01AA09 A01AB13 J01AA07 S02AA08 D06AA04
dapsone	Warnings and precautions	Less-DILI-Concern	3	J04BA02
pyrazinamide	Warnings and precautions	Less-DILI-Concern	3	J04AK01
fenofibrate	Warnings and precautions	Less-DILI-Concern	3	C10AB05
cyclophosphamide	Adverse reactions	Less-DILI-Concern	5	L01AA01
chlorpromazine	Adverse reactions	Less-DILI-Concern	2	N05AA01
doxorubicin	Adverse reactions	Less-DILI-Concern	3	L01DB01
clindamycin	Warnings and precautions	Less-DILI-Concern	3	J01FF01 D10AF01 G01AA10
pyrimethamine	Warnings and precautions	Less-DILI-Concern	3	P01BD01
oxaprozin	Warnings and precautions	Less-DILI-Concern	3	M01AE12
dicloxacillin	Warnings and precautions	Less-DILI-Concern	3	J01CF01
captopril	Adverse reactions	Less-DILI-Concern	7	C09AA01
doxycycline	Adverse reactions	Less-DILI-Concern	3	J01AA02 A01AB22
cefadroxil	Adverse reactions	Less-DILI-Concern	7	J01DB05
hydrochlorothiazide	Adverse reactions	Less-DILI-Concern	2	C03AA03
lisinopril	Adverse reactions	Less-DILI-Concern	3	C09AA03
methylphenidate	Adverse reactions	Less-DILI-Concern	3	N06BA04
quinapril	Adverse reactions	Less-DILI-Concern	3	C09AA06
celecoxib	Warnings and precautions	Less-DILI-Concern	3	M01AH01 L01XX33
fluoxetine	Adverse reactions	Less-DILI-Concern	3	N06AB03
paroxetine	Adverse reactions	Less-DILI-Concern	8	N06AB05
sertraline	Adverse reactions	Less-DILI-Concern	3	N06AB06
amoxicillin	Adverse reactions	Less-DILI-Concern	5	J01CA04
valsartan	Adverse reactions	Less-DILI-Concern	3	C09CA03
simvastatin	Warnings and precautions	Less-DILI-Concern	3	C10AA01
lovastatin	Warnings and precautions	Less-DILI-Concern	3	C10AA02
naproxen	Warnings and precautions	Less-DILI-Concern	3	M01AE02 M02AA12 G02CC02
ibuprofen	Warnings and precautions	Less-DILI-Concern	3	M01AE01 M02AA13 G02CC01 C01EB16
amitriptyline	Adverse reactions	Less-DILI-Concern	5	N06AA09
ranitidine	Adverse reactions	Less-DILI-Concern	5	A02BA02
azithromycin	Adverse reactions	Less-DILI-Concern	7	S01AA26 J01FA10
lamivudine	Warnings and precautions	Less-DILI-Concern	3	J05AF05
verapamil	Warnings and precautions	Less-DILI-Concern	3	C08DA01
bupropion	Adverse reactions	Less-DILI-Concern	5	N06AX12
gabapentin	Warnings and precautions	Less-DILI-Concern	3	N03AX12
atenolol	Adverse reactions	Less-DILI-Concern	4	C07AB03
propafenone	Adverse reactions	Less-DILI-Concern	3	C01BC03
amiloride	Adverse reactions	Less-DILI-Concern	3	C03DB01
propofol	Adverse reactions	Less-DILI-Concern	3	N01AX10
meloxicam	Warnings and precautions	Less-DILI-Concern	3	M01AC06
fluvastatin	Warnings and precautions	Less-DILI-Concern	3	C10AA04
topiramate	Adverse reactions	Less-DILI-Concern	3	N03AX11
pravastatin	Warnings and precautions	Less-DILI-Concern	3	C10AA03
modafinil	Warnings and precautions	Less-DILI-Concern	3	N06BA07
amlodipine	Adverse reactions	Less-DILI-Concern	5	C08CA01
omeprazole	Adverse reactions	Less-DILI-Concern	4	A02BC01
venlafaxine	Adverse reactions	Less-DILI-Concern	7	N06AX16
phenobarbital	Warnings and precautions	Less-DILI-Concern	3	N03AA02
hydralazine	Adverse reactions	Less-DILI-Concern	3	C02DB02
methyltestosterone	Warnings and precautions	Less-DILI-Concern	2	G03EK01 G03BA02
quetiapine	Warnings and precautions	Less-DILI-Concern	3	N05AH04
oxcarbazepine	Warnings and precautions	Less-DILI-Concern	3	N03AF02
ezetimibe	Warnings and precautions	Less-DILI-Concern	3	C10AX09
docetaxel	Warnings and precautions	Less-DILI-Concern	3	L01CD02
octreotide	Warnings and precautions	Less-DILI-Concern	2	H01CB02
norfloxacin	Warnings and precautions	Less-DILI-Concern	3	S01AX12 J01MA06
oxacillin	Warnings and precautions	Less-DILI-Concern	3	J01CF04
salsalate	Warnings and precautions	Less-DILI-Concern	3	N02BA06
succimer	Warnings and precautions	Less-DILI-Concern	2	V09CA02
valdecoxib	Warnings and precautions	Less-DILI-Concern	3	M01AH03
rosuvastatin	Warnings and precautions	Less-DILI-Concern	3	C10AA07

sulfadiazine	Adverse reactions	Less-DILI-Concern	3	J01EC02
lansoprazole	Adverse reactions	Less-DILI-Concern	3	A02BC03
metoprolol	Adverse reactions	Less-DILI-Concern	5	C07AB02
felodipine	Adverse reactions	Less-DILI-Concern	3	C08CA02
cefazolin	Adverse reactions	Less-DILI-Concern	3	J01DB04
cefepime	Adverse reactions	Less-DILI-Concern	3	J01DE01
cefaclor	Adverse reactions	Less-DILI-Concern	3	J01DC04
cefuroxime	Adverse reactions	Less-DILI-Concern	5	J01DC02
valaciclovir	Adverse reactions	Less-DILI-Concern	3	J05AB11
moxifloxacin	Adverse reactions	Less-DILI-Concern	8	J01MA14 S01AX22
nelfinavir	Adverse reactions	Less-DILI-Concern	3	J05AE04
hydroxychloroquine	Adverse reactions	Less-DILI-Concern	7	P01BA02
escitalopram	Adverse reactions	Less-DILI-Concern	7	N06AB10
melphalan	Adverse reactions	Less-DILI-Concern	5	L01AA03
flecainide	Adverse reactions	Less-DILI-Concern	7	C01BC04
ampicillin	Warnings and precautions	Less-DILI-Concern	3	S01AA19 J01CA01
metaxalone	Adverse reactions	Less-DILI-Concern	5	
acamprosate	Adverse reactions	Less-DILI-Concern	3	N07BB03
clopidogrel	Adverse reactions	Less-DILI-Concern	7	B01AC04
clonazepam	Adverse reactions	Less-DILI-Concern	3	N03AE01
progesterone	Adverse reactions	Less-DILI-Concern	2	G03DA04
losartan	Adverse reactions	Less-DILI-Concern	4	C09CA01
clofibrate	Warnings and precautions	Less-DILI-Concern	3	C10AB01
nifedipine	Warnings and precautions	Less-DILI-Concern	3	C08CA05
piroxicam	Warnings and precautions	Less-DILI-Concern	3	M01AC01 M02AA07 S01BC06
quinidine	Adverse reactions	Less-DILI-Concern	3	C01BA01
etoposide	Adverse reactions	Less-DILI-Concern	3	L01CB01
chlorpropamide	Adverse reactions	Less-DILI-Concern	2	A10BB02
imipramine	Adverse reactions	Less-DILI-Concern	3	N06AA02
prochlorperazine	Adverse reactions	Less-DILI-Concern	2	N05AB04
carboplatin	Warnings and precautions	Less-DILI-Concern	3	L01XA02
cisplatin	Warnings and precautions	Less-DILI-Concern	3	L01XA01
mitoxantrone	Warnings and precautions	Less-DILI-Concern	3	L01DB07
ketorolac	Warnings and precautions	Less-DILI-Concern	3	M01AB15 S01BC05
clomiphene	Warnings and precautions	Less-DILI-Concern	3	G03GB02
cloxacillin	Warnings and precautions	Less-DILI-Concern	3	J01CF02
mebendazole	Warnings and precautions	Less-DILI-Concern	3	P02CA01
quinine	Warnings and precautions	Less-DILI-Concern	3	P01BC01
saquinavir	Warnings and precautions	Less-DILI-Concern	3	J05AE01
auranofin	Warnings and precautions	Less-DILI-Concern	3	M01CB03
paclitaxel	Adverse reactions	Less-DILI-Concern	4	L01CD01
vincristine	Adverse reactions	Less-DILI-Concern	8	L01CA02
spironolactone	Adverse reactions	Less-DILI-Concern	2	C03DA01
terbutaline	Adverse reactions	Less-DILI-Concern	3	R03CC03 R03AC03
disopyramide	Adverse reactions	Less-DILI-Concern	2	C01BA03
rifabutin	Adverse reactions	Less-DILI-Concern	5	J04AB04
metoclopramide	Adverse reactions	Less-DILI-Concern	5	A03FA01
trazodone	Adverse reactions	Less-DILI-Concern	5	N06AX05
risperidone	Adverse reactions	Less-DILI-Concern	3	N05AX08
thioguanine	Warnings and precautions	Less-DILI-Concern	2	L01BB03
procarbazine	Warnings and precautions	Less-DILI-Concern	3	L01XB01
ifosfamide	Adverse reactions	Less-DILI-Concern	4	L01AA06
tolazamide	Adverse reactions	Less-DILI-Concern	2	A10BB05
tolmetin	Adverse reactions	Less-DILI-Concern	3	M01AB03 M02AA21
cyproheptadine	Adverse reactions	Less-DILI-Concern	7	R06AX02
aspirin	No match	Less-DILI-Concern	0	N02BA01 B01AC06 A01AD05
phentermine	No match	Less-DILI-Concern	0	A08AA01
fluorouracil	No match	Less-DILI-Concern	0	L01BC02
metformin	No match	Less-DILI-Concern	0	A10BA02
finasteride	No match	Less-DILI-Concern	0	G04CB01 D11AX10
benzonatate	No match	Less-DILI-Concern	0	R05DB01
metronidazole	No match	Less-DILI-Concern	3	J01XD01 D06BX01 A01AB17 P01AB01 G01AF01
vancomycin	No match	Less-DILI-Concern	0	A07AA09 J01XA01
sulfamethizole	No match	Less-DILI-Concern	0	J01EB02 B05CA04 D06BA04 S01AB01

prasugrel hydrochloride	No match	Less-DILI-Concern	0	B01AC22
ketamine	No match	Less-DILI-Concern	0	N01AX03
alendronate	No match	Less-DILI-Concern	0	M05BA04
aliskiren	No match	Less-DILI-Concern	0	C09XA02
aceclofenac	No match	Less-DILI-Concern	0	M02AA25 M01AB16
zoledronate	No match	Less-DILI-Concern	0	M05BA08
abatacept	No match	Less-DILI-Concern	0	L04AA24
vitamin a	Warnings and precautions	Less-DILI-Concern	3	D10AD02 S01XA02 R01AX02
pioglitazone	Warnings and precautions	Less-DILI-Concern	3	A10BG03
rosiglitazone	Warnings and precautions	Less-DILI-Concern	3	A10BG02
donepezil	Adverse reactions	Less-DILI-Concern	3	N06DA02
warfarin	Adverse reactions	Less-DILI-Concern	5	B01AA03
prednisone	Adverse reactions	Less-DILI-Concern	3	H02AB07 A07EA03
chloroquine	Adverse reactions	Less-DILI-Concern	3	P01BA01
montelukast	Adverse reactions	Less-DILI-Concern	3	R03DC03
clotrimazole	Warnings and precautions	Less-DILI-Concern	3	D01AC01 A01AB18 G01AF02
cromoglicate	Adverse reactions	Less-DILI-Concern	3	R03BC01 S01GX01 R01AC01 A07EB01 D11AX17
cefoperazone	Adverse reactions	Less-DILI-Concern	3	J01DD12
amphotericin b	Warnings and precautions	Less-DILI-Concern	3	J02AA01 A01AB04 A07AA07 G01AA03
estradiol	Adverse reactions	Less-DILI-Concern	2	G03CA03
cimetidine	Adverse reactions	Less-DILI-Concern	2	A02BA01
ceftazidime	Adverse reactions	Less-DILI-Concern	3	J01DD02
phenelzine	Adverse reactions	Less-DILI-Concern	3	N06AF03
nafcillin	Warnings and precautions	Less-DILI-Concern	3	J01CF06
famotidine	Adverse reactions	Less-DILI-Concern	3	A02BA03
citalopram	Adverse reactions	Less-DILI-Concern	7	N06AB04
candesartan	Adverse reactions	Less-DILI-Concern	3	C09CA06
irbesartan	Adverse reactions	Less-DILI-Concern	5	C09CA04
telmisartan	Adverse reactions	Less-DILI-Concern	3	C09CA07
benazepril	Adverse reactions	Less-DILI-Concern	4	C09AA07
glimepiride	Adverse reactions	Less-DILI-Concern	7	A10BB12
ketoprofen	Warnings and precautions	Less-DILI-Concern	3	M01AE03 M02AA10
promethazine	Adverse reactions	Less-DILI-Concern	5	R06AD02 D04AA10
loracarbef	Adverse reactions	Less-DILI-Concern	3	J01DC08
tranylcypromine	Adverse reactions	Less-DILI-Concern	3	N06AF04
memantine	Adverse reactions	Less-DILI-Concern	7	N06DX01
penicillin g	Adverse reactions	Less-DILI-Concern	3	J01CE01 S01AA14
deferoxamine	No match	Less-DILI-Concern	0	V03AC01
entacapone	No match	Less-DILI-Concern	0	N04BX02
amphetamine	No match	Less-DILI-Concern	0	N06BA01
chlorcyclizine	No match	Less-DILI-Concern	0	R06AE04
tamsulosin	No match	Less-DILI-Concern	0	G04CA02
mitomycin	No match	Less-DILI-Concern	0	L01DC03
fospropofol	No match	Less-DILI-Concern	0	
anakinra	No match	Less-DILI-Concern	0	L04AC03
darbepoetin alfa	No match	Less-DILI-Concern	0	B03XA02
trastuzumab	No match	Less-DILI-Concern	0	L01XC03
temozolomide	Adverse reactions	Less-DILI-Concern	4	L01AX03
olanzapine	Adverse reactions	Less-DILI-Concern	3	N05AH03
leuprolide	Adverse reactions	Less-DILI-Concern	3	L02AE02
ofloxacin	Warnings and precautions	Less-DILI-Concern	3	S01AX11 J01MA01
meropenem	Warnings and precautions	Less-DILI-Concern	3	J01DH02
chlorambucil	Warnings and precautions	Less-DILI-Concern	3	L01AA02
penicillamine	Warnings and precautions	Less-DILI-Concern	2	M01CC01
letrozole	Warnings and precautions	Less-DILI-Concern	2	L02BG04
adefovir	Warnings and precautions	Less-DILI-Concern	2	J05AF08
etanercept	Warnings and precautions	Less-DILI-Concern	3	L04AB01
methoxsalen	Warnings and precautions	Less-DILI-Concern	3	D05AD02 D05BA02
mirtazapine	Warnings and precautions	Less-DILI-Concern	3	N06AX11
atazanavir	Warnings and precautions	Less-DILI-Concern	3	J05AE08
perphenazine	Warnings and precautions	Less-DILI-Concern	3	N05AB03
teniposide	Warnings and precautions	Less-DILI-Concern	3	L01CB02
tenofovir	Warnings and precautions	Less-DILI-Concern	3	J05AF07
aciclovir	Adverse reactions	Less-DILI-Concern	5	S01AD03 D06BB03 J05AB01

zonisamide	Adverse reactions	Less-DILI-Concern	2	N03AX15
carvedilol	Adverse reactions	Less-DILI-Concern	3	C07AG02
glipizide	Adverse reactions	Less-DILI-Concern	3	A10BB07
anastrozole	Adverse reactions	Less-DILI-Concern	3	L02BG03
glibenclamide	Adverse reactions	Less-DILI-Concern	3	A10BB01
nizatidine	Adverse reactions	Less-DILI-Concern	5	A02BA04
haloperidol	Adverse reactions	Less-DILI-Concern	5	N05AD01
mesalazine	Adverse reactions	Less-DILI-Concern	7	A07EC02
tacrolimus	Adverse reactions	Less-DILI-Concern	5	D11AH01 L04AD02
thalidomide	Adverse reactions	Less-DILI-Concern	4	L04AX02
triptorelin pamoate	Adverse reactions	Less-DILI-Concern	3	L02AE04
methylprednisolone	Adverse reactions	Less-DILI-Concern	3	D10AA02 H02AB04 D07AA01
alprazolam	Adverse reactions	Less-DILI-Concern	7	N05BA12
ivermectin	Adverse reactions	Less-DILI-Concern	3	P02CF01
pantoprazole	Adverse reactions	Less-DILI-Concern	3	A02BC02
prednisolone	Adverse reactions	Less-DILI-Concern	3	R01AD02 S03BA02 D07AA03 S02BA03 H02AB06 S01BA04 C0
tinidazole	Adverse reactions	Less-DILI-Concern	3	P01AB02 J01XD02
bendroflumethiazide	Adverse reactions	Less-DILI-Concern	5	C03AA01
dalteparin sodium	Adverse reactions	Less-DILI-Concern	3	B01AB04
remifentanil	Adverse reactions	Less-DILI-Concern	3	N01AH06
sibutramine	Adverse reactions	Less-DILI-Concern	4	A08AA10
repaglinide	Adverse reactions	Less-DILI-Concern	5	A10BX02
sildenafil	Adverse reactions	Less-DILI-Concern	3	G04BE03
propoxyphene	Adverse reactions	Less-DILI-Concern	3	N02AC04
desloratadine	Adverse reactions	Less-DILI-Concern	3	R06AX27
alfuzosin	Adverse reactions	Less-DILI-Concern	5	G04CA01
baclofen	Adverse reactions	Less-DILI-Concern	3	M03BX01
cefotaxime	Adverse reactions	Less-DILI-Concern	5	J01DD01
cephalexin	Adverse reactions	Less-DILI-Concern	4	J01DB01
cetirizine	Adverse reactions	Less-DILI-Concern	3	R06AE07
chlordiazepoxide	Adverse reactions	Less-DILI-Concern	5	N05BA02
daptomycin	Adverse reactions	Less-DILI-Concern	5	J01XX09
cyclobenzaprine	Adverse reactions	Less-DILI-Concern	5	M03BX08
sitagliptin	Adverse reactions	Less-DILI-Concern	4	A10BH01
cefdinir	Adverse reactions	Less-DILI-Concern	7	J01DD15
pegaspargase	Adverse reactions	Less-DILI-Concern	3	L01XX24
adalimumab	Adverse reactions	Less-DILI-Concern	7	L04AB04
flurazepam	Adverse reactions	Less-DILI-Concern	4	N05CD01
fosfomycin	Adverse reactions	Less-DILI-Concern	6	J01XX01
fosinopril	Adverse reactions	Less-DILI-Concern	3	C09AA09
heparin sodium	Adverse reactions	Less-DILI-Concern	3	C05BA03 S01XA14 B01AB01
isradipine	Adverse reactions	Less-DILI-Concern	3	C08CA03
esomeprazole Sodium	Adverse reactions	Less-DILI-Concern	4	A02BC05
procainamide	Adverse reactions	Less-DILI-Concern	4	C01BA02
ramipril	Adverse reactions	Less-DILI-Concern	7	C09AA05
mefloquine	Adverse reactions	Less-DILI-Concern	7	P01BC02
trandolapril	Adverse reactions	Less-DILI-Concern	4	C09AA10
triamterene	Adverse reactions	Less-DILI-Concern	5	C03DB02
dexmethylphenidate	Adverse reactions	Less-DILI-Concern	7	N06BA11
pentamidine	Warnings and precautions	Less-DILI-Concern	3	P01CX01
bromocriptine	Warnings and precautions	Less-DILI-Concern	3	G02CB01 N04BC01
lomustine	Warnings and precautions	Less-DILI-Concern	3	L01AD02
raltegravir	Warnings and precautions	Most-DILI-Concern	7	J05AX08
gemifloxacin	Warnings and precautions	Less-DILI-Concern	3	J01MA15
carbenicillin	Warnings and precautions	Less-DILI-Concern	3	J01CA03
lenalidomide	Warnings and precautions	Less-DILI-Concern	3	L04AX04
ethionamide	Warnings and precautions	Less-DILI-Concern	3	J04AD03
thiotepa	Warnings and precautions	Less-DILI-Concern	3	L01AC01
rabeprazole	Adverse reactions	Less-DILI-Concern	3	A02BC04
sirolimus	Adverse reactions	Less-DILI-Concern	8	L04AA10
desvenlafaxine	Adverse reactions	Less-DILI-Concern	3	N06AX23
levocetirizine dihydrochloride	Adverse reactions	Less-DILI-Concern	4	R06AE09
enoxaparin	Adverse reactions	Less-DILI-Concern	3	B01AB05
ropinirole	Adverse reactions	Less-DILI-Concern	4	N04BC04

ziprasidone	Adverse reactions	Less-DILI-Concern	3	N05AE04
acebutolol	Adverse reactions	Less-DILI-Concern	3	C07AB04
levetiracetam	Adverse reactions	Less-DILI-Concern	8	N03AX14
alosetron	Adverse reactions	Less-DILI-Concern	4	A03AE01
levothyroxine	Adverse reactions	Less-DILI-Concern	3	H03AA01
linezolid	Adverse reactions	Less-DILI-Concern	3	J01XX08
fondaparinux sodium	Adverse reactions	Less-DILI-Concern	3	B01AX05
cefprozil	Adverse reactions	Less-DILI-Concern	4	J01DC10
chlorothiazide	Adverse reactions	Less-DILI-Concern	2	C03AA04
nelarabine	Adverse reactions	Less-DILI-Concern	4	L01BB07
varenicline	Adverse reactions	Less-DILI-Concern	2	N07BA03
dasatinib	Adverse reactions	Less-DILI-Concern	4	L01XE06
dexlansoprazole	Adverse reactions	Less-DILI-Concern	4	A02BC06
thioridazine	Adverse reactions	Less-DILI-Concern	5	N05AC02
irinotecan	Adverse reactions	Less-DILI-Concern	3	L01XX19
methazolamide	Adverse reactions	Less-DILI-Concern	7	S01EC05
ondansetron	Adverse reactions	Less-DILI-Concern	7	A04AA01
mycophenolate mofetil	Adverse reactions	Less-DILI-Concern	3	
doxepin	Adverse reactions	Less-DILI-Concern	4	N06AA12
canakinumab	Adverse reactions	Less-DILI-Concern	3	L04AC08
riluzole	Warnings and precautions	Most-DILI-Concern	8	N07XX02
cyclosporine	Warnings and precautions	Most-DILI-Concern	7	L04AD01 S01XA18
fenoprofen	Warnings and precautions	Most-DILI-Concern	8	M01AE04
acetazolamide	Warnings and precautions	Most-DILI-Concern	8	S01EC01
nortriptyline	Warnings and precautions	Most-DILI-Concern	8	N06AA10
bexarotene	Warnings and precautions	Most-DILI-Concern	8	L01XX25
sorafenib	Warnings and precautions	Most-DILI-Concern	8	L01XE05
peginterferon alfa-2b	Warnings and precautions	Most-DILI-Concern	4	L03AB10
papaverine	Warnings and precautions	Most-DILI-Concern	5	G04BE02 A03AD01
nimesulide	Withdrawn	Most-DILI-Concern	8	M01AX17
troglitazone	Withdrawn	Most-DILI-Concern	8	A10BG01
amineptine	Withdrawn	Most-DILI-Concern	8	N06AA19
trovafloxacin	Withdrawn	Most-DILI-Concern	8	J01MA13
droxicam	Withdrawn	Most-DILI-Concern	8	M01AC04
tetrabamate	Withdrawn	Most-DILI-Concern	8	
ebrotidine	Withdrawn	Most-DILI-Concern	8	
etravirine	Warnings and precautions	Most-DILI-Concern	8	J05AG04
micafungin	Warnings and precautions	Most-DILI-Concern	7	J02AX05
erythromycin	Warnings and precautions	Most-DILI-Concern	5	D10AF02 S01AA17 J01FA01
zafirlukast	Warnings and precautions	Most-DILI-Concern	8	R03DC01
zileuton	Warnings and precautions	Most-DILI-Concern	5	
chlorzoxazone	Warnings and precautions	Most-DILI-Concern	8	M03BB03
atomoxetine	Warnings and precautions	Most-DILI-Concern	8	N06BA09
ritonavir	Warnings and precautions	Most-DILI-Concern	8	J05AE03
diltiazem	Warnings and precautions	Most-DILI-Concern	4	C08DB01
clarithromycin	Warnings and precautions	Most-DILI-Concern	8	J01FA09
rifampin	Warnings and precautions	Most-DILI-Concern	8	J04AB02
levofloxacin	Warnings and precautions	Most-DILI-Concern	8	J01MA12 S01AX19
testosterone	Warnings and precautions	Most-DILI-Concern	8	G03BA03
mercaptopurine	Warnings and precautions	Most-DILI-Concern	8	L01BB02
acetaminophen	Warnings and precautions	Most-DILI-Concern	5	N02BE01
phenytoin	Warnings and precautions	Most-DILI-Concern	8	N03AB02
allopurinol	Warnings and precautions	Most-DILI-Concern	8	M04AA01
carbamazepine	Warnings and precautions	Most-DILI-Concern	7	N03AF01
labetalol	Warnings and precautions	Most-DILI-Concern	8	C07AG01
minocycline	Warnings and precautions	Most-DILI-Concern	8	J01AA08 A01AB23
nitrofurantoin	Warnings and precautions	Most-DILI-Concern	8	J01XE01
tamoxifen	Warnings and precautions	Most-DILI-Concern	8	L02BA01
fluconazole	Warnings and precautions	Most-DILI-Concern	8	J02AC01 D01AC15
niacin	Warnings and precautions	Most-DILI-Concern	7	C10AD02 C04AC01
sulfasalazine	Warnings and precautions	Most-DILI-Concern	5	A07EC01
indomethacin	Warnings and precautions	Most-DILI-Concern	8	M01AB01 S01BC01 C01EB03 M02AA23
diclofenac	Warnings and precautions	Most-DILI-Concern	8	M01AB05 D11AX18 S01BC03 M02AA15
ciprofloxacin	Warnings and precautions	Most-DILI-Concern	7	S02AA15 S03AA07 J01MA02 S01AX13

methyldopa	Warnings and precautions	Most-DILI-Concern	8	C02AB01
busulfan	Warnings and precautions	Most-DILI-Concern	8	L01AB01
4-aminosalicylic acid	Warnings and precautions	Most-DILI-Concern	5	J04AA01
dactinomycin	Warnings and precautions	Most-DILI-Concern	8	L01DA01
clozapine	Warnings and precautions	Most-DILI-Concern	5	N05AH02
orlistat	Warnings and precautions	Most-DILI-Concern	8	A08AB01
ticlopidine	Warnings and precautions	Most-DILI-Concern	4	B01AC05
azathioprine	Warnings and precautions	Most-DILI-Concern	5	L04AX01
sulindac	Warnings and precautions	Most-DILI-Concern	8	M01AB02
abacavir	Warnings and precautions	Most-DILI-Concern	8	J05AF06
albendazole	Warnings and precautions	Most-DILI-Concern	7	P02CA03
telithromycin	Warnings and precautions	Most-DILI-Concern	8	J01FA15
bicalutamide	Warnings and precautions	Most-DILI-Concern	8	L02BB03
hydroxyurea	Warnings and precautions	Most-DILI-Concern	8	L01XX05
clomipramine	Warnings and precautions	Most-DILI-Concern	8	N06AA04
griseofulvin	Warnings and precautions	Most-DILI-Concern	8	D01AA08 D01BA01
methimazole	Warnings and precautions	Most-DILI-Concern	8	H03BB02
diflunisal	Warnings and precautions	Most-DILI-Concern	8	N02BA11
isotretinoin	Warnings and precautions	Most-DILI-Concern	5	D10AD04 D10BA01
oxaliplatin	Warnings and precautions	Most-DILI-Concern	4	L01XA03
erlotinib	Warnings and precautions	Most-DILI-Concern	8	L01XE03
milnacipran	Warnings and precautions	Most-DILI-Concern	7	N06AX17
asparaginase	Warnings and precautions	Most-DILI-Concern	7	L01XX02
interferon alfa-2b	Warnings and precautions	Most-DILI-Concern	8	L03AB05
infliximab	Warnings and precautions	Most-DILI-Concern	8	L04AB02
gemcitabine	Warnings and precautions	Most-DILI-Concern	8	L01BC05
gemfibrozil	Warnings and precautions	Most-DILI-Concern	4	C10AB04
gefitinib	Warnings and precautions	Most-DILI-Concern	4	L01XE02
lamotrigine	Warnings and precautions	Most-DILI-Concern	7	N03AX09
atorvastatin	Warnings and precautions	Most-DILI-Concern	5	C10AA05
disulfiram	Warnings and precautions	Most-DILI-Concern	8	P03AA04 N07BB01
nilutamide	Warnings and precautions	Most-DILI-Concern	8	L02BB02
terbinafine	Warnings and precautions	Most-DILI-Concern	8	D01BA02 D01AE15
imatinib	Warnings and precautions	Most-DILI-Concern	8	L01XE01
bortezomib	Warnings and precautions	Most-DILI-Concern	7	L01XX32
ethambutol	Warnings and precautions	Most-DILI-Concern	8	J04AK02
tizanidine	Warnings and precautions	Most-DILI-Concern	8	M03BX02
itraconazole	Warnings and precautions	Most-DILI-Concern	8	J02AC02
darunavir	Warnings and precautions	Most-DILI-Concern	8	J05AE10
efavirenz	Warnings and precautions	Most-DILI-Concern	7	J05AG03
duloxetine	Warnings and precautions	Most-DILI-Concern	8	N06AX21
thiabendazole	Warnings and precautions	Most-DILI-Concern	7	D01AC06 P02CA02
tolvaptan	Warnings and precautions	Most-DILI-Concern	8	C03XA01
dronedarone	Warnings and precautions	Most-DILI-Concern	7	C01BD07
febuxostat	Warnings and precautions	Most-DILI-Concern	8	M04AA03
acarbose	Warnings and precautions	Most-DILI-Concern	8	A10BF01
voriconazole	Warnings and precautions	Most-DILI-Concern	8	J02AC03
interferon alfa-2a, recombinant	Warnings and precautions	Most-DILI-Concern	8	L03AB04
interferon beta-1b	Warnings and precautions	Most-DILI-Concern	7	L03AB08
interferon beta-1a	Warnings and precautions	Most-DILI-Concern	7	L03AB07
interferon alfacon-1	Warnings and precautions	Most-DILI-Concern	7	L03AB09
estramustine	Warnings and precautions	Most-DILI-Concern	4	L01XX11
natalizumab	Warnings and precautions	Most-DILI-Concern	5	L04AA23
nandrolone decanoate	Warnings and precautions	Most-DILI-Concern	7	A14AB01
etodolac	Warnings and precautions	Most-DILI-Concern	8	M01AB08
exemestane	Warnings and precautions	Most-DILI-Concern	4	L02BG06
fosphenytoin	Warnings and precautions	Most-DILI-Concern	8	N03AB05
acitretin	Box warning	Most-DILI-Concern	5	D05BB02
ketoconazole	Box warning	Most-DILI-Concern	8	D01AC08 G01AF11 J02AB02
nefazodone	Box warning	Most-DILI-Concern	8	N06AX06
valproic acid	Box warning	Most-DILI-Concern	8	N03AG01
isoniazid	Box warning	Most-DILI-Concern	8	J04AC01
amiodarone	Box warning	Most-DILI-Concern	8	C01BD01
flutamide	Box warning	Most-DILI-Concern	8	L02BB01

zidovudine	Box warning	Most-DILI-Concern	8	J05AF01
methotrexate	Box warning	Most-DILI-Concern	3	L01BA01 L04AX03
dantrolene	Box warning	Most-DILI-Concern	8	M03CA01
oxymetholone	Box warning	Most-DILI-Concern	7	A14AA05
danazol	Box warning	Most-DILI-Concern	8	G03XA01
dacarbazine	Box warning	Most-DILI-Concern	6	L01AX04
tolcapone	Box warning	Most-DILI-Concern	8	N04BX01
leflunomide	Box warning	Most-DILI-Concern	8	L04AA13
mexiletine	Box warning	Most-DILI-Concern	3	C01BB02
bosentan	Box warning	Most-DILI-Concern	7	C02KX01
oxandrolone	Box warning	Most-DILI-Concern	8	A14AA08
lapatinib	Box warning	Most-DILI-Concern	8	L01XE07
propylthiouracil	Box warning	Most-DILI-Concern	8	H03BA02
nevirapine	Box warning	Most-DILI-Concern	8	J05AG01
deferasirox	Box warning	Most-DILI-Concern	7	V03AC03
sunitinib	Box warning	Most-DILI-Concern	8	L01XE04
pazopanib	Box warning	Most-DILI-Concern	8	L01XE11
tipranavir	Box warning	Most-DILI-Concern	8	J05AE09
chlormezanone	Withdrawn	Most-DILI-Concern	8	M03BB02
mefenamic acid	Warnings and precautions	Most-DILI-Concern	8	M01AG01
pentostatin	Box warning	Most-DILI-Concern	3	L01XX08
felbamate	Box warning	Most-DILI-Concern	7	N03AX10
cytarabine	Box warning	Most-DILI-Concern	3	L01BC01
iproniazid	Withdrawn	Most-DILI-Concern	8	N06AF05
bromfenac	Withdrawn	Most-DILI-Concern	8	S01BC11
perhexiline	Withdrawn	Most-DILI-Concern	8	C08EX02
benzbromarone	Withdrawn	Most-DILI-Concern	8	M04AB03
ticrynafen	Withdrawn	Most-DILI-Concern	8	C03CC02
alatrofloxacin mesylate	Withdrawn	Most-DILI-Concern	8	
glafenine	Withdrawn	Most-DILI-Concern	8	N02BG03
alpidem	Withdrawn	Most-DILI-Concern	8	
pirprofen	Withdrawn	Most-DILI-Concern	8	M01AE08
pemoline	Withdrawn	Most-DILI-Concern	8	N06BA05
nomifensine	Withdrawn	Most-DILI-Concern	8	N06AX04
benoxaprofen	Withdrawn	Most-DILI-Concern	8	M01AE06
ibufenac	Withdrawn	Most-DILI-Concern	8	
cinchophen	Withdrawn	Most-DILI-Concern	8	M04AC02
fipexide	Withdrawn	Most-DILI-Concern	8	N06BX05
sulfathiazole	Withdrawn	Most-DILI-Concern	8	D06BA02 J01EB07
sulfacarbamide	Withdrawn	Most-DILI-Concern	8	
triacetyldiphenolisatin	Withdrawn	Most-DILI-Concern	8	
xenazoic acid	Withdrawn	Most-DILI-Concern	8	
lumiracoxib	Withdrawn	Most-DILI-Concern	8	M01AH06
ximelagatran	Withdrawn	Most-DILI-Concern	8	B01AE05
tolrestat	Withdrawn	Most-DILI-Concern	8	A10XA01
moxisylyte	Withdrawn	Most-DILI-Concern	8	G04BE06 C04AX10
oxyphenisatin	Withdrawn	Most-DILI-Concern	8	A06AB01
clomacran	Withdrawn	Most-DILI-Concern	8	
isaxonine	Withdrawn	Most-DILI-Concern	8	
pipamazine	Withdrawn	Most-DILI-Concern	8	
phenoxypropazine	Withdrawn	Most-DILI-Concern	8	
tilbroquinol	Withdrawn	Most-DILI-Concern	8	P01AA05
bendazac	Withdrawn	Most-DILI-Concern	8	M02AA11 S01BC07
benzarone	Withdrawn	Most-DILI-Concern	8	
benziodarone	Withdrawn	Most-DILI-Concern	8	C01DX04
clometacin	Withdrawn	Most-DILI-Concern	8	
cyclofenil	Withdrawn	Most-DILI-Concern	8	G03GB01
exifone	Withdrawn	Most-DILI-Concern	8	
sulcotidil	Withdrawn	Most-DILI-Concern	8	C04AX19
zimelidine	Withdrawn	Most-DILI-Concern	8	N06AB02
mebanazine	Withdrawn	Most-DILI-Concern	8	
nialamide	Withdrawn	Most-DILI-Concern	8	N06AF02
niperotidine	Withdrawn	Most-DILI-Concern	8	A02BA05
mepazine	Withdrawn	Most-DILI-Concern	8	

nitrefazole	Withdrawn	Most-DILI-Concern	8	
sitaxsentan	Withdrawn	Most-DILI-Concern	8	C02KX03
alaproclate	Withdrawn	Most-DILI-Concern	8	N06AB07
alclofenac	Withdrawn	Most-DILI-Concern	8	M01AB06
fialuridine	Discontinued	Most-DILI-Concern	8	
tasosartan	Discontinued	Most-DILI-Concern	8	C09CA05
dermatan	Discontinued	Most-DILI-Concern	8	B01AX04
pralnacasan	Discontinued	Most-DILI-Concern	8	
falnidamol	Discontinued	Most-DILI-Concern	8	
aplaviroc	Discontinued	Most-DILI-Concern	8	
pafuramidine	Discontinued	Most-DILI-Concern	8	
fiduxosin	Discontinued	Most-DILI-Concern	8	
gemtuzumab ozogamicin	Box warning	Most-DILI-Concern	8	L01XC05
eltrombopag olamine	Box warning	Most-DILI-Concern	4	B02BX05
divalproex sodium	Box warning	Most-DILI-Concern	8	
maraviroc	Box warning	Most-DILI-Concern	3	J05AX09
phenoxybenzamine	No match	No-DILI-Concern	0	C04AX02
oxybutynin	No match	No-DILI-Concern	0	G04BD04
diphenhydramine	No match	No-DILI-Concern	0	R06AA02 D04AA32
orphenadrine	No match	No-DILI-Concern	0	M03BC01 N04AB02
clemastine	No match	No-DILI-Concern	0	D04AA14 R06AA04
theophylline	No match	No-DILI-Concern	0	R03DA04
digoxin	No match	No-DILI-Concern	0	C01AA05
liothyronine	No match	No-DILI-Concern	0	H03AA02
caffeine	No match	No-DILI-Concern	0	N06BC01
pyridostigmine	No match	No-DILI-Concern	0	N07AA02
acetylcysteine	No match	No-DILI-Concern	0	R05CB01 S01XA08 V03AB23
aminocaproic acid	No match	No-DILI-Concern	0	B02AA01
vitamin c	No match	No-DILI-Concern	0	S01XA15 G01AD03
leucovorin	No match	No-DILI-Concern	0	V03AF03
raloxifene	No match	No-DILI-Concern	0	G03XC01
ergocalciferol	No match	No-DILI-Concern	0	A11CC01
isosorbide dinitrate	No match	No-DILI-Concern	0	C05AE02 C01DA08
diphenoxylate	No match	No-DILI-Concern	0	A07DA01
pimozide	No match	No-DILI-Concern	0	N05AG02
dextromethorphan	No match	No-DILI-Concern	0	R05DA09
loperamide	No match	No-DILI-Concern	0	A07DA03
isoxsuprine	No match	No-DILI-Concern	0	C04AA01
pseudoephedrine	No match	No-DILI-Concern	0	R01BA02
isoproterenol	No match	No-DILI-Concern	0	R03AB02 R03CB01 C01CA02
primidone	No match	No-DILI-Concern	0	N03AA03
fludrocortisone	No match	No-DILI-Concern	0	H02AA02
fexofenadine	No match	No-DILI-Concern	0	R06AX26
kanamycin	No match	No-DILI-Concern	0	A07AA08 J01GB04 S01AA24
folic acid	No match	No-DILI-Concern	0	B03BB01
thiamine	No match	No-DILI-Concern	0	A11DA01
sorbitol	No match	No-DILI-Concern	0	B05CX02 V04CC01 A06AD18 A06AG07
paromomycin	No match	No-DILI-Concern	0	A07AA06
ammonium chloride	No match	No-DILI-Concern	0	B05XA04 G04BA01
methysergide	No match	No-DILI-Concern	0	N02CA04
methylergonovine	No match	No-DILI-Concern	0	G02AB01
betaine	No match	No-DILI-Concern	0	A16AA06
pyridoxine	No match	No-DILI-Concern	0	A11HA02
budesonide	No match	No-DILI-Concern	0	R03BA02 D07AC09 R01AD05 A07EA06
atropine	No match	No-DILI-Concern	0	A03BA01 S01FA01
pindolol	No match	No-DILI-Concern	0	C07AA03
brompheniramine	No match	No-DILI-Concern	0	R06AB01
primaquine	No match	No-DILI-Concern	0	P01BA03
meperidine	No match	No-DILI-Concern	0	N02AB02
methamphetamine	No match	No-DILI-Concern	0	N06BA03
mecamylamine	No match	No-DILI-Concern	0	C02BB01
meclizine	No match	No-DILI-Concern	0	R06AE05
oxymorphone	No match	No-DILI-Concern	0	
phendimetrazine	No match	No-DILI-Concern	0	

morphine	No match	No-DILI-Concern	0	N02AA01
dofetilide	No match	No-DILI-Concern	0	C01BD04
guaifenesin	No match	No-DILI-Concern	0	R05CA03
hydroxyzine	No match	No-DILI-Concern	0	N05BB01
midodrine	No match	No-DILI-Concern	0	C01CA17
dexbrompheniramine	No match	No-DILI-Concern	0	R06AB06
cyanocobalamin	No match	No-DILI-Concern	0	B03BA01
phytonadione	No match	No-DILI-Concern	0	B02BA01
darifenacin	No match	No-DILI-Concern	0	G04BD10
vorinostat	No match	No-DILI-Concern	0	L01XX38
midazolam	No match	No-DILI-Concern	0	N05CD08
dihydroergotamine	No match	No-DILI-Concern	0	N02CA01
dichlorphenamide	No match	No-DILI-Concern	0	S01EC02
tranexamic acid	No match	No-DILI-Concern	0	B02AA02
l-carnitine	No match	No-DILI-Concern	0	A16AA01
trientine hydrochloride	No match	No-DILI-Concern	1	
phentolamine	No match	No-DILI-Concern	0	G04BE05 C04AB01
guanethidine	No match	No-DILI-Concern	0	S01EX01 C02CC02
dextroamphetamine	No match	No-DILI-Concern	0	N06BA02
chloramphenicol	No match	No-DILI-Concern	0	S02AA01 D10AF03 J01BA01 D06AX02 G01AA05 S03AA08 S0
orciprenaline	No match	No-DILI-Concern	0	R03CB03 R03AB03
chlorpheniramine	No match	No-DILI-Concern	0	R06AB04
dimercaprol	No match	No-DILI-Concern	0	V03AB09
hydrocortisone	No match	No-DILI-Concern	0	H02AB09 S01CB03 A07EA02 C05AA01 A01AC03 S01BA02 D0
minoxidil	No match	No-DILI-Concern	0	C02DC01 D11AX01
etidronic acid	No match	No-DILI-Concern	0	M05BA01
selegiline	No match	No-DILI-Concern	0	N04BD01
terazosin	No match	No-DILI-Concern	0	G04CA03
rizatriptan	No match	No-DILI-Concern	0	N02CC04
penicillin v	No match	No-DILI-Concern	0	J01CE02
mifepristone	No match	No-DILI-Concern	0	G03XB01
tapentadol	No match	No-DILI-Concern	0	N02AX06
benzphetamine	No match	No-DILI-Concern	0	
flavoxate	No match	No-DILI-Concern	0	G04BD02
mitotane	No match	No-DILI-Concern	0	L01XX23
pentazocine	No match	No-DILI-Concern	0	N02AD01
simethicone	No match	No-DILI-Concern	0	
secretin	No match	No-DILI-Concern	0	V04CK01
sucralfate	No match	No-DILI-Concern	0	A02BX02
gamma hydroxybutyric acid	No match	No-DILI-Concern	0	
lanthanum carbonate	No match	No-DILI-Concern	0	V03AE03
trospium	No match	No-DILI-Concern	0	G04BD09
fesoterodine	No match	No-DILI-Concern	0	G04BD11
neomycin	No match	No-DILI-Concern	0	D06AX04 J01GB05 A01AB08 S01AA03 R02AB01 S03AA01 B0
zaleplon	No match	No-DILI-Concern	0	N05CF03
aminophylline	No match	No-DILI-Concern	0	R03DA05
doxylamine	No match	No-DILI-Concern	0	R06AA09
bethanechol	No match	No-DILI-Concern	0	N07AB02
cortisone acetate	No match	No-DILI-Concern	0	H02AB10 S01BA03
dicyclomine	No match	No-DILI-Concern	0	A03AA07
dyphylline	No match	No-DILI-Concern	0	R03DA01
levorphanol	No match	No-DILI-Concern	0	
propantheline	No match	No-DILI-Concern	0	A03AB05
carbinoxamine	No match	No-DILI-Concern	0	R06AA08
methsuximide	No match	No-DILI-Concern	0	N03AD03
mepenzolate	No match	No-DILI-Concern	0	A03AB12
triprolidine	No match	No-DILI-Concern	0	R06AX07
Sodium polystyrene sulfonate	No match	No-DILI-Concern	0	V03AE01
diethylpropion	No match	No-DILI-Concern	0	A08AA03
biperiden	No match	No-DILI-Concern	0	N04AA02
glycopyrrolate	No match	No-DILI-Concern	0	A03AB02
metyrapone	No match	No-DILI-Concern	0	V04CD01
lactulose	No match	No-DILI-Concern	0	A06AD11
calcitriol	No match	No-DILI-Concern	0	A11CC04 D05AX03

calcium acetate	No match	No-DILI-Concern	0	A12AA12
oxycodone	No match	No-DILI-Concern	0	N02AA05
miglitol	No match	No-DILI-Concern	0	A10BF02
tiludronate	No match	No-DILI-Concern	0	M05BA05
tolterodine	No match	No-DILI-Concern	0	G04BD07
sevelamer	No match	No-DILI-Concern	0	V03AE02
salbutamol	No match	No-DILI-Concern	0	R03AC02 R03CC02
benztropine	No match	No-DILI-Concern	0	N04AC01
miglustat	No match	No-DILI-Concern	0	A16AX06
cabergoline	No match	No-DILI-Concern	0	N04BC06 G02CB03
nitazoxanide	No match	No-DILI-Concern	0	P01AX11
ranolazine	No match	No-DILI-Concern	0	C01EB18
cinacalcet	No match	No-DILI-Concern	0	H05BX01
alvimopan	No match	No-DILI-Concern	0	A06AH02
ramelteon	No match	No-DILI-Concern	0	N05CH02
lubiprostone	No match	No-DILI-Concern	0	A07 A06AX03
paliperidone	No match	No-DILI-Concern	0	N05AX13
tetrahydrobiopterin	No match	No-DILI-Concern	0	A16AX07
codeine	No match	No-DILI-Concern	0	R05DA04
colistimethate	No match	No-DILI-Concern	0	
polymyxin b sulfate	No match	No-DILI-Concern	0	S02AA11 S03AA03 J01XB02 S01AA18 A07AA05
estazolam	No match	No-DILI-Concern	0	N05CD04
riboflavin	No match	No-DILI-Concern	0	A11HA04
levonorgestrel	No match	No-DILI-Concern	0	G03AC03
methylscopolamine bromide	No match	No-DILI-Concern	0	A03BB03 S01FA03
nitroglycerin	No match	No-DILI-Concern	0	C05AE01 C01DA02
ambenonium	No match	No-DILI-Concern	0	N07AA30
ephedrine	No match	No-DILI-Concern	0	R01AB05 S01FB02 R01AA03 R03CA02
pyrantel	No match	No-DILI-Concern	0	P02CC01
quazepam	No match	No-DILI-Concern	0	N05CD10
topotecan	No match	No-DILI-Concern	0	L01XX17
scopolamine	No match	No-DILI-Concern	0	N05CM05 A04AD01 S01FA02
carisoprodol	No match	No-DILI-Concern	0	M03BA02
pamabrom	No match	No-DILI-Concern	0	
bisacodyl	No match	No-DILI-Concern	0	A06AG02 A06AB02
procyclidine	No match	No-DILI-Concern	0	N04AA04
meprobamate	No match	No-DILI-Concern	0	N05BC01
penbutolol	No match	No-DILI-Concern	0	C07AA23
edrophonium	No match	No-DILI-Concern	0	
dopamine	No match	No-DILI-Concern	0	C01CA04
physostigmine	No match	No-DILI-Concern	0	V03AB19 S01EB05
streptomycin	No match	No-DILI-Concern	0	J01GA01 A07AA04
adenosine	No match	No-DILI-Concern	0	C01EB10
dobutamine	No match	No-DILI-Concern	0	C01CA07
flumazenil	No match	No-DILI-Concern	0	V03AB25
lidocaine	No match	No-DILI-Concern	0	D04AB01 S02DA01 N01BB02 C05AD01 C01BB01 R02AD02 S0
bacitracin	No match	No-DILI-Concern	0	D06AX05 R02AB04
l-arginine	No match	No-DILI-Concern	0	B05XB01
acetylcholine chloride	No match	No-DILI-Concern	0	S01EB09
vinblastine	No match	No-DILI-Concern	0	L01CA01
saxagliptin	No match	No-DILI-Concern	0	A10BH03
dexchlorpheniramine	No match	No-DILI-Concern	0	R06AB02
protriptyline	No match	No-DILI-Concern	0	N06AA11
nabilone	No match	No-DILI-Concern	0	A04AD11
almotriptan	No match	No-DILI-Concern	0	N02CC05
frovatriptan	No match	No-DILI-Concern	0	N02CC07
dutasteride	No match	No-DILI-Concern	0	G04CB02
ibandronate	No match	No-DILI-Concern	0	M05BA06
temsirolimus	No match	No-DILI-Concern	0	L01XE09
fludarabine	No match	No-DILI-Concern	0	L01BB05
loratadine	No match	No-DILI-Concern	0	R06AX13
reserpine	No match	No-DILI-Concern	0	C02AA02
trihexyphenidyl	No match	No-DILI-Concern	0	N04AA01
mepivacaine	No match	No-DILI-Concern	0	N01BB03

nalbuphine	No match	No-DILI-Concern	0	N02AF02
sufentanil	No match	No-DILI-Concern	0	N01AH03
cisatracurium besylate	No match	No-DILI-Concern	0	M03AC11
droperidol	No match	No-DILI-Concern	0	N01AX01 N05AD08
alfentanil	No match	No-DILI-Concern	0	N01AH02
dexpanthenol	No match	No-DILI-Concern	0	A11HA30 S01XA12 D03AX03
mannitol	No match	No-DILI-Concern	0	B05CX04 A06AD16 B05BC01
aminohippurate	No match	No-DILI-Concern	0	V04CH30
corticotropin	No match	No-DILI-Concern	0	H01AA01
indocyanine green	No match	No-DILI-Concern	0	
iothalamate sodium	No match	No-DILI-Concern	0	V08AA04
dextrose	No match	No-DILI-Concern	0	B05CX01 V06DC01 V04CA02
sodium lactate	No match	No-DILI-Concern	0	
protirelin	No match	No-DILI-Concern	0	V04CJ02
sincalide	No match	No-DILI-Concern	0	V04CC03
cupric chloride	No match	No-DILI-Concern	0	
ioversol	No match	No-DILI-Concern	0	V08AB07
corticorelin ovine triflutate	No match	No-DILI-Concern	0	V04CD04
sodium thiosulfate	No match	No-DILI-Concern	0	
nesiritide	No match	No-DILI-Concern	0	C01DX19
ganirelix acetate	No match	No-DILI-Concern	0	H01CC01
pramlintide	No match	No-DILI-Concern	0	A10BX05
gadofosveset trisodium	No match	No-DILI-Concern	0	V08CA11
exenatide	No match	No-DILI-Concern	0	A10BX04
dimenhydrinate	No match	No-DILI-Concern	0	R06AA52
doxapram	No match	No-DILI-Concern	0	R07AB01
methylnaltrexone bromide	No match	No-DILI-Concern	0	A06AH01
naloxone	No match	No-DILI-Concern	0	V03AB15
clevidipine	No match	No-DILI-Concern	0	C08CA16
regadenoson	No match	No-DILI-Concern	0	C01EB21
ferumoxytol	No match	No-DILI-Concern	0	
bendamustine	No match	No-DILI-Concern	0	L01AA09
romidepsin	No match	No-DILI-Concern	0	L01XX39
pancrelipase	No match	No-DILI-Concern	0	A09AA02
ecallantide	No match	No-DILI-Concern	0	B06AC03
histamine diphosphate	No match	No-DILI-Concern	0	V04CG03
hyaluronidase	No match	No-DILI-Concern	0	B06AA03
protamine sulfate	No match	No-DILI-Concern	0	
norepinephrine	No match	No-DILI-Concern	0	C01CA03
epinephrine	No match	No-DILI-Concern	0	S01EA01 R01AA14 C01CA24 R03AA01 A01AD01 B02BC09
succinylcholine	No match	No-DILI-Concern	0	M03AB01
thyrotropin alfa	No match	No-DILI-Concern	0	V04CJ01 H01AB01
diatrizoate	No match	No-DILI-Concern	0	V08AA01
oxytocin	No match	No-DILI-Concern	0	H01BB02
cosyntropin	No match	No-DILI-Concern	0	H01AA02
pancuronium	No match	No-DILI-Concern	0	M03AC01
pentagastrin	No match	No-DILI-Concern	0	V04CG04
calcitonin salmon	No match	No-DILI-Concern	0	
pentetic acid	No match	No-DILI-Concern	0	
somatropin recombinant	No match	No-DILI-Concern	0	H01AC01
fluorescein	No match	No-DILI-Concern	0	S01JA01
carboprost tromethamine	No match	No-DILI-Concern	0	G02AD04
vecuronium	No match	No-DILI-Concern	0	M03AC03
atracurium	No match	No-DILI-Concern	0	M03AC04
ethanolamine oleate	No match	No-DILI-Concern	0	C05BB01
rubidium chloride rb-82	No match	No-DILI-Concern	0	
urofollitropin	No match	No-DILI-Concern	0	G03GA04
teriparatide	No match	No-DILI-Concern	0	H05AA02
gadopentetate dimeglumine	No match	No-DILI-Concern	0	V08CA01
pegademase bovine	No match	No-DILI-Concern	0	L03AX04
fenoldopam	No match	No-DILI-Concern	0	C01CA19
mivacurium	No match	No-DILI-Concern	0	M03AC10
gadoteridol	No match	No-DILI-Concern	0	V08CA04
rocuronium	No match	No-DILI-Concern	0	M03AC09

ferumoxides	No match	No-DILI-Concern	0	
porfimer	No match	No-DILI-Concern	0	L01XD01
ibutilide	No match	No-DILI-Concern	0	C01BD05
polyethylene glycol 3350	No match	No-DILI-Concern	0	
eptifibatide	No match	No-DILI-Concern	0	B01AC16
bivalirudin	No match	No-DILI-Concern	0	B01AE06
argatroban	No match	No-DILI-Concern	0	B01AE03
tirofiban	No match	No-DILI-Concern	0	B01AC17
gadoversetamide	No match	No-DILI-Concern	0	V08CA06
perflutren	No match	No-DILI-Concern	0	
amifostine	No match	No-DILI-Concern	0	V03AF05
amikacin	No match	No-DILI-Concern	0	J01GB06 D06AX12 S01AA21
gadobenate dimeglumine	No match	No-DILI-Concern	0	V08CA08
conivaptan	No match	No-DILI-Concern	0	C03XA02
urokinase	No match	No-DILI-Concern	0	B01AD04
benzylpenicilloyl polylysine	No match	No-DILI-Concern	0	
sterile water	No match	No-DILI-Concern	0	
botulinum toxin type a	No match	No-DILI-Concern	0	M03AX01
alteplase	No match	No-DILI-Concern	0	B01AD02 S01XA13
epoetin alfa	No match	No-DILI-Concern	0	B03XA01
sargramostim	No match	No-DILI-Concern	0	L03AA09
abciximab	No match	No-DILI-Concern	0	B01AC13
esmolol	No match	No-DILI-Concern	0	C07AB09
oprelvekin	No match	No-DILI-Concern	0	L03AC02
basiliximab	No match	No-DILI-Concern	0	L04AC02
palivizumab	No match	No-DILI-Concern	0	J06BB16
reteplase	No match	No-DILI-Concern	0	B01AD07
botulinum toxin type b	No match	No-DILI-Concern	0	M03AX01
etomidate	No match	No-DILI-Concern	0	N01AX07
tenecteplase	No match	No-DILI-Concern	0	B01AD11
rasburicase	No match	No-DILI-Concern	0	V03AF07
omalizumab	No match	No-DILI-Concern	0	R03DX05
ibritumomab tiuxetan	No match	No-DILI-Concern	0	V10XX02
pegfilgrastim	No match	No-DILI-Concern	0	L03AA13
bevacizumab	No match	No-DILI-Concern	0	L01XC07
palifermin	No match	No-DILI-Concern	0	V03AF08
galsulfase	No match	No-DILI-Concern	0	A16AB08
alglucosidase alfa	No match	No-DILI-Concern	0	A16AB07
idursulfase	No match	No-DILI-Concern	0	A16AB09
eculizumab	No match	No-DILI-Concern	0	L04AA25
romiplostim	No match	No-DILI-Concern	0	B02BX04
ofatumumab	No match	No-DILI-Concern	0	L01XC10
iodipamide meglumine	No match	No-DILI-Concern	0	V08AC04
l-methionine	No match	No-DILI-Concern	0	V03AB26
vitamin e	No match	No-DILI-Concern	0	A11HA03
apomorphine	No match	No-DILI-Concern	0	N04BC07 G04BE07
norethindrone acetate	No match	No-DILI-Concern	0	
exametazime	No match	No-DILI-Concern	0	
erythropoietin	No match	No-DILI-Concern	0	B03XA01
Argipressin	No match	No-DILI-Concern	0	H01BA06
sodium bicarbonate	No match	No-DILI-Concern	0	B05XA02 B05CB04
plerixafor	No match	No-DILI-Concern	0	L03AX16
tyramine	No match	No-DILI-Concern	0	
Hetastarch	No match	No-DILI-Concern	0	
imiglucerase	No match	No-DILI-Concern	0	A16AB02
agalsidase beta	No match	No-DILI-Concern	0	A16AB04
gadolinium ethoxybenzyl DTPA	No match	No-DILI-Concern	0	V08CA10
Levoleucovorin	No match	No-DILI-Concern	0	
levomefolate calcium	No match	No-DILI-Concern	0	
hyaluronidase recombinant human	No match	No-DILI-Concern	0	
daunorubicin	No match	No-DILI-Concern	0	L01DB02
nystatin	No match	No-DILI-Concern	0	D01AA01 G01AA01 A07AA02
nitroprusside	No match	No-DILI-Concern	0	C02DD01
alemtuzumab	No match	No-DILI-Concern	0	L01XC04

panitumumab	No match	No-DILI-Concern	0	L01XC08
rilonacept	No match	No-DILI-Concern	0	L04AC04
ustekinumab	No match	No-DILI-Concern	0	L04AC05

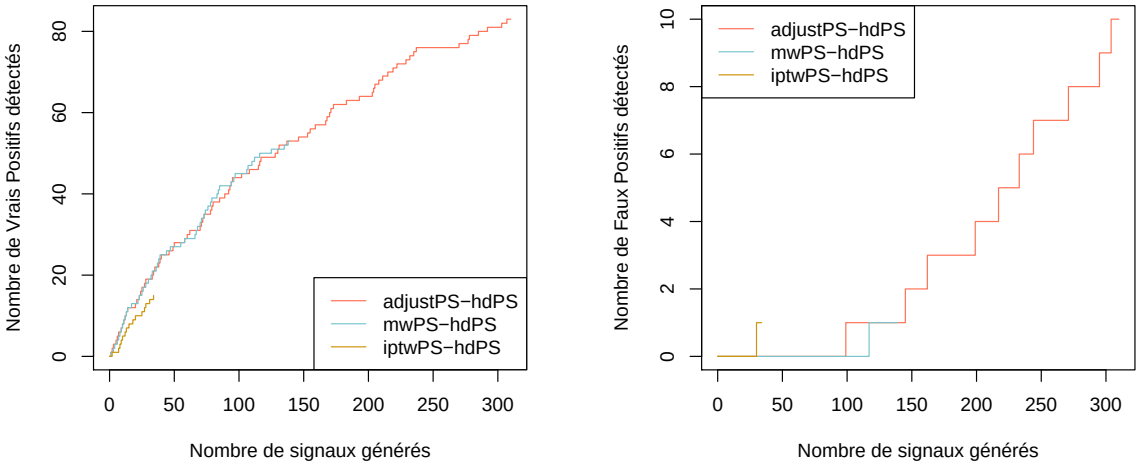
Tableau A.2 – Codes PT associés aux mot-clés définis par Chen *et al.* (2011) pour définir un DILI, après traduction de ces mot-clés en requêtes standardisées du dictionnaire MedDRA par l'équipe Inserm 1219.

Standardised MedDRA Queries	Preferred Term	Code Preferred Term
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Acute hepatic failure	10000804
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Acute on chronic liver failure	10077305
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Acute yellow liver atrophy	10070815
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Cholestatic liver injury	10067969
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Coma hepatic	10010075
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Drug-induced liver injury	10072268
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatectomy	10061997
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic atrophy	10019637
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic encephalopathy	10019660
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic encephalopathy prophylaxis	10066599
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic failure	10019663
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic infiltration eosinophilic	10064668
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic lesion	10061998
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic necrosis	10019692
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic steato-fibrosis	10077215
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatic steatosis	10019708
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatitis fulminant	10019772
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatobiliary disease	10062000
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatocellular foamy cell syndrome	10053244
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatocellular injury	10019837
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatorenal failure	10019845
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatorenal syndrome	10019846
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Hepatotoxicity	10019851
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Liver dialysis	10076640
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Liver disorder	10024670
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Liver injury	10067125
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Liver transplant	10024714
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Minimal hepatic encephalopathy	10076204
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Mixed liver injury	10066758
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Non-alcoholic fatty liver	10029530
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Non-alcoholic steatohepatitis	10053219
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Reye's syndrome	10039012
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Steatohepatitis	10076331
Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ)	Subacute hepatic failure	10056956
Hepatitis, non-infectious (SMQ)	Allergic hepatitis	10071198
Hepatitis, non-infectious (SMQ)	Autoimmune hepatitis	10003827
Hepatitis, non-infectious (SMQ)	Hepatitis	10019717
Hepatitis, non-infectious (SMQ)	Hepatitis acute	10019727
Hepatitis, non-infectious (SMQ)	Hepatitis cholestatic	10019754
Hepatitis, non-infectious (SMQ)	Hepatitis fulminant	10019772
Hepatitis, non-infectious (SMQ)	Hepatitis toxic	10019795
Hepatitis, non-infectious (SMQ)	Non-alcoholic steatohepatitis	10053219
Hepatitis, non-infectious (SMQ)	Steatohepatitis	10076331
Liver related investigations, signs and symptoms (SMQ)	Alanine aminotransferase abnormal	10001547
Liver related investigations, signs and symptoms (SMQ)	Alanine aminotransferase increased	10001551
Liver related investigations, signs and symptoms (SMQ)	Ammonia abnormal	10001942
Liver related investigations, signs and symptoms (SMQ)	Ammonia increased	10001946
Liver related investigations, signs and symptoms (SMQ)	Aspartate aminotransferase abnormal	10003477
Liver related investigations, signs and symptoms (SMQ)	Aspartate aminotransferase increased	10003481
Liver related investigations, signs and symptoms (SMQ)	Bile output abnormal	10051344
Liver related investigations, signs and symptoms (SMQ)	Bile output decreased	10051343
Liver related investigations, signs and symptoms (SMQ)	Bilirubin conjugated abnormal	10067718
Liver related investigations, signs and symptoms (SMQ)	Bilirubin conjugated increased	10004685
Liver related investigations, signs and symptoms (SMQ)	Bilirubin urine present	10077356
Liver related investigations, signs and symptoms (SMQ)	Biopsy liver abnormal	10004792

Liver related investigations, signs and symptoms (SMQ)	Blood bilirubin abnormal	10058477
Liver related investigations, signs and symptoms (SMQ)	Blood bilirubin increased	10005364
Liver related investigations, signs and symptoms (SMQ)	Blood bilirubin unconjugated increased	10005370
Liver related investigations, signs and symptoms (SMQ)	Bromosulphthalein test abnormal	10006408
Liver related investigations, signs and symptoms (SMQ)	Computerised tomogram liver abnormal	10078360
Liver related investigations, signs and symptoms (SMQ)	Foetor hepaticus	10052554
Liver related investigations, signs and symptoms (SMQ)	Galactose elimination capacity test abnormal	10059710
Liver related investigations, signs and symptoms (SMQ)	Galactose elimination capacity test decreased	10059712
Liver related investigations, signs and symptoms (SMQ)	Gamma-glutamyltransferase abnormal	10017688
Liver related investigations, signs and symptoms (SMQ)	Gamma-glutamyltransferase increased	10017693
Liver related investigations, signs and symptoms (SMQ)	Hepatic enzyme abnormal	10062685
Liver related investigations, signs and symptoms (SMQ)	Hepatic enzyme increased	10060795
Liver related investigations, signs and symptoms (SMQ)	Hepatic function abnormal	10019670
Liver related investigations, signs and symptoms (SMQ)	Hepatic hypertrophy	10076254
Liver related investigations, signs and symptoms (SMQ)	Hepatic pain	10019705
Liver related investigations, signs and symptoms (SMQ)	Hepatobiliary scan abnormal	10066195
Liver related investigations, signs and symptoms (SMQ)	Hepatomegaly	10019842
Liver related investigations, signs and symptoms (SMQ)	Hepatosplenomegaly	10019847
Liver related investigations, signs and symptoms (SMQ)	Hyperammonaemia	10020575
Liver related investigations, signs and symptoms (SMQ)	Hyperbilirubinaemia	10020578
Liver related investigations, signs and symptoms (SMQ)	Hypertransaminasaemia	10068237
Liver related investigations, signs and symptoms (SMQ)	Liver function test abnormal	10024690
Liver related investigations, signs and symptoms (SMQ)	Liver scan abnormal	10061947
Liver related investigations, signs and symptoms (SMQ)	Liver tenderness	10024712
Liver related investigations, signs and symptoms (SMQ)	Perihepatic discomfort	10054125
Liver related investigations, signs and symptoms (SMQ)	Total bile acids increased	10064558
Liver related investigations, signs and symptoms (SMQ)	Transaminases abnormal	10062688
Liver related investigations, signs and symptoms (SMQ)	Transaminases increased	10054889
Liver related investigations, signs and symptoms (SMQ)	Ultrasound liver abnormal	10045428
Liver related investigations, signs and symptoms (SMQ)	Urine bilirubin increased	10050792
Cholestasis and jaundice of hepatic origin (SMQ)	Bilirubin excretion disorder	10061009
Cholestasis and jaundice of hepatic origin (SMQ)	Cholaemia	10048611
Cholestasis and jaundice of hepatic origin (SMQ)	Cholestasis	10008635
Cholestasis and jaundice of hepatic origin (SMQ)	Cholestatic liver injury	10067969
Cholestasis and jaundice of hepatic origin (SMQ)	Cholestatic pruritus	10064190
Cholestasis and jaundice of hepatic origin (SMQ)	Drug-induced liver injury	10072268
Cholestasis and jaundice of hepatic origin (SMQ)	Hepatitis cholestatic	10019754
Cholestasis and jaundice of hepatic origin (SMQ)	Hyperbilirubinaemia	10020578
Cholestasis and jaundice of hepatic origin (SMQ)	Icterus index increased	10021209
Cholestasis and jaundice of hepatic origin (SMQ)	Jaundice	10023126
Cholestasis and jaundice of hepatic origin (SMQ)	Jaundice cholestatic	10023129
Cholestasis and jaundice of hepatic origin (SMQ)	Jaundice hepatocellular	10023136
Cholestasis and jaundice of hepatic origin (SMQ)	Mixed liver injury	10066758
Cholestasis and jaundice of hepatic origin (SMQ)	Ocular icterus	10058117

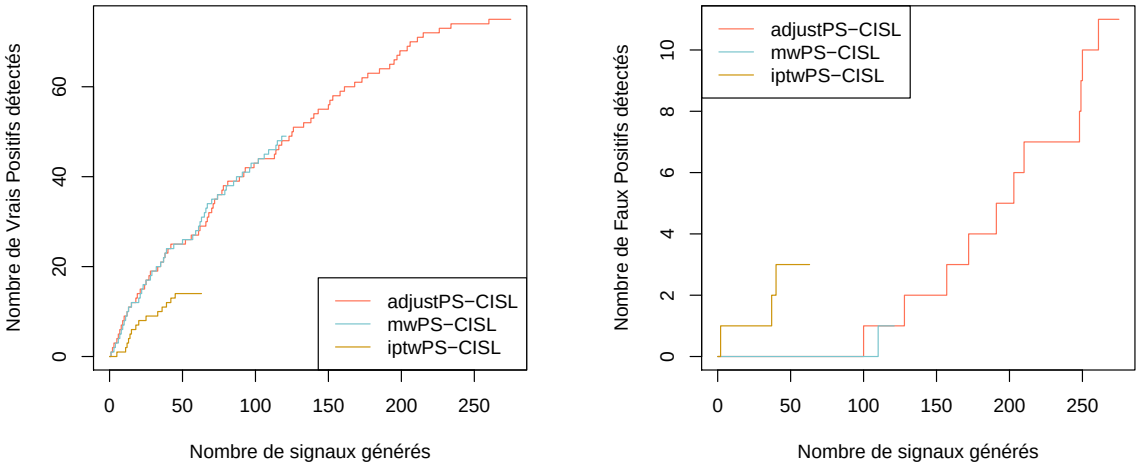
Annexe B

Résultats complémentaires pour l'utilisation des scores de propension en grande dimension pour la détection automatisée concernant l'évènement indésirable *drug-induced liver injury* dans la BNPV



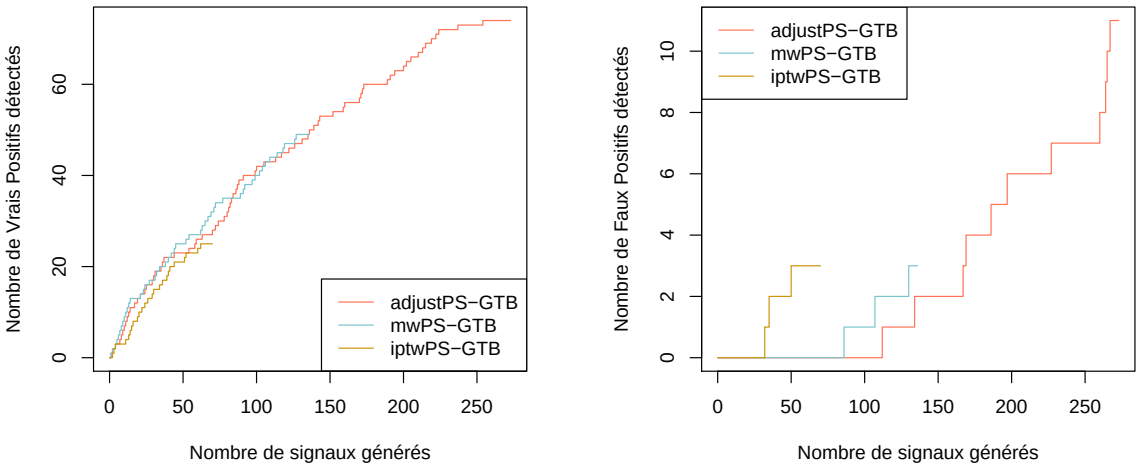
(a) (b)

FIGURE B.1 – Nombre de (a) vrais positifs et (b) faux positifs détectés selon le nombre de signaux générés par les approches basées sur le score de propension : ajustement, pondération avec les poids *matching weights*, pondération *inverse probability of treatment weighting* où les scores sont estimés avec hdPS : adjustPS-hdPS, mwPS-hdPS, iptwPS-hdPS.
Pour toutes ces approches, les signaux sont ordonnés selon les p-valeurs corrigées pour la multiplicité des tests.



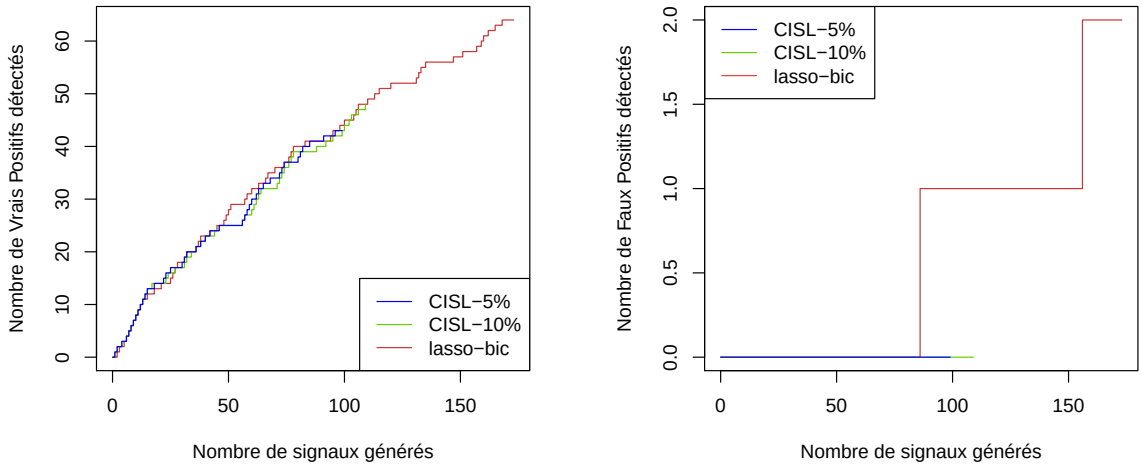
(a) (b)

FIGURE B.2 – Nombre de (a) vrais positifs et (b) faux positifs détectés selon le nombre de signaux générés par les approches basées sur le score de propension : ajustement, pondération avec les poids *matching weights*, pondération *inverse probability of treatment weighting* où les scores sont estimés avec CISL : adjustPS-CISL, mwPS-CISL, iptwPS-CISL.
Pour toutes ces approches, les signaux sont ordonnés selon les p-valeurs corrigées pour la multiplicité des tests.



(a) (b)

FIGURE B.3 – Nombre de (a) vrais positifs et (b) faux positifs détectés selon le nombre de signaux générés par les approches basées sur le score de propension : ajustement, pondération avec les poids *matching weights*, pondération *inverse probability of treatment weighting* où les scores sont estimés avec GTB : adjustPS-GTB, mwPS-GTB, iptwPS-GTB.
Pour toutes ces approches, les signaux sont ordonnés selon les p-valeurs corrigées pour la multiplicité des tests.



(a) (b)

FIGURE B.4 – Nombre de (a) vrais positifs et (b) faux positifs détectés selon le nombre de signaux générés par les approches basées sur des régression multiples pénalisées lasso : lasso-bic, CISL-5% et CISL-10%.
Pour lasso-bic, les signaux sont ordonnées selon leur p-valeur dans le modèle de régression non pénalisé. Pour CISL-5% et CISL-10% les signaux sont ordonnées selon leur valeur du quantile à 5% ou 10% de la distribution obtenue avec CISL.

Annexe C

Résultats complémentaires du travail de détection automatisée concernant l'évènement indésirable *drug-induced liver injury* dans la base de l'EGB

Approches	Nombre de signaux	Nombre de signaux à statut connu	Vrais positifs			Faux positifs		
			n	PPV (%)	Sensibilité (%)	n	FDP (%)	Spécificité (%)
Paramétrage 1 (211, 13, 20)								
cco-univ	0	0	0	NA	0,00	0	NA	100,00
cco-aic	31	4	1	25,00	7,69	3	75,00	85,00
cco-bic	1	0	0	NA	0,00	0	NA	100,00
Paramétrage 2 (297, 16, 26)								
cco-univ	17	5	1	20,00	6,25	4	80,00	84,62
cco-aic	41	7	1	14,29	6,25	6	85,71	76,92
cco-bic	10	3	1	33,33	6,25	2	66,67	92,31
Paramétrage 3 (211, 12, 20)								
cco-univ	1	0	0	NA	0,00	0	NA	100,00
cco-aic	30	5	1	20,00	8,33	4	80,00	80,00
cco-bic	4	1	1	100,00	8,33	0	0,00	100,00
Paramétrage 4 (293, 14, 25)								
cco-univ	10	2	1	50,00	7,14	1	50,00	96,00
cco-aic	35	6	1	16,67	7,14	5	83,33	80,00
cco-bic	10	2	1	50,00	7,14	1	50,00	96,00
Paramétrage 5 (271, 13, 24)								
cco-univ	1	0	0	NA	0,00	0	NA	100,00
cco-aic	55	11	4	36,36	30,77	7	63,64	70,83
cco-bic	10	1	1	100,00	7,69	0	0,00	100,00
Paramétrage 6 (362, 19, 29)								
cco-univ	40	8	2	25,00	10,53	6	75,00	79,31
cco-aic	62	11	2	18,18	10,53	9	81,82	68,97
cco-bic	22	3	1	33,33	5,26	2	66,67	93,10
Paramétrage 7 (310, 17, 26)								
cco-univ	3	1	1	100,00	5,88	0	0,00	100,00
cco-aic	68	9	1	11,11	5,88	8	88,89	69,23
cco-bic	12	2	1	50,00	5,88	1	50,00	96,15
Paramétrage 8 (395, 21, 30)								
cco-univ	60	11	2	18,18	9,52	9	81,82	70,00
cco-aic	78	13	3	23,08	14,29	10	76,92	66,67
cco-bic	22	5	1	20,00	4,76	4	80,00	86,67

Tableau C.1 – Résultats de la détection de signaux pour les approches basées sur le *case-crossover* pour tous les paramétrages. Pour chaque paramétrage, il est précisé : le nombre de médicaments considéré induit par la mise en forme des données, le nombre de témoins positifs présents, le nombre de témoins négatifs présents.
PPV : Valeur Prédictive Positive
FDP : Proportion de Fausses Découvertes

Paramétrage	Nombre de fenêtres d'observation	Nombre moyen de médicaments par fenêtres	Nombre moyen de séquences	Nombre de médicaments différents	Nombre de témoins positifs	Nombre de témoins négatifs
m3-no	28	13,86 (5,72)	3,99 (1,46)	88	2	8
m3-o	82	13,99 (5,85)	4,01 (1,44)	117	2	10
m6-no	14	70,21 (12,35)	5,63 (3,35)	204	6	20
m6-o	79	68,37 (10,40)	5,61 (3,24)	252	13	26
m12-no	7	200,86 (14,10)	9,34 (8,51)	344	17	30
m12-o	73	195,11 (12,60)	9,22 (8,25)	403	19	36

Tableau C.2 – Nombre de fenêtres d'observation selon les différentes paramétrages de la *sequence symmetry analysis*. Effectifs moyen et écart type des médicaments par fenêtre, effectifs moyen et écart type des séquences sur l'ensemble de la période d'observation et effectifs de témoins négatifs et témoins positifs selon DILIrank pour chaque paramétrage.

Approches	Nombre de signaux	Nombre de signaux à statut connu	Vrais positifs			Faux positifs		
			n	PPV (%)	Sensibilité (%)	n	FDP (%)	Spécificité (%)
Paramétrage m3-no (88, 2, 8)								
IC exact	0	0	0	NA	0,00	0	NA	100,00
IC *bootstrap*	9	0	0	NA	0,00	0	NA	100,00
Paramétrage m3-o (117, 2, 10)								
IC exact	0	0	0	NA	0,00	0	NA	100,00
IC *bootstrap*	17	1	1	100,00	50,00	0	0,00	100,00
Paramétrage m6-no (204, 6, 20)								
IC exact	0	0	0	NA	0,00	0	NA	100,00
IC *bootstrap*	20	6	2	33,33	33,33	4	66,67	80,00
Paramétrage m6-o (252, 13, 26)								
IC exact	0	0	0	NA	0,00	0	NA	100,00
IC *bootstrap*	68	8	4	50,00	30,77	4	50,00	84,62
Paramétrage m12-no (344, 17, 30)								
IC exact	0	0	0	NA	0,00	0	NA	100,00
IC *bootstrap*	27	5	2	40,00	11,76	3	60,00	90,00
Paramétrage m12-o (403, 19, 36)								
IC exact	1	0	0	NA	0,00	0	NA	100,00
IC *bootstrap*	121	16	5	31,25	26,32	11	68,75	69,44

TABLEAU C.3 – Résultats de la détection de signaux pour les approches basées sur la *sequence symmetry analysis* pour tous les paramétrages. Pour chaque paramétrage, il est précisé : le nombre de médicaments considéré induit par la mise en forme des données, le nombre de témoins positifs présents, le nombre de témoins négatifs présents.
PPV : Valeur Prédictive Positive
FDP : Proportion de Fausses Découvertes

Annexe D

Publications et soumissions

D.1 Article publié dans la revue *Frontiers in Pharmacology*

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 18 September 2018
doi: 10.3389/fphar.2018.01010

Propensity Score-Based Approaches in High Dimension for Pharmacovigilance Signal Detection: an Empirical Comparison on the French Spontaneous Reporting Database

Émeline Courtois[1], Antoine Pariente[2], Francesco Salvo[2], Étienne Volatier[1], Pascale Tubert-Bitter[1†] and Ismaïl Ahmed[1†]*

[1] Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases, INSERM, UVSQ (Université Paris-Saclay), Institut Pasteur, Villejuif, France, [2] Bordeaux Population Health Research Center, Pharmacoepidemiology Team (UMR 1219), INSERM, University of Bordeaux, Bordeaux, France

OPEN ACCESS

Edited by:
Marie-Christine Jaulent, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Robert L. Lins, Retired, Antwerpen, Belgium
Juergen Landes, Ludwig-Maximilians-Universität München, Germany

***Correspondence:**
Émeline Courtois
emeline.courtois@inserm.fr

† These authors share last authorship

Specialty section:
This article was submitted to Pharmaceutical Medicine and Outcomes Research, a section of the journal Frontiers in Pharmacology

Received: *29 November 2017*
Accepted: *20 August 2018*
Published: *18 September 2018*

Citation:
Courtois É, Pariente A, Salvo F, Volatier É, Tubert-Bitter P and Ahmed I (2018) Propensity Score-Based Approaches in High Dimension for Pharmacovigilance Signal Detection: an Empirical Comparison on the French Spontaneous Reporting Database. Front. Pharmacol. 9:1010. doi: 10.3389/fphar.2018.01010

Classical methods used for signal detection in pharmacovigilance rely on disproportionality analysis of counts aggregating spontaneous reports of a given adverse drug reaction. In recent years, alternative methods have been proposed to analyze individual spontaneous reports such as penalized multiple logistic regression approaches. These approaches address some well-known biases resulting from disproportionality methods. However, while penalization accounts for computational constraints due to high-dimensional data, it raises the issue of determining the regularization parameter and eventually that of an error-controlling decision rule. We present a new automated signal detection strategy for pharmacovigilance systems, based on propensity scores (PS) in high dimension. PSs are increasingly used to assess a given association with high-dimensional observational healthcare databases in accounting for confusion bias. Our main aim was to develop a method having the same advantages as multiple regression approaches in dealing with bias, while relying on the statistical multiple comparison framework as regards decision thresholds, by considering false discovery rate (FDR)-based decision rules. We investigate four PS estimation methods in high dimension: a gradient tree boosting (GTB) algorithm from machine-learning and three variable selection algorithms. For each (drug, adverse event) pair, the PS is then applied as adjustment covariate or by using two kinds of weighting: inverse proportional treatment weighting and matching weights. The different versions of the new approach were compared to a univariate approach, which is a disproportionality method, and to two penalized multiple logistic regression approaches, directly applied on spontaneous reporting data. Performance was assessed through an empirical comparative study conducted on a reference signal set in the French national pharmacovigilance database (2000–2016) that was recently proposed for drug-induced liver injury. Multiple regression approaches performed better in detecting true positives and false positives. Nonetheless, the performances of the PS-based methods using

matching weights was very similar to that of multiple regression and better than with the univariate approach. In addition to being able to control FDR statistical errors, the proposed PS-based strategy is an interesting alternative to multiple regression approaches.

Keywords: pharmacovigilance, signal detection, propensity score in high dimension, spontaneous reports, penalized multiple regression, FDR

1. INTRODUCTION

Once a drug is introduced on the market, many people are exposed to it in real-life conditions, which can be very different from those evaluated in clinical trials. The goal of pharmacovigilance is to detect as quickly as possible potential adverse reactions which could be induced by drug exposure. To achieve this challenging task, health authorities collect and monitor spontaneous reports of suspected adverse events (AEs), mainly from practitioners. In France, such a pharmacovigilance database is maintained by the National Agency for the Safety of Drugs and Health Products (*Agence Nationale de Sécurité du Médicament et des Produits de Santé*, ANSM). It contained around 431,000 reports at the end of July 2016. In recent years, about 36,000 reports have been reported annually.

To exploit this amount of data, several automated signal detection tools have been developed. The term signal refers to a (drug, AE) pair whose association is highlighted by a signal detection method. The aim of such methods is to draw attention to unexpected associations by acting as hypothesis generators. The signals thus generated must be further investigated by pharmacovigilance experts in order to draw definite conclusions. The most common methods used by health agencies are disproportionality methods (Almenoff et al., 2007). These methods rely on a three-step strategy: for each (drug, AE) pair (1) determine the number of reports involving this specific pair; (2) construct a measure of the degree to which the observed count of reports exceeds the expected count if independence applies (the disproportionality measure); and (3) decide that a signal has occurred if this measure exceeds a specified threshold value. The procedure is conducted for all (drug, AE) pairs simultaneously (Ahmed et al., 2015). More recently, these disproportionality methods were extended by integrating a false discovery rate (FDR) estimation procedure in order to take into account comparison multiplicity (Benjamini and Hochberg, 1995; Ahmed et al., 2010).

To address the shortcomings of disproportionality methods in dealing with confounding like masking effect and co-prescription bias (Harpaz et al., 2012), multiple logistic regression approaches have been proposed. While being more computationally intensive, they have been shown through empirical studies to be promising signal detection methods (Caster et al., 2010; Harpaz et al., 2013; Marbac et al., 2016; Ahmed et al., 2018). Instead of considering aggregated data as disproportionality methods do, these newer approaches analyse spontaneous reports directly: the observation becomes the individual report, the outcome is the presence/absence of a given AE, and the covariates are all drug presence indicators (Caster et al., 2010). Because of the very large number of potential covariates, a lasso version of the logistic regression has been used (Tibshirani, 1996). It consists in maximizing the log-likelihood of a classical multiple logistic regression model, penalized by a function proportional to the L_1 norm of the regression coefficients. This penalty makes it possible to shrink some coefficients associated with the covariates to exactly zero. Cross validation is commonly used to choose the penalty value and it achieves good performances in terms of prediction. However, determining the best penalization parameter in the variable selection framework is a challenging task.

Another way to deal with confounding bias issues is to summarize all the information into a propensity score (PS). PSs are classically used in epidemiological studies investigating one targeted association (i.e., one drug, one AE) to deal with confounding like indication bias. In recent years, this methodology has been used with high-dimensional health-care databases, like claims data, by collapsing all the numerous covariates present in these databases into a single value which is easier to manipulate (Seeger et al., 2005; Schneeweiss et al., 2009). Tatonetti et al. (2012) investigated the use of one PS strategy in the context of spontaneous reporting data and illustrated its interest by comparing it to a disproportionality method.

In this paper, we develop several PS-based approaches for signal detection from spontaneous reporting and compare them to penalized multiple regression approaches. We built the PSs with four different algorithms adapted to high dimensional settings and for each of them, we considered three different integration strategies to perform signal detection. Our proposed PS-based methods include a signal detection rule relying on FDR estimation. This comparison is performed with the French national pharmacovigilance database (2000–2016) using a large and recently published reference set pertaining to a common adverse reaction: drug-induced liver injury (DILI) (Chen et al., 2011, 2016).

2. METHODS

2.1. Reference Methods

The most common proposed methods for signal detection are disproportionality methods, which rely on the aggregated form of the data. These methods aim to detect higher-than-expected combinations of drugs and events in the database. Nevertheless, two kinds of biases are created by the univariate feature of disproportionality methods: masking and the co-reporting confounding effect (Harpaz et al., 2012). In order to address these two bias issues, other approaches in pharmacovigilance have

lately emerged that rely on multiple logistic regressions (Caster et al., 2010). Individual spontaneous reports are directly analyzed instead of aggregated counts of drug-event combinations. We denote by I the total number of drugs and J the total number of AEs in the database. Let Y_j indicate the presence or absence of AE_j, and X_i denote the presence or absence of $drug_i$. The corresponding logistic model for each AE_j is

$$\text{logit}(\Pr(Y_j = 1)) = \beta_0^j + \sum_{i=1}^{I} \beta_i^j X_i. \quad (1)$$

In spontaneous reporting databases, the number of drug covariates I is very large. To deal with this high-dimensional issue and in order to derive the most parsimonious model, a penalized lasso logistic regression was implemented (Tibshirani, 1996). This consists in maximizing the log-likelihood of model (1) minus a penalization term

$$pen_j(\lambda) = \lambda\, |\beta^j|_1 = \lambda \sum_{i=1}^{I} |\beta_i^j|. \quad (2)$$

Depending on the penalization parameter λ, the lasso regression shrinks most of the coefficients associated to the covariates to exactly zero, so these covariates are not retained in the model. Usually cross validation is used to select the best value of λ in terms of prediction error, but this does not achieve good performance in a variable selection framework which is that of signal detection. Here, we considered two alternative strategies to this penalization parameter selection: the BIC-Lasso and the class-imbalance subsampling lasso (CISL) (Ahmed et al., 2018).

2.1.1. BIC-Lasso

This method uses the Bayesian Information Criterion. First, a lasso regression is computed for a predefined set of penalization parameter values λ_ks. To each tested λ_k is associated a set of variables selected by the lasso regression. The BIC is then calculated from a classical logistic regression model for each of these sets:

$$\text{BIC} = -2\ln(L) + k\ln(N) \quad (3)$$

where L is the likelihood of the regression model, N the number of observations, and k the number of parameters in the model. We declared as signals all drugs positively associated with the outcome in the model that minimizes the BIC.

2.1.2. CISL

Another way to get around the penalization parameter selection issue is the stability selection method (Meinshausen and Bühlmann, 2010). To take into account the sparsity of pharmacovigilance databases with low frequencies of the outcomes, Ahmed et al. (2018) proposed a variation of this method: the CISL algorithm. In this method, samples are drawn following an nonequiprobable sampling scheme with replacement to allow a better representation of individuals who experienced the outcome of interest. For a given set of penalization parameter values, lasso logistic regressions are computed in each of these samples. The CISL procedure consists in computing the quantity:

$$\widehat{\pi}_i^b = \frac{1}{E} \sum_{\eta=1}^{E} \mathbf{1}[\widehat{\beta}_i^{\eta,b} > 0], \quad (4)$$

where E is the maximum value of covariates selected by all the lasso regressions, $\eta \in \{1, .., E\}$ is the number of predictors selected and $\widehat{\beta}_i^{\eta,b}$ is the regression coefficient estimated by the logistic lasso for drug i, on sample $b \in \{1, .., B\}$, for a model including η covariates. An empirical distribution of $\widehat{\pi}_i^b$ for each drug is obtained over all B samples. A covariate is finally selected by the CISL method if a given quantile of the distribution of $\widehat{\pi}_i^b$ is non-zero. In this work, we considered the covariate sets established with the 5% and the 10% quantiles of these distributions.

In the following we refer to these two methods as the multiple regression approaches.

2.2. Propensity Score Approaches

The PS is defined as the probability of being exposed to a treatment given the observed covariates (Rosenbaum and Rubin, 1983). Conditionally on the PS, treatment exposure and the observed covariates are independent: patients with similar PSs will have on average similar covariate distributions between the exposed and unexposed subjects. This relies on the strong assumption that there are no unmeasured confounders. This means that all variables that affect both treatment assignment and outcome, thus potentially inducing spurious associations, have been measured. The PS is used in the context of observational studies as a balancing score that allows for non-randomized trials to reproduce the conditions of a randomized experiment by making observed baseline characteristics comparable in two different treatment groups. Thus, it addresses confounding biases like indication bias. In this framework, the PS is unknown and has to be estimated from observed data, usually using a logistic regression. Each patient is then assigned a predicted probability of being exposed that is calculated from the PS regression.

2.2.1. Propensity Score Modeling

It is recommended to include predictors and confounders in the PS model i.e., covariates that are related to the outcome or to the outcome and the exposure. It is also strongly advised to avoid instrumental variables i.e., variables that are only related to the exposure (Patorno et al., 2013). In practice, it is sometimes recommended to make *a priori* variable selection according to expert knowledge (Brookhart et al., 2006). In the high-dimensional framework such as health-care databases where thousands of covariates are entered, this approach cannot hold. The high-dimensional propensity score (hdPS) algorithm was proposed to deal with this kind of difficulty (Schneeweiss et al., 2009). For a given exposure, the choice of candidate covariates to include in the PS model relies on empirical assessment of covariates prevalence and strength of association with the exposure and with the outcome of interest. The algorithm ranks all input variables by their potential for confounding and the

investigator must choose the top-ranked variables to be included in the model for estimating the PS.

In the present work we aimed at computing the PS for each drug in the database by selecting among all the other drugs those to be included in the PS estimation model. We implemented four PS-estimation methods. We considered three variable selection algorithms: hdPS and the two other variable selection methods described above: the BIC-Lasso and the CISL [with the positive constraint for β replaced by a non-zero constraint in Equation (4)]. Using the BIC-lasso methodology, each covariate associated with the smallest BIC models was selected. Using CISL, all covariates with a 10% percentile of its distribution greater than zero were selected. Using hdPS, the 20 top-ranked variables were selected. For the sake of completeness, we also implemented a PS-estimation method which relies on a machine-learning algorithm, as was recently proposed by Lee et al. (2010): we considered a gradient tree boosting (GTB) algorithm (Hastie et al., 2009).

2.2.2. Use of the Propensity Score

Once the PS is estimated, four methodologies may be used to remove confounding in estimating the treatment effect: (1) adjustment on the PS; (2) stratification on the PS; (3) matching on the PS; (4) weighting on the PS. Adjustment on the PS consists in considering the estimated PS as a covariate and including it as an additional variable, with the exposure, in the regression model on the outcome. Stratification on the PS is based on a population stratification according to the quantiles of the PS distribution. The treatment effect is then estimated in each subgroup of patients. Matching on the PS consists in matching one treated subject to one (or more) untreated subject which have similar values of their estimated PS. Classically, subjects are matched with a nearest neighbor matching algorithm within a specified caliper distance equal to 0.2 of the standard deviation of the estimated PS. Weighting on the PS amounts to assigning to each subject a weight that is calculated from the estimated PS. The treatment effect is estimated on the pseudo population which is built according to these weights (Austin, 2011).

In this study, we considered two of these four different PS methodologies. First we implemented an adjustment on the PS. For every (AE_j, $drug_i$) pair of interest, we computed the following logistic regression model:

$$\text{logit}(\Pr(Y_j = 1)) = \beta_0^j + \beta_i^j X_i + \widetilde{\beta}_i^j \widehat{e}_i \tag{5}$$

where $\widehat{e}_i$ is the estimated PS of $drug_i$ for all the observations.

We also implemented the inverse probability of treatment weighting (IPTW, Austin, 2011). For a given $drug_i$, the weights to be attributed to each observation are defined by

$$w_i^{IPTW} = \frac{X_i}{\widehat{e}_i} + \frac{1 - X_i}{1 - \widehat{e}_i}. \tag{6}$$

Because some drugs have been reported very rarely, and do not have many reports in common with AEs, matching subjects on PS may cause dramatic losses in terms of subjects who experienced both the exposure and the outcome. Keeping this constraint in mind, we performed another kind of weighting to mimic the classical matching by targeting the same estimand as pair-matching on the PS. The proposed weights are referred to the matching weights (MW) (Li and Greene, 2013; Franklin et al., 2017).

$$w_i^{MW} = \frac{\min(\widehat{e}_i, 1 - \widehat{e}_i)}{X_i \widehat{e}_i + (1 - X_i)(1 - \widehat{e}_i)}. \tag{7}$$

For these two weighting approaches, we then computed univariate weighted logistic regressions by considering w_i^{IPTW} or w_i^{MW}.

For all these PS-based approaches, the signal detection rule is the one applied in the hypothesis testing framework. To take the multiplicity of the tested hypotheses into account, we applied an FDR controlling procedure adapted to the one-sided null hypothesis setting that stands in pharmacovigilance (Ahmed et al., 2010). An FDR adjusted p-value is estimated for each comparison tested. A signal is considered to occur if this p-value corrected for multiplicity is below an FDR threshold.

3. MATERIAL AND COMPARISON SETTINGS

To assess the performances of our PS-based approaches, we compared their ability to correctly classify true and false signals to multiple regression approaches presented above and a univariate approach, using the French pharmacovigilance database on a recently proposed reference set pertaining to DILI adverse events.

3.1. Comparison Set-Up

We compared the ability of each method to detect true signals and not to detect false signals. For the BIC-Lasso approach, we declare as signals all drugs positively associated with the outcome in the model that minimizes the BIC. All drugs selected by CISL are considered as a signal since the positive association with the outcome is taken into account in the algorithm. For all PS-based approaches, a 5% level FDR controlling procedure was used to determine which drug-AE associations are considered as signals. We also included a method based on a simple univariate logistic regression: for each AE_j, $j \in \{1, .., J\}$ and each $drug_i$, $i \in \{1, .., I\}$ we computed:

$$\text{logit}(\Pr(Y_j = 1)) = \beta_0^j + \beta_i^j X_i. \tag{8}$$

This method was implemented to serve as a reference level. It can be assimilated to a reporting odds ratio method, which is a disproportionality method for signal detection (van Puijenbroek et al., 2002). We applied the same FDR controlling procedure to this univariate approach.

All the analyses were performed with R version 3.4.0 (R Core Team, 2017). All the logistic regressions were computed with the speedglm R package v0.3–2. All lasso regressions were implemented using the glmnet R package v2.0–10. GTB algorithms for PS estimation were implemented with the xgb.train function from the xgboost R package 0.6–4. The default values of the xgb.train function were used, except for the learning rate, which was fixed at 0.1.

Table 1 summarizes the main advantages and disadvantages of the different signal detection methods presented in this section. For the sake of clarity, we refer to the multiple regression approaches in the following as BIC-Lasso, CISL-5%, CISL-10% and to the univariate approach as Univ. The terms adjustPS-BIC, mwPS-BIC, iptwPS-BIC, adjustPS-CISL, mwPS-CISL, iptwPS-CISL, adjustPS-GTB, mwPS-GTB, iptwPS-GTB, and adjustPS-hdPS, mwPS-hdPS, iptwPS-hdPS refer to the PS approaches implemented, the way PS was estimated (BIC, CISL, GTB or hdPS) and how PS was taken into account (adjustment, MW, or IPTW).

3.2. The French Pharmacovigilance Database

The performances of our approaches were evaluated and compared by an empirical analysis using data from the French pharmacovigilance database. Drugs are coded according to the 5th level of the Anatomical Therapeutic Chemical (ATC) hierarchy, and the AEs to the Preferred Term (PT) level of the Medical Dictionary for Regulatory Activities (MedDRA).

We applied some filters to the data and considered (i) only products which are known to be a drug (leaving out vaccines, phytotherapy, homeotherapy, dietary supplement, oligotherapy and enzyme inhibitor); (ii) reactions which are targeted as AEs (overdoses or medication errors are not taken into account). Drugs are listed according to their active substances, which are encoded with the ATC classification and additional codes for substances that do not have a corresponding code. We extracted data over the period 2000–2016, which represents 382,484 reports with 5,906 different AEs and 2,344 different drugs.

3.3. Set of Reference Signals

The set of reference signals that we used to compare the performances of the methods is the DILIrank set recently established by Chen et al. (2011) which considers a specific adverse event: DILI. After determining a list of keywords related to the DILI event, this set was developed by text-mining the FDA-approved drug labels. A severity level was associated to each keyword. Drugs were classified into three DILI-related categories according to where keywords appear in the labeling section (Warnings & Precautions section, for example) of the FDA-approved drug labels, and the severity level of the involved keywords: most DILI concern, less DILI concern and no DILI concern. In 2016, this classification was refined by including information from other data sources to assess the causal relationship between each considered drug and a DILI event. Only drugs confirmed as DILI cause were retained (Chen et al., 2016).

We translated DILI labeling keywords into PT codes from the MedDRA classification: 91 PT codes were considered as DILI related. Each spontaneous report involving one of these 91 PT was thus considered as a reported DILI event, which translated into 22,440 DILI reports in the French pharmacovigilance database. To avoid numerical issues due to very low drug-reporting frequency, we limited the comparison to drugs which have more than three reports in common with a DILI, and have more than ten reports in total. This restriction applied, we considered 922 different drugs out of the 2,344 original drugs from the spontaneous reporting database. Considering these drugs, the final DILIrank set accounted for 417 (drug, DILI) pairs: 90 no DILI concern pairs, 213 less DILI concern pairs and 114 most DILI concern pairs. We considered as negative controls the no DILI concern pairs, and as positive controls the most DILI concern ones, which involve drugs associated with severe DILI outcomes.

4. RESULTS

Performances of the methods were assessed in terms of number of signals detected, number of true positives and number of false negatives identified, sensitivity and specificity, positive predictive value (PPV) and false discovery proportion (FDP). **Table 2** summarizes the results. AdjustPS-BIC, adjustPS-CISL, adjustPS-GTB, and adjustPS-hdPS detected more signals than the other PS-based approaches with 308, 275, 273, and 310 signals respectively. All the iptwPS-based approaches led to the lowest number of signals with 35, 63, 70, 34 signals for iptwPS-BIC, iptwPS-CISL, iptwPS-GTB, and iptwPS-hdPS. The mwPS-based approaches gave an intermediate number of detected signals with 147, 121, 136, and 139 signals for mwPS-BIC, mwPS-CISL, mwPS-GTB, and mwPS-hdPS respectively. Multiple regression approaches detected between 99 and 173 signals. The univariate approach detected the largest number of signals: 359.

Multiple regression approaches gave the highest positive predictive values (PPV): 96.97, 100, and 100% for BIC-Lasso,

TABLE 1 | Main advantages and disadvantages of the compared signal detection methods.

Methods	Advantages	Disadvantages
Univariate approach (disproportionality method)	Very fast computation time Detection threshold based on classical test theory (*p*-values, FDR correction)	Do not account for multiple exposures
Penalized multiple logistic regression methods	Fast computation time Account for multiple exposures	Detection threshold not relying on classical test theory
Propensity-score based methods	Account for multiple exposures Detection threshold based on classical test theory (*p*-values, FDR correction)	Long calculation time for the propensity score estimation step

TABLE 2 | Number of signals detected for each method.

Method	Number of generated signals	Number of signals with known status (positive or negative)	True positives signals			False positives signals		
			n	PPV	Sensitivity	*n*	FDP	Specificity
Univ	359	105	86	81.90	75.44	19	18.10	78.89
BIC-lasso	173	66	64	96.97	56.14	2	3.03	97.78
CISL-5%	99	43	43	100.00	37.72	0	0.00	100.00
CISL-10%	109	48	48	100.00	42.11	0	0.00	100.00
adjustPS-BIC	308	96	81	84.38	71.05	15	15.62	83.33
mwPS-BIC	147	53	52	98.11	45.61	1	1.89	98.89
iptwPS-BIC	35	14	13	92.86	11.40	1	7.14	98.89
adjustPS-CISL	275	86	75	87.21	65.79	11	12.79	87.78
mwPS-CISL	121	50	49	98.00	42.98	1	2.00	98.89
iptwPS-CISL	63	17	14	82.35	12.28	3	17.65	96.67
adjustPS-GTB	273	85	74	87.06	64.91	11	12.94	87.78
mwPS-GTB	136	52	49	94.23	42.98	3	5.77	96.67
iptwPS-GTB	70	28	25	89.29	21.93	3	10.71	96.67
adjustPS-hdPS	310	93	83	89.25	72.81	10	10.75	88.89
mwPS-hdPS	139	54	53	98.15	46.49	1	1.85	98.89
iptwPS-hdPS	34	16	15	93.75	13.16	1	6.25	98.89

For the BIC-lasso a signal is a positive association in the selected model, for CISL-5% and CISL-10% a signal is a selected variable obtained according to a distribution quantile (5% or 10%) in the CISL methodology. For Univ and all the PS-based approaches, a signal is an association FDR significant. The number of true positives signals is the number of positives controls detected by the method. The number of false positives signals is the number of negatives controls detected by the method.
PPV: Positive Predictive Value.
FDP: False Discovery Proportion.

CISL-5% and CISL-10% respectively. All the mwPS-based approaches gave similar results with a PPV around 98%, except for mwPS-GTB with a PPV equal to 94.23%. The univariate approach had the lowest PPV: 81.90%. AdjustPS-based approaches had slightly better performances: 84.38, 87.21, 87.06, and 89.25% for adjustPS-BIC, adjustPS-CISL, adjustPS-GTB, and adjustPS-hdPS respectively. Depending on the method used to estimate the PS, iptwPS-based approaches performed differently with PPVs ranging from 82.35% for iptwPS-CISL to 93.75% for iptwPS-hdPS.

Univ, adjustPS-BIC, adjustPS-CISL, adjustPS-GTB, and adjustPS-hdPS had the best performance in terms of sensitivity: between 64.91% for adjustPS-GTB and 75.44% for Univ, but they had the poorest specificity: between 78.9% for Univ and 88.9% for adjsutPS-hdPS. On the other hand, multiple regression approaches and mwPS-based approaches had very good specificity. BIC-Lasso, CISL-5% and CISL-10% had a specificity equal to 96.97, 100, and 100% respectively. MwPS-BIC, mwPS-CISL, mwPS-GTB, and mwPS-hdPS had a specificity equal to 98.11, 98.00, 94.23, and 98.15% respectively. Multiple regression approaches had sensitivity between 37.72% for CISL-5% and 56.14% for BIC-Lasso. MwPS-based approaches had sensitivity between 42.98% for mwPS-CISL and mwPS-GTB, and 46.49% for mwPS-hdPS. IptwPS-based approaches had good specificity but very low sensitivity.

Figure 1 shows the number of true positives or true negatives according to the number of signals generated by Univ, BIC-Lasso and mwPS-BIC, where signals are ranked according to their *p*-values for BIC-Lasso, and according to their adjusted *p*-values for Univ and mwPS-BIC. BIC-Lasso and mwPS-BIC were chosen to be represented here since they showed either better or similar performances to the other multiple regression or PS-based approaches respectively (**Figures S1**–**S3**). Regarding true positive signal detection, BIC-Lasso performed the best and Univ the worst, mwPS-BIC having a comparable performance to BIC-Lasso. Concerning false positive signal detection, mwPS-BIC performed the best since it detected the first false positive the latest. Univ detected far more signals than mwPS-BIC and BIC-Lasso, but its true positive detection rate was worse on its latest detected signals. Indeed, 36% of its first 150 generated signals were true positive controls compared to 15% of its last 209 signals. The same trend was observed for false positive detection: Univ detected 2% of false positives within its first 150 generated signals and about 7% over its last 209 signals.

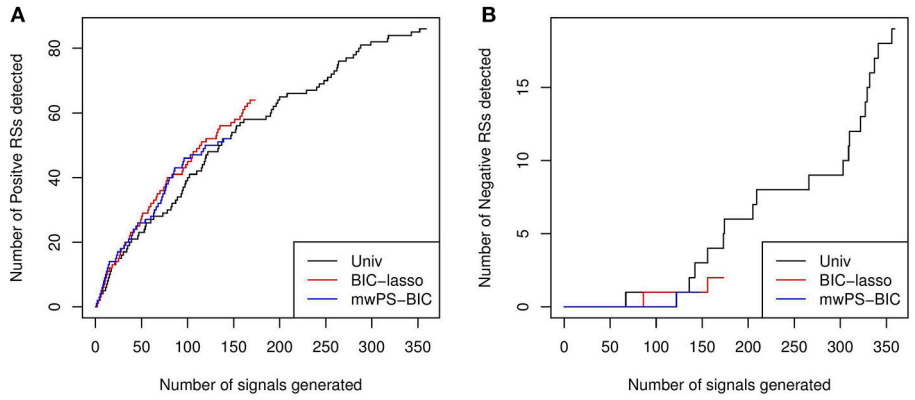


FIGURE 1 | (A) Number of positive reference signals detected according to number of signals generated by BIC-Lasso, mwPS-BIC and Univ, where signals are ranked in ascending order by their associated *p*-values for BIC-Lasso and by their adjusted *p*-values for mwPS-BIC and Univ. **(B)** Number of negative reference signals detected according to number of signals generated by BIC-lasso, mwPS-BIC and Univ, where signals are ranked in ascending order by their associated *p*-values for BIC-Lasso and by their adjusted *p*-values for mwPS-BIC and Univ.

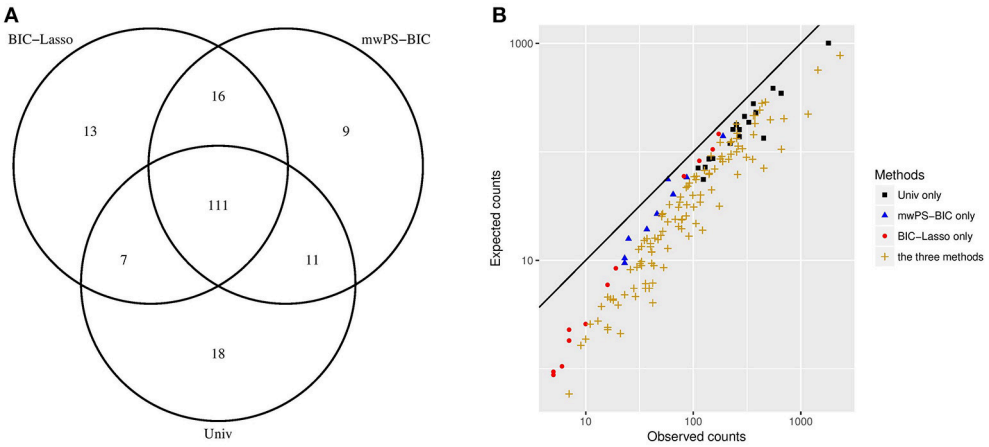


FIGURE 2 | (A) Distribution of the first 147 signals generated between Univ, BIC-Lasso and mwPS-BIC. **(B)** Observed counts vs. expected counts of signals generated by Univ only, BIC-Lasso only, mwPS-BIC only and by the three methods, considering their first 147 generated signals. Observed counts *n* are number of reports which involved the signal considered, expected counts *e* are those expected if independence applies between the drug and the AE that form the signal considered . They are calculated as follows: for a signal ($drug_i$, AE_j) $e_{i,j} = \frac{N_i \times N_j}{N}$ where N_i, N_j are the observed counts of $drug_i$ and AE_j respectively, and *N* the total number of observations.

Looking more specifically at the PS-based approaches, they behaved similarly within a given PS-estimation method (**Figures S1**, **S2**). Considering the three PS methodologies, mwPS-based approaches performed better for an equal number of signals generated (**Figure S3**).

Figure 2A shows the overlap between the first 147 signals generated by Univ, BIC-Lasso and mwPS-BIC and **Figure 2B** the observed counts *n* vs. the expected counts *e* of some of these signals according to the method(s) by which they were detected. 111 signals were detected by all methods: **Figure 2B** shows that these signals are characterized by a greater observed risk ratio ($\frac{n}{e}$) and/or a greater support of the data (large *n*) in comparison to signals detected solely by one method. **Figure 2B** also shows that the methods tend to focus on different types of signals: in particular the BIC-Lasso highlighted some signals with relatively weak support of the data ($n < 10$) whereas Univ highlighted signals with low observed risk but a large number of reports. mwPS-BIC had an intermediary behavior.

5. DISCUSSION

Development of novel signal detection methods is crucial to improve the responsiveness and the efficiency of post-marketing surveillance systems. The main issue is to develop techniques that give a reasonable number of signals for further analysis by experts, while providing a list of relevant signals with the fewest possible false associations. Using spontaneous reporting

databases is difficult because of their high dimension and extreme imbalance. This study compared recently proposed multiple approaches and a set of methods based on the PS in a signal detection framework. To do so, we used a recently developed set of reference signals pertaining to liver injury: the DILIrank set.

The best performing PS-based approach is the one which relies on the MW. Its performance in terms of true/false signal detection, specificity and sensitivity is very close to that of the multiple regression approaches, which performs the best. It also generates roughly the same number of signals as BIC-Lasso and CISL, the latter being conservative for highly reported adverse events, as is the case for DILI (Ahmed et al., 2018). The mwPS-based approaches have an intermediary behavior between multiple regression approaches and univariate approach in the type of signal generated.

In accordance with the literature, adjustment on the PS does not achieve the best performances (Stuart, 2010). The adjustment approaches had similar behavior to the univariate approach. In particular, they generated very large numbers of signals in comparison to the other PS-based approaches and multiple regression approaches, and they had poor specificity and good sensitivity.

The iptwPS-based approaches performed extremely poorly. Unlike the MW, weights from IPTW are not normalized and could be very large for untreated individuals with low PS. This numerical instability due to high weights with IPTW weighting has already been reported (Yoshida et al., 2017). To avoid this issue, a solution is to do weight-trimming: weights greater than a given value are assigned this value (Seeger et al., 2017).

In addition to the well-known variable selection algorithm in PS estimation, hdPS, here we implemented three other PS-estimation methods: two based on variable selection algorithms, BIC-Lasso or CISL, and one derived from a machine-learning algorithm: GTB. A drawback with the hdPS algorithm in our setting is that the PS obtained with this method is AE dependent: for each drug to be included in the PS, its association with the AE under study has to be computed. This can become very time-consuming when screening several thousands of AEs. On the contrary, the three other PS-estimation methods have a computational advantage: once the PS is estimated for a given drug, it can be used to test associations with any AEs. When PS is estimated with the covariates selected by BIC-Lasso, CISL or with a GTB algorithm, PS-based approaches are competitive with multiple regression approaches in terms of calculation time. Regressions like weighted univariate logistic regressions can be easily performed. Depending on the PS methodology used (adjustment or weighting), results are roughly comparable according to the way PS is estimated.

A major constraint in developing signal detection methods is to obtain a reliable set of reference signals to assess performances. Here we used a recently established set, the DILIrank set pertaining to a common adverse event, which is not the case for all the AEs in the database. Furthermore, three levels of DILI risk assessment are defined in DILIrank for drugs: most DILI concern, less DILI concern and no DILI concern. We chose to consider only most DILI concern drugs as true positive controls, leaving less DILI concern drugs, because we decided to focus on drugs which could lead to severe DILI related outcomes.

The results suggest that the proposed PS-based methodology is an interesting complement to other existing methods. In a sense, it combines the main strengths of both univariate and multiple regression approaches: it makes it possible to account for co-reported drugs while using multiple hypothesis testing theory as regards the detection threshold. Further studies on alternative reference sets and simulation studies will be useful to confirm its potential for automated signal detection in pharmacovigilance.

AUTHOR CONTRIBUTIONS

ÉC, IA, and PT-B conceived and designed the study. ÉV contributed to the design of the work. AP and FS contributed to the acquisition and the interpretation of data for the work. ÉC performed the computations. ÉC, IA, PT-B, and ÉV discussed the results. ÉC drafted the manuscript with support from IA and PT-B. All authors critically revised the work and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank the regional pharmacovigilance centers and the ANSM for providing the pharmacovigilance database dataset.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2018.01010/full#supplementary-material

REFERENCES

Ahmed, I., Dalmasso, C., Haramburu, F., Thiessard, F., Broët, P., and Tubert-Bitter, P. (2010). False discovery rate estimation for frequentist pharmacovigilance signal detection methods. *Biometrics* 66, 301–309. doi: 10.1111/j.1541-0420.2009.01262.x

Ahmed, I., Pariente, A., and Tubert-bitter, P. (2018). Class-imbalanced subsampling lasso algorithm for discovering adverse drug reactions. *Statist Methods Med. Res.* 27, 785–797. doi: 10.1177/0962280216643116

Ahmed, I., Bégaud, B., and Tubert-Bitter, P. (2015). "chapter 13: evaluation of post-marketing safety using spontaneous reporting databases," in *Statistical Methods for Evaluating Safety in Medecine Product Development*, ed A. L. Gould (Chichester, UK: Wiley), 332–344.

Almenoff, J.S., Pattishall, E.N., Gibbs, T.G., Dumouchel, W., Evans, S. J., and Yuen, N. (2007). Novel statistical tools for monitoring the safety of marketed drugs. *Clin. Pharmacol. Therapeut.* 82, 157–166. doi: 10.1038/sj.clpt.6100258

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational Studies. *Multivar. Behav. Res* 46. 399–424. doi: 10.1080/00273171.2011.568786

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* 57, 289–300.

Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J., and Stürmer, T. (2006). Variable selection for propensity score models. *Am. J. Epidemiol.* 163, 1149–1156. doi: 10.1093/aje/kwj149

Caster, O., Norén, G. N., Madigan, D., and Bate, A. (2010). Large-Scale regression-based pattern discovery: the example of screening the WHO global drug safety database. *Statist. Anal. Data Minig* 3, 197–208. doi: 10.1002/sam.10078

Chen, M., Suzuki, A., Thakkar, S., Yu, K., and Hu, C. (2016). DILIrank : the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Disc. Tod.* 21, 648–653. doi: 10.1016/j.drudis.2016.02.015

Chen, M., Vijay, V., Shi, Q., Liu, Z., and Fang, H. (2011). FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Disc. Tod.* 16, 697–703. doi: 10.1016/j.drudis.2011.05.007

Franklin, J. M., Eddings, W., Austin, P. C., Stuart, E. A., and Schneeweiss, S. (2017). Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. *Stat. Med.* 36, 1946–1963. doi: 10.1002/sim.7250

Harpaz, R., Dumouchel, W., Lependu, P. Bauer-Mehren A., Ryan, P., and Shah, N.H. (2013). Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin. Pharmacol. Therap.* 93, 539–546. doi: 10.1038/clpt.2013.24

Harpaz, R., Dumouchel, W., Shah, N. H., Madigan, D., Ryan, P., and Friedman, C. (2012). Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin. Pharmacol. Therapeut.* 91, 1010–1021. doi: 10.1038/clpt.2012.50

Hastie, T., Tibshirani, R., and Friedman, J. (2009). "Boosting and additive trees," in *The Elements of Statistical Learning*, eds T. Hastie, R. Tibshirani and J. Friedman (New York, NY: Springer), 337–387.

Lee, B. K., Lessler, J., and Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statist. Med.* 29, 337–346. doi: 10.1002/sim.3782

Li, L., and Greene, T. (2013). A weighting analogue to pair matching in propensity score analysis. *Int. J. Biostatist.* 9, 215–234. doi: 10.1515/ijb-2012-0030

Marbac, M., Tubert-Bitter, P., and Sedki, M. (2016). Bayesian model selection in logistic regression for the detection of adverse drug reactions. *Biometr. J.* 58, 1376–1389. doi: 10.1002/bimj.201500098

Meinshausen, N., and Bühlmann, P. (2010). Stability selection. *J. R. Statist. Soc.* 72, 417–473. doi: 10.1111/j.1467-9868.2010.00740.x

Patorno, E., Grotta, A., Bellocco, R., and Schneeweiss, S. (2013). Propensity score methodology for confounding control in health care utilization databases. *Epidemiol. Biostatist. Public Health* 10:e8940-1. doi: 10.2427/8940

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55. doi: 10.1093/biomet/70.1.41

Schneeweiss, S., Rassen, J. A., Glynn, R. J., Avorn, J., Mogun, H., and Brookhart, M. A. (2009). High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 20, 512–522. doi: 10.1097/EDE.0b013e3181a663cc

Seeger, J. D., Bykov, K., Bartels, D. B., Huybrechts, K., and Schneeweiss, S. (2017). Propensity score weighting compared to matching in a study of dabigatran and warfarin. *Drug Saf.* 40, 169–181. doi: 10.1007/s40264-016-0480-3

Seeger, J. D., Williams, P. L., and Walker, A. M. (2005). An application of propensity score matching using claims data. *Pharmacoepidemiol. Drug Saf.* 14, 465–476. doi: 10.1002/pds.1062

Stuart, E. A. (2010). Matching methods for causal inference : a review and a look forward. *Stat. Sci.* 25, 1–21. doi: 10.1214/09-STS313

Tatonetti, N. P., Ye, P. P., Altman, R. B., and Daneshjou, R. (2012). Data-driven prediction of drug effects and interactions. *Sci. Transl. Med.* 4:125ra31. doi: 10.1126/scitranslmed.3003377

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Statist. Soc.* 58, 267–288.

van Puijenbroek, E. P., Bate, A., Leufkens, H. G., Lindquist, M., Orre, R., and Egberts, A. C. (2002). A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol. Drug Saf.* 11, 3–10. doi: 10.1002/pds.668

Yoshida, K., Hernández-Díaz, S., Solomon, D. H., Jackson, J. W., Gagne, J. J., Glynn, R. J., et al. (2017). Matching weights to simultaneously compare three. *Epidemiology* 28, 387–395. doi: 10.1097/EDE.0000000000000627

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers

Supplementary Material: Propensity score-based approaches in high dimension for pharmacovigilance signal detection: an empirical comparison on the French spontaneous reporting database

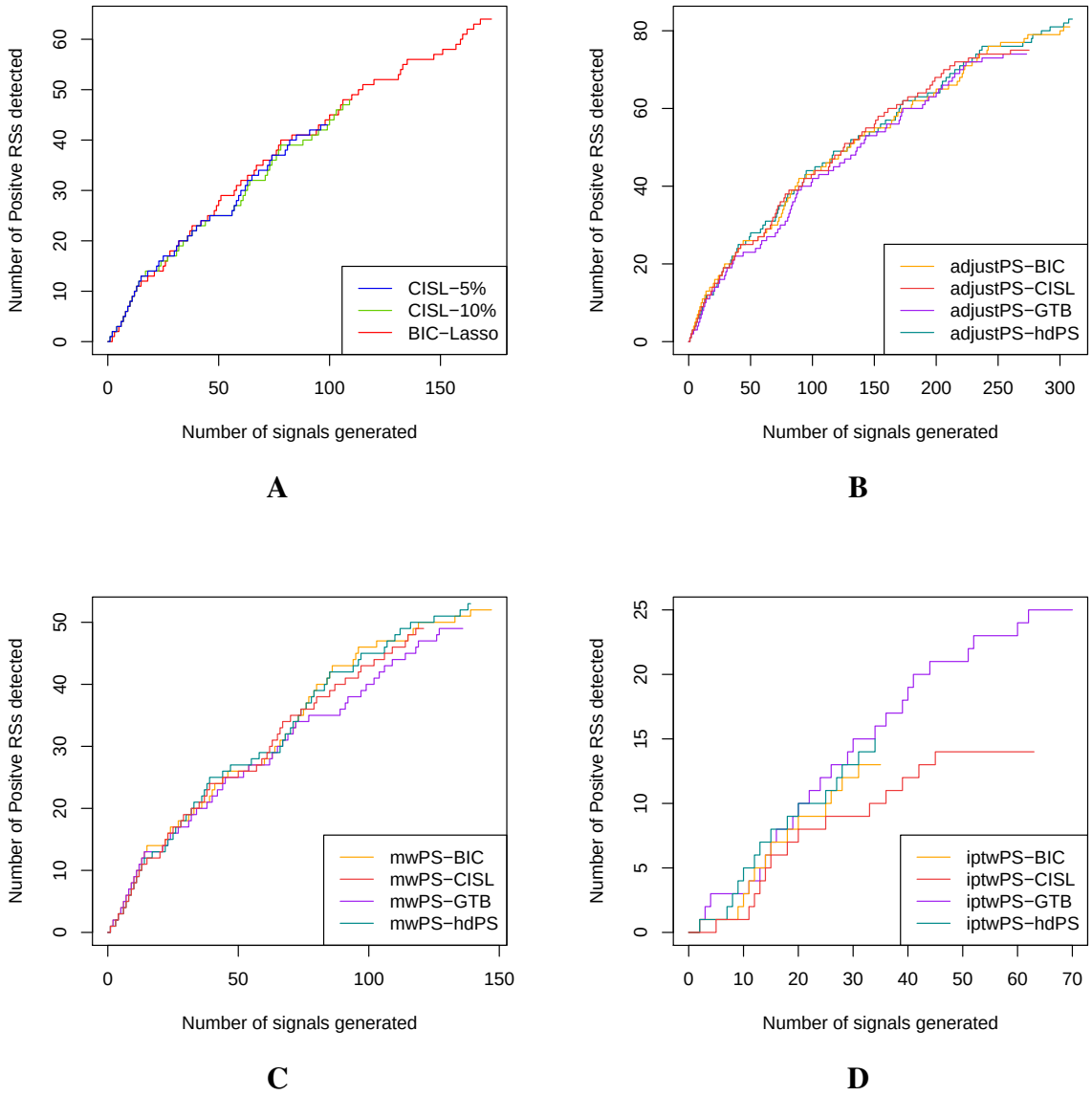


Figure S1: **(A):** Number of positive reference signals detected according to number of signals generated by multiple regression approaches: BIC-lasso, CISL-5% and CISL-10%, where signals are ranked in ascending order by their associated p-values for BIC-Lasso, and in decreasing order of their 5% or 10% output distribution quantile value for CISL-5% and CISL-10%. **(B):** Number of positive reference signals detected according to number of signals generated by adjustment on propensity score methods, depending on how the propensity score was estimated: adjustPS-BIC, adjustPS-CISL, adjustPS-GTB and adjustPS-hdPS, where signals are ranked in ascending order by their adjusted p-values. **(C):** Number of positive reference signals detected according to number of signals generated by weighting with matching weights on propensity score methods, depending on how the propensity was estimated: mwPS-BIC, mwPS-CISL, mwPS-GTB and mwPS-hdPS, where signals are ranked in ascending order by their adjusted p-values. **(D):** Number of positive reference signals detected according to number of signals generated by inverse probability of treatment weighting on propensity score methods, depending on how the propensity was estimated: iptwPS-BIC, iptwPS-CISL, iptwPS-GTB and iptwPS-hdPS, where signals are ranked in ascending order by their adjusted p-values.

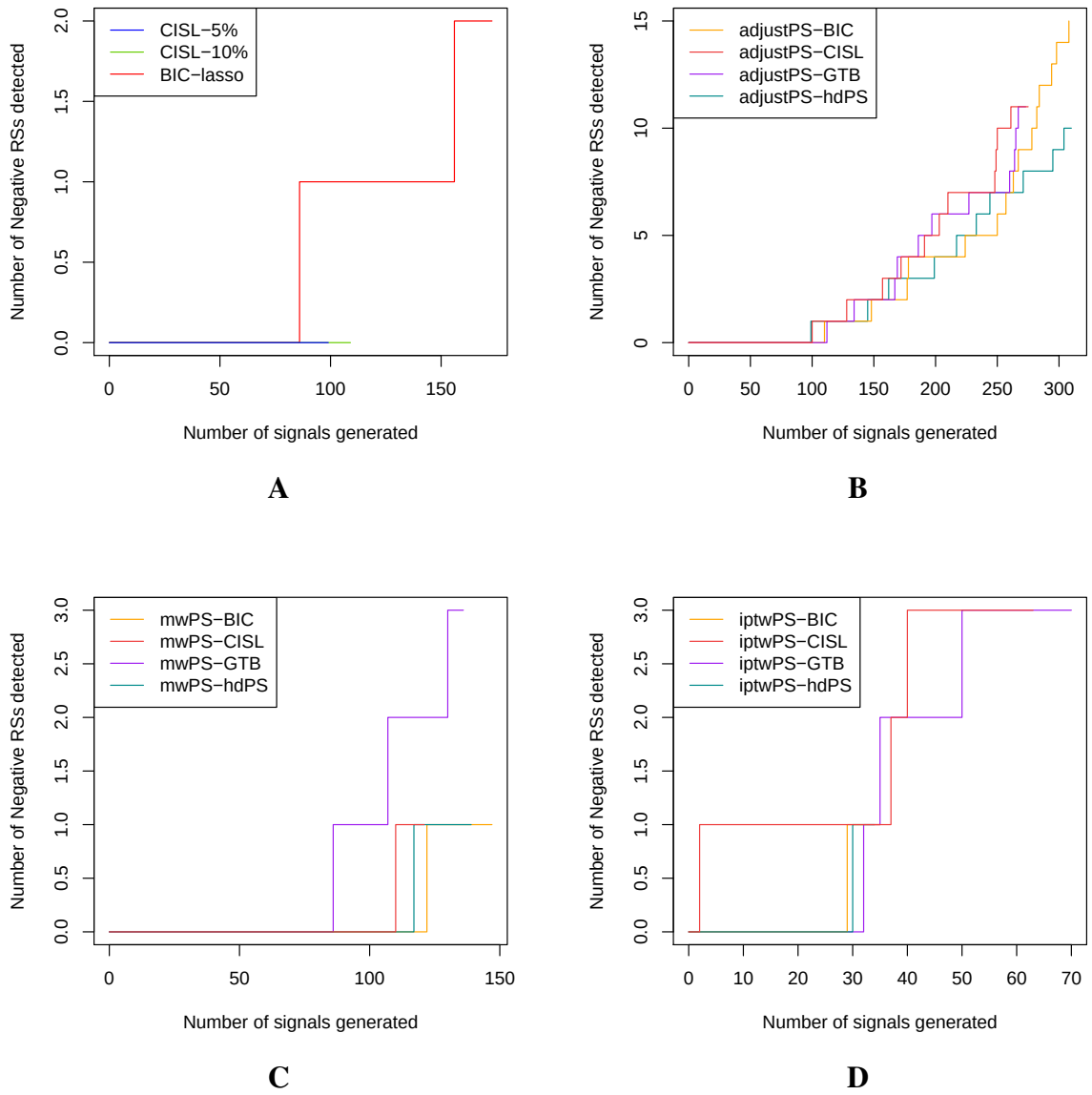


Figure S2: **(A):** Number of negative reference signals detected according to number of signals generated by multiple regression approaches: BIC-lasso, CISL-5% and CISL-10%, where signals are ranked in ascending order by their associated p-values for BIC-Lasso, and in decreasing order of their 5% or 10% output distribution quantile value for CISL-5% and CISL-10%. **(B):** Number of negative reference signals detected according to number of signals generated by adjustment on propensity score methods, depending on how the propensity score was estimated: adjustPS-BIC, adjustPS-CISL, adjustPS-GTB and adjustPS-hdPS, where signals are ranked in ascending order by their adjusted p-values. **(C):** Number of negative reference signals detected according to number of signals generated by weighting with matching weights on propensity score methods, depending on how the propensity was estimated: mwPS-BIC, mwPS-CISL, mwPS-GTB and mwPS-hdPS, where signals are ranked in ascending order by their adjusted p-values. **(D):** Number of negative reference signals detected according to number of signals generated by inverse probability of treatment weighting on propensity score methods, depending on how the propensity was estimated: iptwPS-BIC, iptwPS-CISL, iptwPS-GTB and iptwPS-hdPS, where signals are ranked in ascending order by their adjusted p-values.

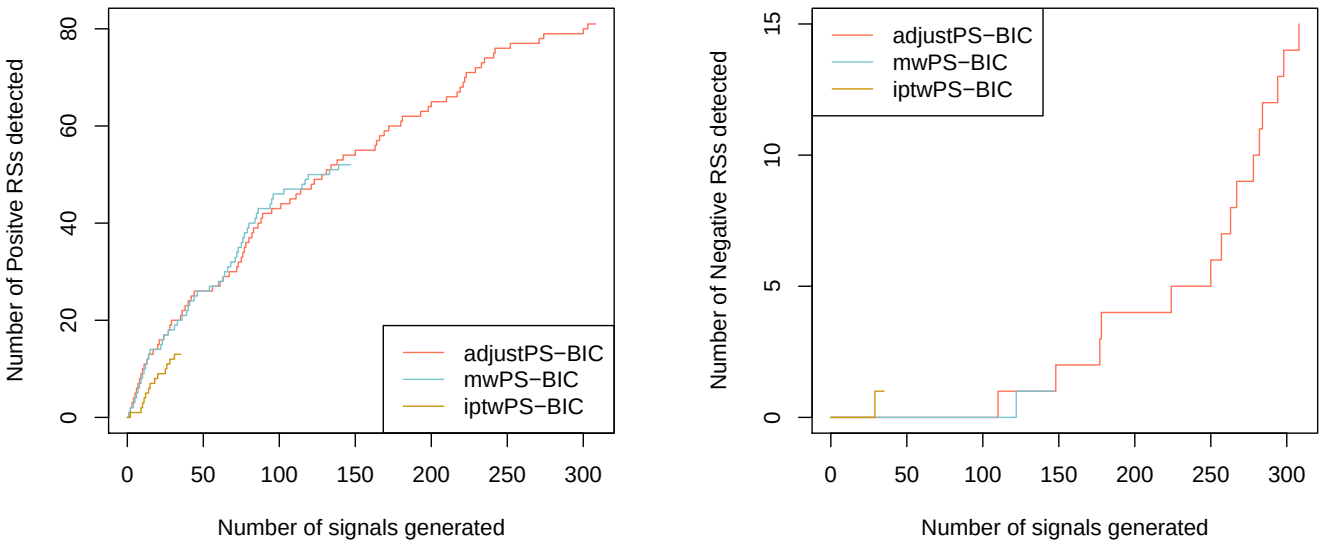


Figure S3: **(A):** Number of positive reference signals detected according to number of signals generated by adjustment on propensity score methods, weighting with matching weights on propensity score methods and inverse probability of treatment weighting on propensity score methods, where the score is estimated with BIC-Lasso methodology: adjustPS-BIC, mwPS-BIC, iptwPS-BIC, where signals are ranked in ascending order by their adjusted p-values. **(B):** Number of negative reference signals detected according to number of signals generated by adjustment on propensity score methods, weighting with matching weights on propensity score methods and inverse probability of treatment weighting on propensity score methods, where the score is estimated with BIC-Lasso methodology: adjustPS-BIC, mwPS-BIC, iptwPS-BIC, where signals are ranked in ascending order by their adjusted p-values.

D.2 Article soumis à la revue *Statistical Methods in Medical Research*

New adaptive lasso approaches for variable selection in automated pharmacovigilance signal detection

Émeline Courtois*,[1], Pascale Tubert-Bitter †,[1], and Ismaïl Ahmed ‡,[1]

[1]Inserm, Université Paris Saclay, Inserm U1018 - Center for Epidemiology and Population Health (CESP), Team High-Dimensional Biostatistics for Drug Safety and Genomics.

Short title: Adaptive lasso approaches in pharmacovigilance

Abstract

Adverse effects of drugs are often identified after market introduction. Post-marketing pharmacovigilance aims to detect them as early as possible and relies on spontaneous reporting systems collecting suspicious cases. Signal detection tools have been developed to mine these large databases and counts of reports are analysed with disproportionality methods. To address disproportionality method biases, recent methods applied to individual observations taking into account all exposures for the same patient. In particular, the logistic lasso provides an efficient variable selection framework, yet the choice of the regularization parameter is a challenging issue and the lasso variable selection may be inconsistent. We propose a new signal detection methodology based on the adaptive lasso in the context of high dimension. We derived two new adaptive weights from (i) a lasso regression using the Bayesian Information Criterion (BIC), and (ii) the class-imbalanced subsampling lasso, an extension of stability selection. The BIC is used in the adaptive lasso stage for variable selection. We performed an extensive simulation study and an application to real data, where we compared our methods to the existing adaptive lasso, and recent detection approaches based on lasso regression or propensity scores. Our methods compared very well. They are implemented in R package `adapt4pv`.

Keywords: adaptive logistic lasso, BIC, class imbalanced subsampling lasso, variable selection, drug safety signal, spontaneous reporting, simulations, propensity score, drug-induced liver injury

1 Introduction

Because the conditions of exposure of an active drug in real life are very different from those of clinical trials, the adverse effects of drugs are often identified once they are introduced on the market. This may be due to a complex interaction with subcategories of population, or to

*Corresponding author: `emeline.courtois@inserm.fr`

†These authors share last authoship

‡

a long latency period after exposure. Post-marketing pharmacovigilance aims to detect as early as possible these adverse effects that have not been identified during the safety assessment stages of drug development. Pharmacovigilance systems rely on large databases of individual case safety reports of adverse events (AEs) suspected to be drug-induced. Many countries currently have a spontaneous reporting system as well as supranational entities such as the European Medicines Agency or the Uppsala Monitoring Centre in charge of pharmacovigilance for the World Health Organization. In France, the national pharmacovigilance database is maintained by the National Agency for the Safety of Drugs and Health Products (*Agence Nationale de Sécurité du Médicament et des Produits de Santé*, ANSM). It contained around 450 000 reports at the end of December 2017. Currently, about 36 000 reports are reported annually.

Several automated signal detection tools have been developed to mine these large amounts of data in order to highlight suspicious AE-drug combinations. To draw definite conclusions, these signals need further expert investigations or additional studies. Thus, it is important to generate a reasonable number of signals with as few false associations as possible for further analysis. Classical signal detection methods are based on disproportionality analyses of counts aggregating patients' reports for each drug-AE pair. [1, 2, 3, 4] These methods have been extended to account for multiple comparison testing in order to provide alternative signal ranking and detection thresholds based on false discovery rate (FDR) estimates. [5, 6, 7] Other methods for aggregated counts relying on likelihood ratio tests have also been proposed. [8, 9]

Disproportionality methods are subject to the masking effect bias and do not account for co-prescription. [10, 11, 12, 13] In recent years, multiple logistic regression-based signal detection methods which rely on lasso penalization [14, 15] have been proposed to address these limitations. Unlike disproportionality methods, they are directly applied to individual spontaneous reports rather than to aggregated counts. For an observation, the outcome is the presence or absence of a given AE, and the covariates are all drug presence indicators. The objective of pharmacovigilance therefore pertains to the variable selection framework by aiming to identify the drugs potentially associated with AE among the multitude of candidate

covariates. The drug exposure matrix is thus large, binary and extremely sparse, and there is also a large imbalance between the presence and absence of a given AE. More recently, signal detection methods based on propensity scores (PS) in high dimension have also been proposed as an alternative to address disproportionality method biases. [16, 17, 18]

Lasso penalization is a computationally efficient way to perform regression in high dimension. [19] The parsimony induced by the L_1 norm is also an appealing feature of this algorithm. Nevertheless, while cross-validation is classically used for the purpose of prediction, it is more complicated to choose the best regularization parameter controlling the sparsity of the model in the variable selection framework. Furthermore, it has been shown that there is no proper regularization parameter that allows the lasso to enjoy the oracle properties defined by Fan and Li. [20] This means for instance that the lasso variable selection may be inconsistent. Subsampling strategies such as stability selection [21] have been proposed to lessen the importance of the choice of regularization parameter, and the class-imbalanced subsampling lasso (CISL) [15] was specifically designed to account for the large imbalance in spontaneous reporting data.

The adaptive lasso is an alternative approach to improve the variable selection properties of the lasso. [22] It consists in using adaptive weights (AWs) for penalizing covariates differently in the L_1 penalty. Originally, AWs were derived from coefficients estimated by maximum likelihood. A high-dimensional version of the adaptive lasso has been proposed by Bühlmann and Van De Geer [23] in the linear case, in which the AWs are derived from the coefficients obtained by a first lasso regression. Huang et al. [24] proposed the same approach in the logistic case. They also showed in another work that in the linear case, AWs derived from univariate regression coefficients result in good recovery properties under certain conditions. [25]

In this work, we present a new automated signal detection strategy based on the adaptive lasso which aims at improving the guidance of the variable selection operated by the lasso through adaptive penalty weights specific to each covariate. This new strategy also involves the use of the Bayesian Information Criterion (BIC). We propose two new AWs derived from (i) a lasso logistic regression for which the regularization parameter is chosen using the BIC,

and (ii) CISL. These AWs are then incorporated into a lasso logistic regression using the BIC to choose the regularization parameter. We compare both versions of our approach to (i) more classical implementations of the adaptive lasso in high dimension, [23, 24, 25] (ii) lasso regressions considering cross-validation, BIC or permutations [26, 27] for choosing the regularization parameter, (iii) CISL, and (iv) the propensity score in high dimension-based approaches. We use extensive simulations exploiting real drug exposure data from the French pharmacovigilance database in order to preserve the sparsity of the covariates. We also present an empirical study on the French national database using a large and recently published reference set pertaining to drug-induced liver injuries (DILI). [28, 29]

2 Methods

2.1 The logistic lasso

Let N denote the number of spontaneous reports (i.e. the number of observations) and P the total number of drug covariates. Let $\mathbf{X}$ denote the $N \times P$ binary matrix of drug exposures and $\mathbf{y}$ the N-vector of binary responses that indicates the presence or absence of the AE of interest. For $i \in \{1, ..., N\}$, the corresponding multiple logistic model is

$$\text{logit}(\Pr(y_i = 1|x_i)) = \beta_0 + \sum_{p=1}^{P} \beta_p \, x_{ip}, \tag{1}$$

where β_0 is the intercept and $\boldsymbol{\beta}$ is a P-vector of regression coefficients associated with drug covariates. Although we are not in the $P >> N$ context, P is typically very large, which can cause some numerical problems with classical regression. The penalized logistic lasso consists in estimating:

$$(\widehat{\beta}_0, \widehat{\boldsymbol{\beta}})_\lambda = \text{argmax}_{(\beta_0, \boldsymbol{\beta})}\{l((\beta_0, \boldsymbol{\beta}), \mathbf{y}, \mathbf{X}) - \text{pen}(\lambda)\},$$

where l is the log-likelihood of model (1), λ is the regularization parameter and $\text{pen}(\lambda)$ is defined as

$$\text{pen}(\lambda) = \lambda|\boldsymbol{\beta}|_1 = \lambda \sum_{p=1}^{P} |\beta_p|. \tag{2}$$

We denote by $\widehat{\boldsymbol{\beta}}_\lambda$ the P-vector in the lasso estimate $(\widehat{\beta}_0, \widehat{\boldsymbol{\beta}})_\lambda$. Thanks to the L_1 penalty in (2), some coefficients of $\widehat{\boldsymbol{\beta}}_\lambda$ are shrunk to exactly zero, so the covariates associated with these

coefficients are not retained in the model. By controlling the amount of penalization, the λ parameter in the lasso regression is closely related to the number of non-zero estimated coefficients. Since the aim in pharmacovigilance is to detect deleterious associations with the outcome, we are only interested in covariates with a positive associated penalized coefficient in $\widehat{\boldsymbol{\beta}}_\lambda$.

2.1.1 Penalization parameter selection

We considered three strategies for selecting the penalization parameter: cross-validation, BIC and permutations. One round of cross-validation involves partitioning the dataset into n_f subsets, called folds: $n_f - 1$ are used as a training set, i.e. the model is estimated on this set, and the remaining fold is used as a validation set where a prediction performance metric (e.g. area under curve or deviance) is calculated. This procedure is repeated so that each fold is used exactly once as a validation set. An average value of the performance metric and a standard deviation are then calculated over the n_f obtained values. In the lasso regression context, cross-validation is performed for each tested λ value. The selected λ according to cross-validation is the one with the best result in terms of the prediction performance metric selected. In this work, we used the deviance, and we set n_f to 5.

An alternative strategy for selecting the penalization parameter is to use the BIC. [15] For each tested λ, we implemented the BIC as follows:

$$\text{BIC}_\lambda = -2l_\lambda + \text{df}(\lambda)\ \ln(N), \tag{3}$$

where l_λ is the log-likelihood of the classical multiple logistic regression model, which includes the set of covariates with a non-zero coefficient in $\widehat{\boldsymbol{\beta}}_\lambda$, and $\text{df}(\lambda) = |\widehat{\boldsymbol{\beta}}_\lambda \neq 0|$. If different λs lead to the same subset of retained covariates, then the non-penalized models resulting from these λs are the same, as is the BIC. Consequently, this approach selects the subset of covariates that leads to the classical model which minimizes the BIC defined in (3), rather than selecting a particular λ.

An approach based on permutations for selecting the penalization parameter in lasso regression was proposed by Sabourin et al. [27] based on the suggestion of Ayers and Cordell. [26] Denoting π as any permutation of $\{1, ..., N\}$, let $\mathbf{y}_{\pi_l} = (y_{\pi(1)}, ..., y_{\pi(N)})$ be a permuted

version of the outcome $\mathbf{y}$ with $1 \leqslant l \leqslant K$. A lasso regression is performed for each of these permutations by regressing $\mathbf{y}_{\pi_l}$ on the original data set $\mathbf{X}$. One then obtains $\lambda_{max}(\mathbf{y}_{\pi_l})$, i.e. the smallest value of the penalty parameter, such that no covariate is selected in the lasso regression on $\mathbf{y}_{\pi_l}$. As in Sabourin et al., [27] we used the median value of $(\lambda_{max}(\mathbf{y}_{\pi_1}), ..., \lambda_{max}(\mathbf{y}_{\pi_K}))$ in a lasso regression performed with the original outcome $\mathbf{y}$. In this work, we set $K = 20$.

In the following, we refer to the approach involving cross-validation, BIC or permutation to choose the penalty parameter as lasso-cv, lasso-bic and lasso-perm, respectively.

2.1.2 Lasso and subsampling

To circumvent the penalization parameter selection issue in lasso regression, Meinshausen and Bühlman proposed the stability selection algorithm. [21] Briefly, it consists of perturbing the data by subsampling many times, implementing lasso regression on these subsamples randomly drawn without replacement, and choosing covariates that occur in a large fraction of the resulting selected sets induced by the lasso path of regularization. Ahmed et al. proposed a variation of this method to account for the large imbalance of the outcome that occurs in pharmacovigilance databases: the CISL algorithm. [15] In CISL, subsamples are drawn following a nonequiprobable sampling scheme with replacement in order to allow a better representation of individuals who experienced the outcome of interest. Lasso regressions are performed in each of these samples and the following quantity is computed:

$$\widehat{\pi}_p^b = \frac{1}{E} \sum_{\eta=1}^{E} \mathbb{1}[\widehat{\beta}_p^{\eta,b} > 0], \tag{4}$$

where E is the maximum number of covariates selected by all the lasso regressions, $\eta \in \{1, .., E\}$ is the number of covariates selected and $\widehat{\beta}_p^{\eta,b}$ is the regression coefficient estimated by the logistic lasso for drug p, on sample $b \in \{1, .., B\}$ for a model including η covariates. Thus, for each drug, an empirical distribution of $\widehat{\pi}_p^b$ is obtained over all B samples. The drug covariate is then selected if a given quantile of the distribution of $\widehat{\pi}_p^b$ is non-zero. In this work, we considered the covariate sets established with the 10% quantiles of these distributions following Ahmed et. al.'s recommendation. [15]

2.2 Adaptive lasso

As defined by Fan and Li, [20] an optimal procedure in statistical learning should have the following oracle properties: (i) identifies the right subset of true predictors, and (ii) produces unbiased estimates. In their work, they showed that the lasso procedure does not enjoy these oracle properties. Indeed, there are some scenarios in which the lasso variable selection could be inconsistent. Furthermore, with an equal penalty for all covariates, the lasso tends to overpenalize the relevant ones and to produce biased estimates for true large coefficients. To overcome this drawback, Zou [22] proposed the adaptive lasso in which AWs are used to penalize covariates differently in the L_1 penalty:

$$\text{pen}(\lambda) = \lambda \sum_{p=1}^{P} w_p |\beta_p|.$$

The penalty applied to the covariate p is defined by $\lambda_p = \lambda \times w_p$. The higher the value of the weight w_p, the more the variable p is penalized and the less likely the variable is to be included in the model. By assigning a higher penalty to small coefficients and a lower penalty to large ones, the adaptive lasso makes it possible to consistently select the right model and produce unbiased estimates. Thus, Zou showed in his work that under certain conditions for the AWs, the adaptive lasso enjoys the oracle properties.

To build the AWs, Zou proposed to use an initial consistent estimator of $\boldsymbol{\beta}^*$, the P-vector of regression coefficients. To this end, he considered $\widehat{\boldsymbol{\beta}}_p^{mle}$ the maximum likelihood estimate for covariate p and defined the associated penalty weight $w_p = \frac{1}{|\widehat{\beta}_p^{mle}|^\gamma}$, with $\gamma > 0$. However, in the high-dimensional context, it is non-trivial to find a consistent estimate for constructing the AWs since computing the maximum likelihood is not feasible.

In the linear case, Bühlmann and Van De Geer [23] proposed to use the penalized regression coefficients estimated by a lasso regression to determine these AWs considering $\gamma = 1$. In both the lasso and the adaptive lasso, the penalization parameter λ was selected through cross-validation. This two-stage procedure, which involves an initial lasso step with cross-validation, was proposed in the logistic case under the name of iterated lasso. [24] By denoting $\widehat{\boldsymbol{\beta}}^{lcv}$ the P-vector of lasso regression coefficients determined with lasso-cv, the AWs

associated with the drug covariate p in the adaptive lasso stage are defined as:

$$w_p^{lcv} = \begin{cases} \dfrac{1}{|\widehat{\beta}_p^{lcv}|} & \text{if} \quad \widehat{\beta}_p^{lcv} \neq 0 \\ \infty & \text{if} \quad \widehat{\beta}_p^{lcv} = 0 \end{cases}$$

Thus, a covariate that has not been selected with lasso-cv in the first stage is automatically excluded in the adaptive lasso stage. In the following, we refer to this approach as adapt-cv.

In the linear case, Huang et al. [25] showed that under certain conditions, using univariate regression coefficients to determine the AWs presents nice properties. By denoting $\widehat{\beta}_p^{univ}$ the univariate coefficient associated to drug covariate p, we defined the following AWs associated with the drug covariate p in the adaptive lasso stage:

$$w_p^{univ} = \frac{1}{|\widehat{\beta}_p^{univ}|}.$$

As in the work of Huang et al., we chose the penalisation parameter according to cross-validation in the adaptive lasso stage. We refer to this approach as adapt-univ.

2.3 Extending adaptive lasso for pharmacovigilance

Since the aim in pharmacovigilance is to select the right subset of drugs associated with an AE, we sought to develop a signal detection approach by enhancing the performance of the adaptive lasso in terms of variable selection. To this end, we first use the BIC as defined in section 2.1.1 to identify the final subset of covariates in the adaptive lasso stage instead of cross-validation. We also propose two new AWs.

The first one consists in using the BIC in the first stage. By denoting $\widehat{\boldsymbol{\beta}}^{lbic}$ the P-vector of unpenalized regression coefficients estimated in the first stage with lasso-bic, we define the following AWs associated with the drug covariate p in the adaptive lasso stage by:

$$w_p^{lb} = \begin{cases} \dfrac{1}{|\widehat{\beta}_p^{lbic}|} & \text{if} \quad \widehat{\beta}_p^{lbic} \neq 0 \\ \infty & \text{if} \quad \widehat{\beta}_p^{lbic} = 0 \end{cases}$$

A covariate that has not been selected by the lasso-bic in the first stage is automatically excluded in the adaptive lasso stage.

The second proposed AWs are derived from the CISL approach. We first compute CISL by considering a non-zero constraint in calculating of the quantity (4) instead of the original

positive constraint:

$$\widehat{\tau}_p^b = \frac{1}{E}\sum_{\eta=1}^{E} \mathbb{1}[\widehat{\beta}_p^{\eta,b} \neq 0].$$

This quantity measures the proportion to which a variable has been selected in E first models provided by the lasso regularization path. We define AW for covariate p according to the B-vector $\widehat{\boldsymbol{\tau}}_p$ as:

$$w_p^{cisl} = \begin{cases} \dfrac{1}{B} & \text{if} \quad \forall b \in \{1,...,B\}\ \widehat{\tau}_p^b > 0 \\ \infty & \text{if} \quad \forall b \in \{1,...,B\}\ \widehat{\tau}_p^b = 0 \\ 1 - \dfrac{1}{B}\sum_{b=1}^{B} \mathbb{1}[\widehat{\tau}_p^b > 0] & \text{otherwise.} \end{cases}$$

Thus, the more $\widehat{\boldsymbol{\tau}}_p$ is non-null over the B subsamples, the smaller is its associated AW.
In the following, we refer to these approaches as adapt-bic and adapt-cisl.

2.4 Propensity score approaches

The propensity score (PS) is defined as the probability of being exposed to a drug of interest given the observed covariates. [30] It is a balancing score, which means that conditionally on the PS, treatment exposure and the observed covariates are independent, so it is possible to deal with measured confounding. Recently, this methodology was extended to the high-dimensional context to exploit large healthcare databases. In this framework, covariate selection algorithms are used to automatically select potential confounders for inclusion in the PS estimation model of a given drug exposure. [31]
In Courtois et al., [17] several PS-based approaches were proposed in the context of signal detection from spontaneous reporting data. These approaches consisted in estimating a PS for each drug reported in the database. The PSs were built by selecting among all the other drugs those to be included in the PS estimation model. Because of the large number of candidate covariates to be included in these models, covariate selection algorithms were used and compared. Here we used the lasso-bic approach presented in section 2.1.1. Following Courtois et al, we accounted for these PSs in the final regression model through adjustment and weighting with two different weightings for the latter: Inverse Probability of Treatment Weighting (IPTW) [32] and Matching Weights (MW). [33] We also investigated the weights

truncation approach with IPTW. This consists in assigning to individuals whose corresponding weight is below the rth percentile or above the $(1-r)$th percentile of weights, the value of the rth or $(1-r)$th percentile, respectively. [34] Here we chose to set $r = 2.5\%$. For a given PS-based approach, each drug was evaluated using one-sided hypothesis testing. To account for multiple testing, we used the procedure proposed by Benjamini and Yekutelli [35] to control the FDR under arbitrary dependence assumptions. We set the FDR level at 5%. In the following, we refer to the adjustment on the PS, the weighting on the PS with weights IPTW, IPTW with truncation, and MW as ps-adjust, ps-iptw, ps-iptwT and ps-mw, respectively.

3 Simulation study

We performed a simulation study to assess the performances of the proposed adaptive lasso strategies and to compare all the methods described in section 2 with respect to FDR, which is our primary criterion of interest, and sensitivity across a wide range of data-generating scenarios.

3.1 Comparison set-up

We simulated the occurrence of a given AE according to a logistic regression model $y_i \sim \text{Bernoulli}(\alpha_i)$ with $\alpha_i = \frac{1}{1+\exp(-\beta_0 - \sum_{p=1}^{P} \beta_p x_{ip})}$. As for the drug exposure matrix, we used the French pharmacovigilance database for the period 2000-2017 which contains 452 914 individual reports and 2 378 different drugs (see section 4.1 for a description of the data). For each replication of each scenario, we randomly selected 100 000 individual reports. For each of these datasets, we randomly selected a subset of 500 drugs among those reported more than 10 times. We investigated 27 scenarios that differed according to:

- the value of the intercept β_0, the latter being used to simulate outcomes of varying scarcity $\beta_0 = -2, -4, -6$;
- the number of drugs associated with the outcome: $n_{TP} = 0, 5, 20$;
- the value of the regression coefficients for the n_{TP} true predictors: $\beta_{TP} = 1, 2$;

- the reporting frequency of the true predictors (if any): frequent (at least 100 reports) or rare (between 20 and 100 reports).

Note that for each scenario, the n_{TP} true predictors were chosen randomly for each of the 500 replications.

In order to measure the relevance of using the BIC with the adaptive lasso, we included in the comparison a method based on the same AWs as adapt-univ but instead of cross-validation, we used the BIC to perform variable selection in the adaptive lasso stage. In the following we refer to this approach as adapt-univ-bic. For the sake of clarity, Table 1 summarizes all the implemented signal detection approaches based on the adaptive lasso. For each approach, it details how the AWs are obtained and what variable selection method is used in the adaptive lasso stage.

Method	Construction of AWs	Criterion for variable selection in adaptive lasso stage
adapt-cv	lasso-cv	cross-validation
adapt-univ	univariate coefficients	cross-validation
adapt-univ-bic	univariate coefficients	BIC
adapt-bic	lasso-bic	BIC
adapt-cisl	CISL	BIC

Table 1: Characteristics of signal detection approaches based on adaptive lasso. AWs: Adaptive Weights.

We declared as signals all drugs positively associated with the outcome for all the lasso and adaptive lasso-based approaches. For PS-based approaches, we applied a supplementary filter by considering only drugs which had more than three reports in common with the outcome. Drugs discarded by the filter had their associated p-value set to one. [36] For the sake of completeness, we also included a disproportionality method in the comparison: the Reporting Fisher's Exact Test (RFET). [5] Compared to the the more classical Reporting Odds Ratio (ROR) and Proportional Reporting Ratio (PRR), the RFET does not rely on asymptotic assumptions which are often not met given the low number of observed counts. As for the PS-based approaches, RFET was implemented on drugs with more than three reports in common with the outcome, and one-sided p-values were considered. We applied the multiple testing correction procedure to RFET presented in section 2.4.

In total, we compared 14 signal detection approaches: one disproportionality method, four lasso-based approaches, five adaptive lasso-based approaches and four PS-based approaches. All these approaches (except RFET) are implemented in the R package `adapt4pv` available on the CRAN. All the analyses were performed with R version 3.6.0. All the logistic regressions were computed with the `speedglm` R package v0.3-2 designed to handle sparse matrices efficiently. All lasso regressions were implemented using the `glmnet` R package v3.0-2.

3.2 Results

Table 2 shows the average number of drug covariates and the average number of true predictors kept after discarding covariates with fewer than three reports in common with the simulated outcome per scenario. Table 2 also shows the average number of cases per scenario. As the number of cases decreased, the number of covariates retained after filtering decreased, which also included true predictors. When the true predictors were rarely reported and the outcome was particularly rare, there were no true predictors retained after filtering (scenarios 24, 25, 26).

We first compared the performances of our proposed approaches versus the other adaptive lasso-based approaches. Figure 1 shows the average FDR and sensitivity of the approaches listed in Table 1 for scenarios 1 to 15, i.e. scenarios in which there are no true predictors (scenario 1-3) and scenarios with true predictors frequently reported (scenario 4-15). Adapt-cv and adapt-univ showed a high sensitivity at the cost of a very high FDR, especially for adapt-univ. By comparing adapt-univ and adapt-univ-bic, we see that using the BIC to perform variable selection in the adaptive lasso stage made it possible to reduce the FDR significantly across all the scenarios. By comparing approaches which rely on BIC in the adaptive lasso stage, namely adapt-univ-bic, adapt-bic and adapt-cisl, we see that approaches based on our proposed AWs performed better overall, since they showed both a lower FDR and a higher sensitivity than adapt-univ-bic. These differences in performance were particularly noticeable in scenarios 4, 8, 10, and 13 to 15. Nonetheless, adapt-univ-bic had a slightly lower FDR than adapt-bic and adapt-cisl when there were no true predictors (scenarios 1-3), with an FDR around 0.10 for adapt-cisl and adapt-bic, and around 0.05 for adap-univ-bic.

Scenario	β_0	n_{TP}	β_{TP}	Average number of covariates retained after filtering	Average number of true predictors retained after filtering	Average number of cases
No true predictors						
1	-2	0	0	389	NA	11 923
2	-4	0	0	165	NA	1 798
3	-6	0	0	30	NA	247
True predictors reported more than 100 times						
4	-2	5	1	394	5	12 327
5	-2	5	2	400	5	12 937
6	-2	20	1	407	20	13 589
7	-2	20	2	424	20	15 956
8	-4	5	1	172	5	1 879
9	-4	5	2	184	5	2 083
10	-4	20	1	193	20	2 155
11	-4	20	2	237	20	3 121
12	-6	5	1	34	2	259
13	-6	5	2	44	4	294
14	-6	20	1	49	11	300
15	-6	20	2	95	18	507
True predictors reported between 20 to 100 times						
16	-2	5	1	391	5	11 958
17	-2	5	2	392	5	12 013
18	-2	20	1	396	20	12 064
19	-2	20	2	400	20	12 284
20	-4	5	1	167	2	1 805
21	-4	5	2	171	4	1 822
22	-4	20	1	175	8	1 827
23	-4	20	2	191	17	1 897
24	-6	5	1	30	0	248
25	-6	5	2	31	0	251
26	-6	20	1	31	0	251
27	-6	20	2	35	2	263

Table 2: Average number of drug covariates/ true predictors retained after filtering and average number of cases according to scenario settings.

Among our proposals, adapt-cisl generally performed better than adapt-bic with a lower FDR and a slightly higher sensitivity, in particular in scenarios 12 to 15 when the outcome was rare. In these scenarios, the FDR of adapt-cisl ranged from 0.01 to 0.10 and its sensitivity ranged from 0.06 to 0.72, while the FDR of adapt-bic ranged from 0.03 to 0.12 and its sensitivity ranged from 0.03 to 0.68. Simulation results for scenarios 16 to 27, i.e. for true predictors reported between 20 and 100 times, are shown in Figure 2 for all these approaches. The same behaviours were observed in terms of false discoveries when the true predictors were rare. Unsurprisingly, all the approaches showed a lower sensitivity in these scenarios

compared to scenarios 4 to 15. Adapt-univ-bic showed a lower FDR when the true predictors were less reported but overall our proposals adapt-cisl and adapt-univ performed the best, with an FDR that remained low and a good sensitivity, as in scenarios 1 to 15.
In supplementary materials, Table S1 shows the average number of signals generated by adapt-cv, adapt-univ, adapt-univ-bic, adapt-bic, adapt-cisl across all the scenario settings. For all these approaches, as the outcome and true predictors became rarer, the number of signals generated decreased. Adapt-cv and adapt-univ generated far more signals on average than adapt-univ-bic, adapt-bic and adapt-cisl. The latter generated about the same number of signals, which was close to the number of true predictors.

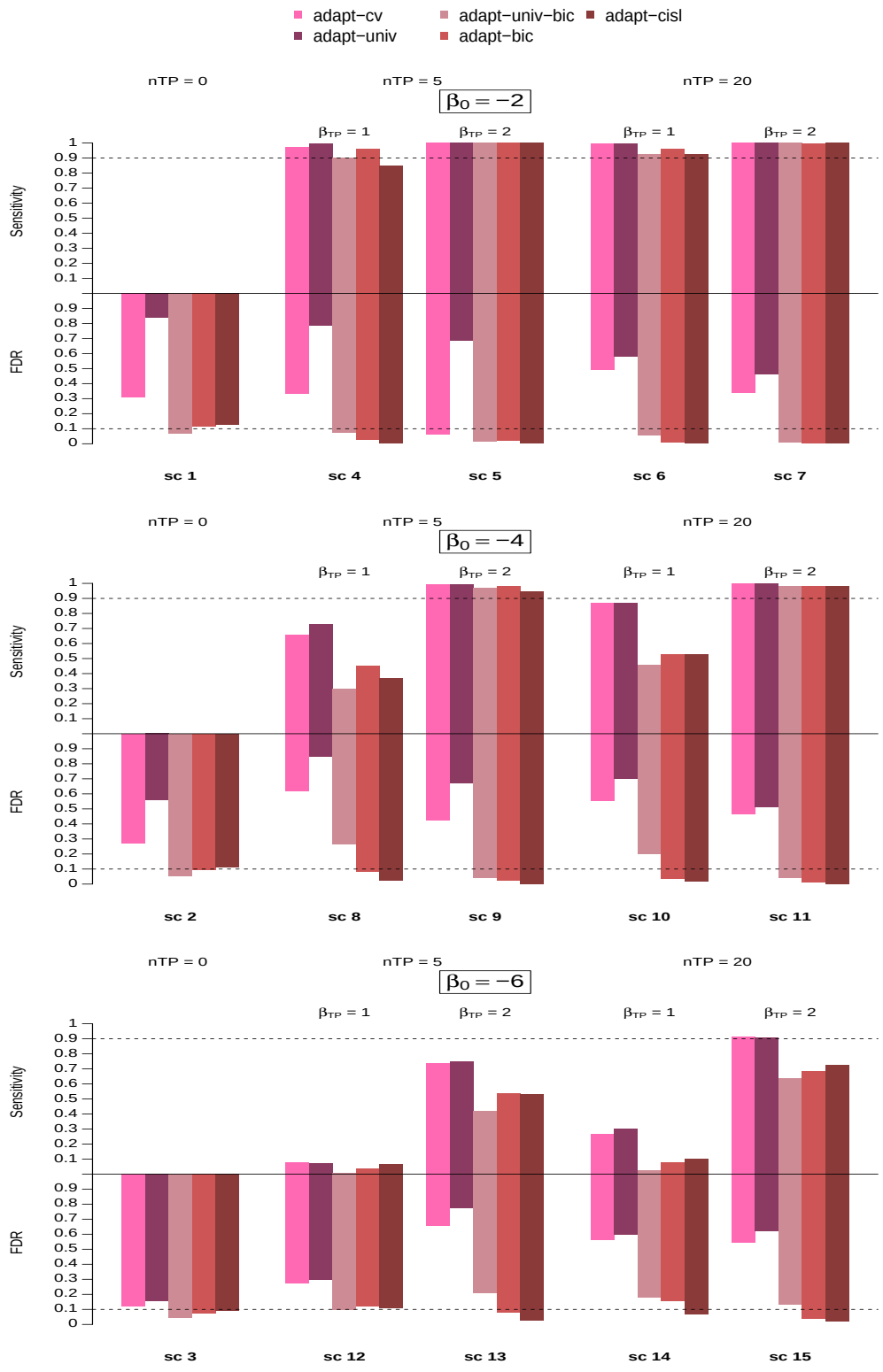


Figure 1: Sensitivity and False Discovery Rate of signal detection approaches based on adaptive lasso across scenarios 1 to 15.

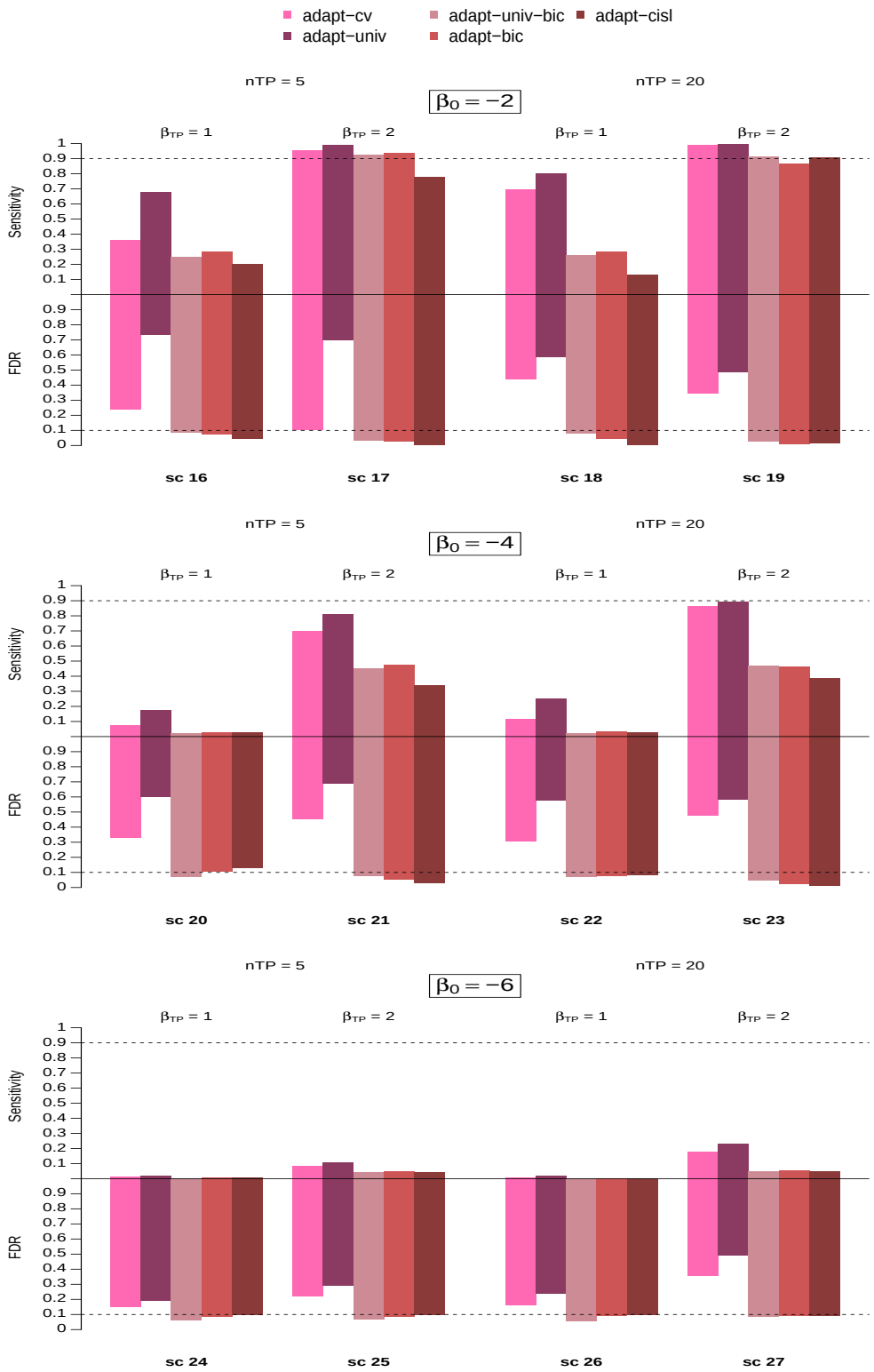


Figure 2: Sensitivity and False Discovery Rate of signal detection approaches based on adaptive lasso across scenarios 16 to 27.

Since adapt-cisl and adapt-bic showed the best performances among the adaptive lasso-based methods, we retained only these two methods for the remaining comparisons with the other

approaches. Figure 3 shows the average FDR and sensitivity of our two approaches, RFET, lasso-based and PS-based approaches when true predictors were reported more than 100 times (scenarios 4-15). Figure 4 shows the same results when true predictors were reported between 20 and 100 times (scenarios 16-27). Results for scenarios where there were no true predictors (1-3) are shown in Table 3. Lasso-cv, ps-iptwT had the best performances in terms of sensitivity but they both showed a high FDR across all the scenarios. To a lesser extent, RFET and ps-adjust showed the same behaviour in scenarios 7,11, 15 with an FDR up to 0.62 for RFET and up to 0.34 for adjust-ps. RFET also had this behaviour in scenarios 4 to 6. In all other scenarios, they both showed a rather low FDR. Overall, ps-adjust performed better than RFET both in terms of sensitivity and FDR in scenarios 4 to 15, but RFET reached a lower FDR in scenarios 16 to 27. On the other hand, ps-mw was very conservative: its FDR remained very low across the different scenarios, sometimes even much lower than the expected 5% fixed threshold. Its sensitivity dropped and became null as soon as the outcome became rarer and the true predictors reporting frequencies decreased (scenarios 12-15, 16, 18 and 20-27). The ps-iptw approach performed very poorly across all the scenarios with a very low sensitivity and an extremely high FDR. The lasso-based approaches other than lasso-cv showed good performances. Among them, lasso-perm performed worse with a high FDR when there were no true predictors (scenarios 1-3) with an FDR around 0.35, or when the outcome was rare, both for frequent and rare true predictors (scenarios 12-15 and 24-27) with an FDR up to 0.38. CISL and lasso-bic showed very good performances with both an acceptable sensitivity and a low FDR in most scenarios. When the outcome was rare, i.e. $\beta_0 = -6$, CISL showed an increase in its FDR. This increase was noticeable in scenarios 3, 12 and especially in scenarios 24 to 26, where CISL showed an FDR between 0.15 and 0.20. Overall, lasso-bic had a slightly higher sensitivity and FDR than CISL. Although lasso-bic had a fairly stable behaviour, it showed a surprising increase in its FDR in scenarios 14 and 20, with an FDR at 0.15 and 0.10, respectively.

Compared to lasso-bic, our proposals showed an equivalent or lower sensitivity and a lower FDR in all scenarios. In particular, this difference in FDR was noticeable in scenarios 12 to 15 comparing adapt-cisl to lasso-bic. Like lasso-bic, adapt-bic and adapt-cisl showed an increase

in terms of FDR for scenario 20 with an FDR of 0.12 for adapt-cisl and 0.10 for adapt-bic.

Methods	False Discovery Rate		
	scenario 1: $\beta_0 = -2$	scenario 2: $\beta_0 = -4$	scenario 3: $\beta_0 = -6$
RFET	0.00	0.01	0.00
lasso-cv	0.31	0.28	0.12
lasso-bic	0.12	0.10	0.07
lasso-perm	0.37	0.41	0.35
CISL	0.06	0.04	0.18
adapt-bic	0.11	0.10	0.07
adapt-cisl	0.13	0.11	0.09
ps-adjust	0.01	0.15	0.14
ps-mw	0.00	0.00	0.00
ps-iptw	0.99	1.00	1.00
ps-iptwT	0.23	0.35	0.29

Table 3: Average number of generated signals and False Discovery Rate of all signal detection approaches across scenarios 1 to 3, e.g. for scenarios where $n_{TP} = 0$.

In supplementary materials, Table S2 and Table S3 show the average number of signals generated by RFET, the lasso-based and the PS-based approaches across all the scenario settings. The number of generated signals for all approaches considered here decreased with the scarcity of the outcome and true predictors. This was particularly the case for ps-mw which did not generate any signals for scenarios 20 to 27. All the approaches except ps-iptw had a low number of generated signals on average when $n_{TP} = 0$. For all other scenarios, the average number of signals generated was consistent with the observed performance in terms of true and false discoveries. Approaches such as lasso-cv, ps-iptwT, ps-iptw, RFET and ps-adjust generated too many signals compared to the number of true predictors, particularly when they were highly reported (scenarios 4-15). Lasso-bic, lasso-perm, CISL and to a lesser extent ps-mw behaved like our proposals by generating a number of signals close to the number of true predictors across all the settings.

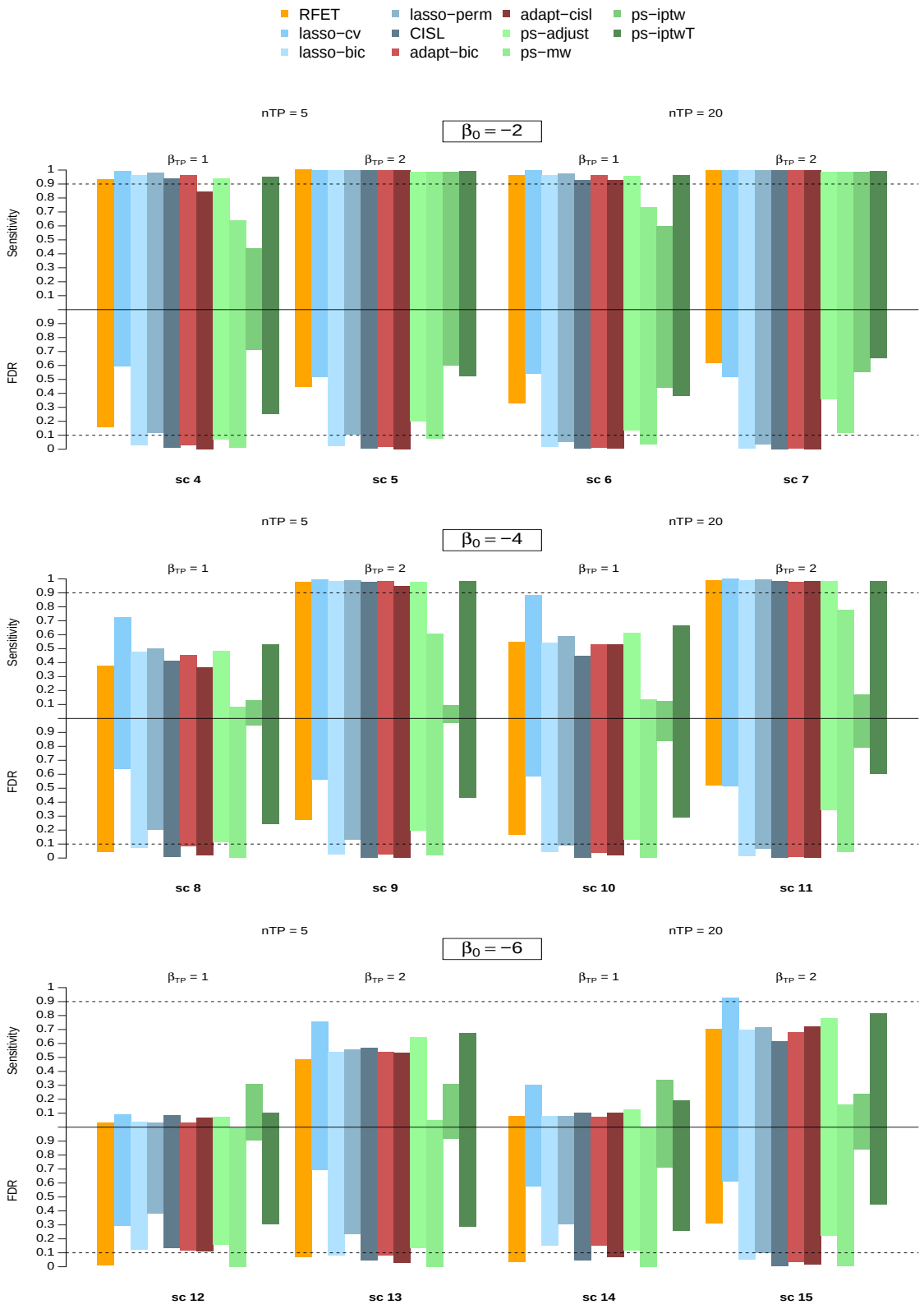


Figure 3: Sensitivity and False Discovery Rate of all signal detection approaches across scenarios 4 to 15.

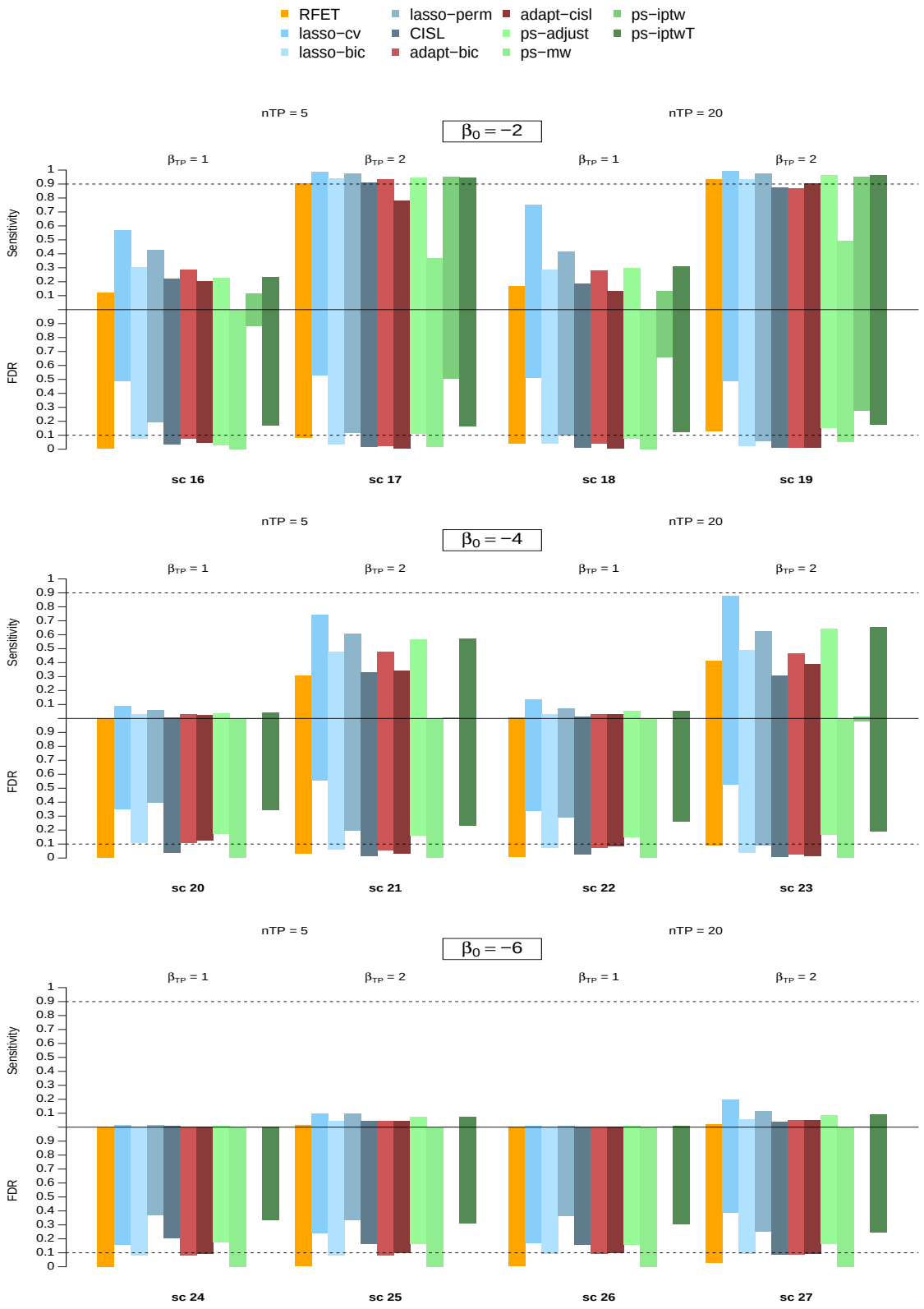


Figure 4: Sensitivity and False Discovery Rate of all signal detection approaches across scenarios 16 to 27.

4 Real-world data analysis

4.1 The French pharmacovigilance database

We applied the aforementioned signal detection approaches to the French pharmacovigilance data extracted from 1 January 2000 to 29 December 2017. We discarded spontaneous reports involving (i) drugs recorded as vaccines, phytotherapy, homeotherapy, dietary supplements, oligotherapy or enzyme inhibitors, (ii) reactions recorded as overdoses or medication errors. Drugs are listed according to their active substance which is coded with the 5th level of the Anatomical Therapeutic Chemical (ATC) hierarchy. AEs are coded according to the Preferred Term (PT) level of the Medical Dictionary for Regulatory Activities (MedDRA). This extraction of the French pharmacovigilance database included 452 914 reports with 6 617 different AEs and 2 378 different drugs.

4.2 Comparison set-up

To assess the performances of these approaches, we used a reference signal set pertaining to the adverse event Drug-Induced Liver Injury (DILI). [28, 29] The set was established by text-mining the FDA-approved drug labels with a list of keywords related to the DILI event. A level of DILI severity was assigned to each keyword: mild, moderate or severe DILI. According to where keywords appeared in the labelling section of the FDA-approved drug labels, drugs were classified in two DILI-related categories: "less-DILI-concern" and "most-DILI-concern". If no keywords were found in the label, drugs were considered as "no-DILI-concern". The majority of "most-DILI-concern" drugs were associated with severe DILI. This classification was refined later to assess the causal relationship between each drug and a DILI event using other data sources. Only drugs confirmed as a cause of DILI were retained. We translated the list of keywords used to define a DILI event into Preferred Terms (PT) codes from the MedDRA classification. If a spontaneous report involved at least one of the PT codes, it was considered as a reported DILI event. This resulted in considering 25 187 DILI reports in the French pharmacovigilance database. We considered the "no-DILI-concern" drugs as true negatives, and the "most-DILI-concern" drugs as true positives.

Over the study period, the database consisted of 1 692 different drugs reported more than 10

Table 4: Performance of each method in terms of number of signals, False Discovery Proportion (FDP), specificity and sensitivity. Operating characteristics are calculated based on drugs with known status.

Method	Number of generated signals	Number of signals with known status (positive or negative)	Number of false positive signals	FDP (%)	Specificity (%)	Sensitivity (%)
RFET	249	82	13	15.9	93.6	51.9
lasso-cv	220	79	5	6.3	97.5	55.6
lasso-bic	170	65	5	7.7	97.5	45.1
lasso-perm	158	60	4	6.7	98.0	42.1
CISL	109	48	2	4.2	99.0	34.6
adapt-cv	188	70	5	7.1	97.5	48.9
adapt-univ	179	65	4	6.2	98.0	45.9
adapt-univ-bic	163	61	4	6.6	98.0	42.9
adapt-bic	151	60	4	6.7	98.0	42.1
adapt-cisl	153	60	2	3.3	99.0	43.6
ps-adjust	209	71	6	8.5	97.0	48.9
ps-mw	86	40	0	0.0	100.0	30.1
ps-iptw	49	15	3	20.0	98.5	9.0
ps-iptwT	260	84	12	14.3	94.1	54.1

times. Among these drugs, 1 136 had more than three reports in common with a DILI. As in the simulation work, RFET and all the PS-based signal detection approaches were implemented on these 1 136 drugs and the remaining 556 had their p-value set to one. In the end, the DILI reference signal set contained 203 true negative controls and 133 true positive controls among the 1 692 drugs. Of the 1 136 drugs retained after filtering, the reference signal set contained 123 true negative controls and 119 true positive ones.

4.3 Results

Table 4 summarizes the results of all the methods in terms of generated signals, False Discovery Proportion (FDP), specificity and sensitivity derived from the DILI reference signal set. Despite the wide variability in terms of number of generated signals, we observe that 10 methods out of 14 achieved a rather comparable balance between false positives and sensitivity as regards the reference set. As in the simulations, adapt-cisl and adap-bic showed good performance in terms of false discoveries, at the cost of lower sensitivity. Some methods such as lasso-cv and adapt-univ showed better performance than in the simulations. Among all the compared methods, adapt-cisl showed the best performances with only two false positives out of 60 signals with known status.

Figure 5-A shows the overlap between signals generated by adapt-cisl, adapt-bic and lasso-bic

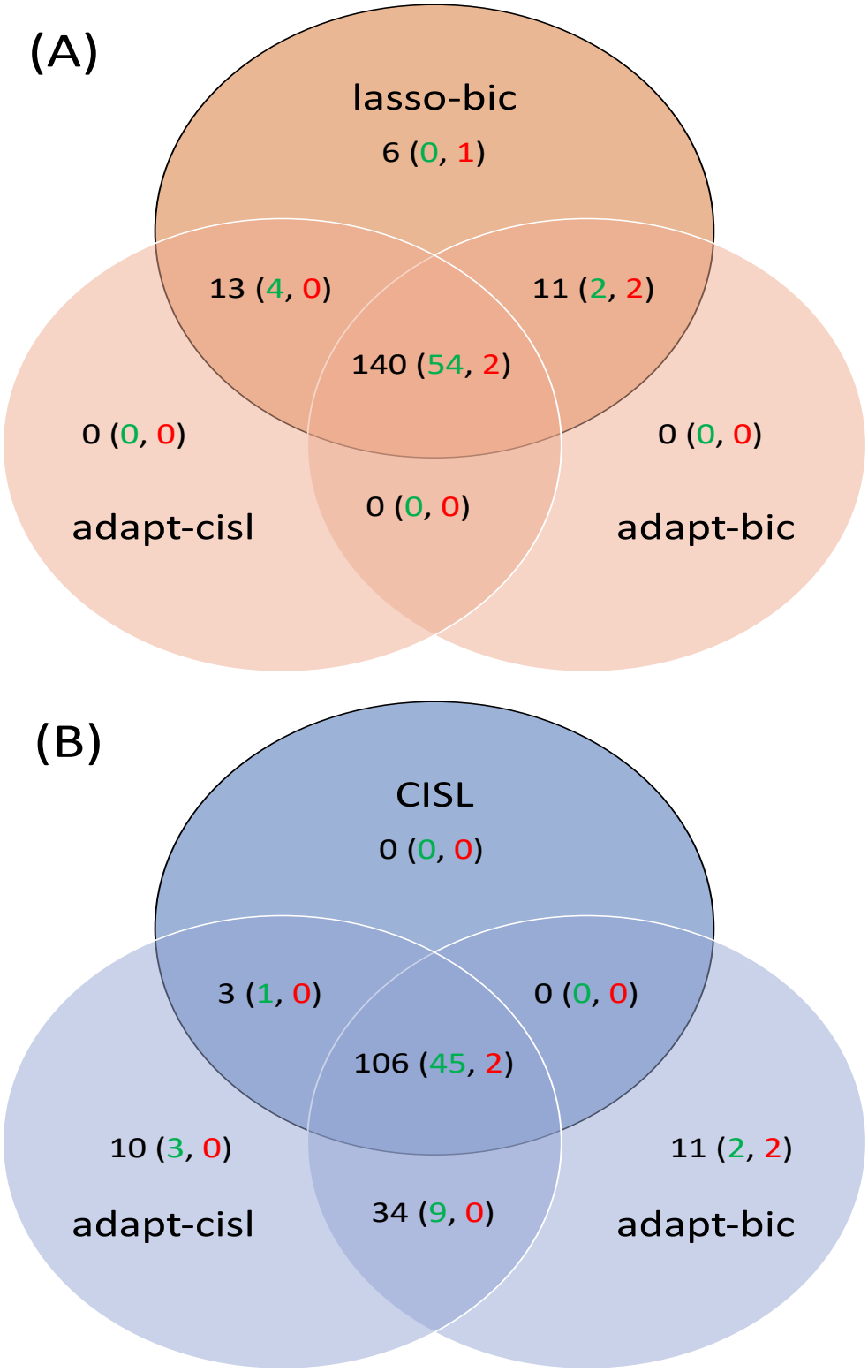


Figure 5: Distribution of signals generated by adapt-cisl, adapt-bic and **(A)** lasso-bic; **(B)** CISL. Among signals generated, true positives are in green and false positives in red.

and Figure 5-B shows this overlap for adapt-cisl, adapt-bic and CISL. Among the signals generated, true positives and false positives according to the reference set are also represented. Figure 5-A shows that all the signals generated by our two proposals were also generated by lasso-bic: 140 signals were common to the three methods, 13 were generated by adapt-cisl and lasso-bic, and 11 were generated by adapt-bic and lasso-bic. Six signals were generated by lasso-bic only, of which none were known to be positive and one was a known negative. Among the signals generated only by adapt-cisl or adapt-bic and common to lasso-bic, adapt-cisl generated four true signals and no false positive, whereas adapt-bic was a little less efficient with two true positives and two false positives. There were 54 true positives and two false positives among the 140 signals generated by the three methods. Figure 5-B shows that all signals generated by CISL were also generated by adapt-cisl with 106 signals common for the three methods (CISL, adapt-cisl, adapt-bic), three in common between CISL and adapt-cisl, and 10 additional signals generated by adapt-cisl only with three true positives and no false positives. Adapt-bic did not share any signals with CISL only and it generated 11 signals on its own with two true positives and two false positives, i.e. the same generated by lasso-bic. Overall, adapt-cisl performed well since its only two false positives among associations with known status were shared with the three other methods and no additional false positives occurred by itself. This was not the case lasso-bic and adapt-bic.

5 Discussion

The development of novel signal detection methods is crucial for improving the responsiveness and the efficiency of post-marketing surveillance systems. In this work we propose new approaches for signal detection based on an appropriate methodology for variable selection: the adaptive lasso. In addition to defining adaptive penalty weights in the context of high dimension, we used the BIC to perform variable selection. To assess the performances of our strategies, we performed an extensive simulation study conducted for multiple scenario configurations and an application to real data, where we compared our approaches to other implementations of the adaptive lasso in high dimension found in the literature, as well as to other detection approaches recently proposed based either on lasso regression or on PSs. We

also developed an R package available on CRAN that implement all the methods compared in the present work.

By comparing all the adaptive lasso-based approaches including our two proposals, adapt-bic and adapt-cisl, we first demonstrate that our defined AWs and the use of the BIC for variable selection are relevant for signal detection. Then, a broader comparison that includes state-of-the-art signal detection approaches shows that our proposals are particularly competitive. Our approaches show very satisfactory performance in terms of false discoveries, which is our main criterion of interest. Compared to lasso regression where BIC is used to perform variable selection, an approach we called lasso-bic here, our proposals tend to show a lower FDR at the cost of a slightly lower sensitivity. For adapt-bic, this result is not surprising since by construction, the covariates selected by this approach are a sub-sample of the covariates selected by the lasso-bic approach.

Our work also confirms that CISL is a relevant signal detection approach. The choice of the quantile, which we set at 10%, seems appropriate for a large number of settings, except when the outcome and true predictors are rare. As expected, cross-validation is not appropriate for signal detection in light of the poor performance in terms of FDR of all the approaches based on this criterion: lasso-cv, adapt-cv and adapt-univ. The use of the permutation method with lasso regression, implemented as lasso-perm, did not show fully satisfactory results with poor performance in some scenarios.

Among the PS-based approaches for signal detection, results are concordant with our previous work [17] so we further investigated them. Weighting on the propensity score with matching weights performed very well when the outcome was frequent but became very conservative as the outcome became rarer. This is a disadvantage since it is common in pharmacovigilance datasets to have very few reported outcomes. Adjustment on the PS did not achieve the best performance for signal detection since it led to a high sensitivity and quite a high FDR among several simulation settings. The ps-iptw approach showed very poor performances in all settings. As discussed in our previous work, these results can be explained by a potential numerical instability of weights, as already reported in the literature. [37] Performing truncation of those weights improves these results, but it is still an unsatisfactory signal

detection approach since it leads to a significant number of false discoveries.
Overall, the results obtained in the application to real data confirm those obtained in simulations. Among all the approaches tested, our adapt-cisl approach showed the best compromise between a very low FDR, an acceptable sensitivity and a reasonable number of generated signals.

Using the BIC as a criterion to select the penalisation parameter in lasso regression for variable selection has been widely studied. [38, 39, 40, 41] In particular, Chen and Chen [42] defined the extended Bayesian Information Criterion (eBIC), which is suitable for model selection in large model spaces. With this criterion, a term is added to the original BIC to correct for the prior probability of the different possible models in order to promote small dimension models. The BIC is a particular case of the eBIC. Chen and Chen showed that this criterion is particularly relevant in the large-P-small-N configuration. Considering that here we are in the $P << N$ situation, implementing the original BIC to perform variable selection seemed reasonable. In the case of the adaptive lasso, Hui et al. [43] developed the Extended Regularized Information Criterion (ERIC) to perform variable selection. They argued that the BIC cannot account for the prior information carried by the AWs. In their work, they considered the BIC defined with the penalised log likelihood. By using the BIC which is based on the unpenalized likelihood, we avoid some of the issues raised by Hui et al. However, it would now be interesting to compare the variable selection for the adaptive lasso operating with the original BIC versus ERIC in our context.
To preserve the specificities of pharmacovigilance datasets, i.e. a large size and sparsity, we based our simulation work on real data. This strategy has already been used in the literature to simulate large health care data. [34] We varied the number and the frequency of true predictors and their strength of association with the outcome, the rarity of which we also varied. However, with this strategy, it was not possible to vary the correlation structure between the true predictors and the other variables. An improvement would be to develop a more complex strategy of data generation to allow this parameter to be varied.
A major issue in the development of signal detection methods is the lack of reliable and sufficiently large sets of reference signals to evaluate performance in real-life conditions. Here

we considered a set of reference signals pertaining to a common adverse event: DILI. Although this set is very broad, performance can only be evaluated on signals with known status. As a consequence, it may result in an underestimation of the FDR.

The simulation and the application results suggest that the adaptive lasso is an appealing methodology for pharmacovigilance when the adaptive penalty weights are cleverly chosen, and when an appropriate variable selection criterion is used. Although the BIC does not make it possible to control the FDR, we are confident that it is relevant in view of the results of our simulations. Finally, our approaches do not require much more computation time than lasso-based approaches, and is still far less than the time needed for PS-based approaches. An interesting development could consist in integrated external relevant information through adaptive penalty weighting in the pharmacovigilance context. Under the weighted lasso designation, this technique has proved to be very attractive in an area such as genomics for improving penalised regression performances in terms of prediction and variable selection. [44, 45]

5.1 Acknowledgements

The authors thank the regional pharmacovigilance centers and the ANSM for providing the pharmacovigilance database dataset.

5.2 Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

5.3 Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Eugène P. van Puijenbroek, Andrew Bate, Hubert G. M. Leufkens, Marie Lindquist, Roland Orre, and Antoine C. G. Egberts. A comparison of measures of disproportion-

ality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiology and drug safety*, 11(1):3–10, 2002.

[2] S. J. W. Evans, P. C. Waller, and S. Davis. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*, 10(6):483–486, 2001.

[3] A. Bate, M. Lindquist, I. R. Edwards, S. Olsson, R. Orre, A. Lansner, and R. M. De Freitas. A Bayesian neural network method for adverse drug reaction signal generation. *European Journal of Clinical Pharmacology*, 54(4):315–321, 1998.

[4] W. Dumouchel. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician*, 53(3):177–190, 1999.

[5] Ismaïl Ahmed, Cyril Dalmasso, Françoise Haramburu, Frantz Thiessard, Philippe Broët, and Pascale Tubert-Bitter. False discovery rate estimation for frequentist pharmacovigilance signal detection methods. *Biometrics*, 66(1):301–309, 2010.

[6] Ismaïl Ahmed, Françoise Haramburu, Annie Fourrier-Réglat, Frantz Thiessard, Carmen Kreft-Jais, Ghada Miremont-Salamé, Bernard Bégaud, and Pascale Tubert-Bitter. Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting. *Statistics in Medicine*, 28(April):1774–1792, 2009.

[7] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[8] Lan Huang, Jyoti Zalkikar, and Ram C. Tiwari. A likelihood ratio test based method for signal detection with application to FDA's drug safety data. *Journal of the American Statistical Association*, 106(496):1230–1241, 2011.

[9] Yuxin Ding, Marianthi Markatou, and Robert Ball. An evaluation of statistical approaches to postmarketing surveillance. *Statistics in Medicine*, 2020.

[10] Mickael Arnaud, Francesco Salvo, Ismaïl Ahmed, Philip Robinson, Nicholas Moore, Bernard Bégaud, Pascale Tubert-Bitter, and Antoine Pariente. A Method for the Minimization of Competition Bias in Signal Detection from Spontaneous Reporting Databases. *Drug Safety*, 39:251–260, 2016.

[11] June Almenoff, Joseph M. Tonning, A. Lawrence Gould, Ana Szarfman, Manfred Hauben, Rita Ouellet-Hellstrom, Robert Ball, Ken Hornbuckle, Louisa Walsh, Chuen Yee, Susan T. Sacks, Nancy Yuen, Vaishali Patadia, Michael Blum, Mike Johnston, Charles Gerrits, Harry Seifert, and Karol LaCroix. Perspectives on the use of data mining in pharmacovigilance. *Drug Safety*, 28(11):981–1007, 2005.

[12] R. Harpaz, W. Dumouchel, N. H. Shah, D. Madigan, P. Ryan, and C. Friedman. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clinical Pharmacology and Therapeutics*, 91(6):1010–1021, 2012.

[13] A. Pariente, P. Avillach, F. Salvo, F. Thiessard, G. Miremont-Salamé, A. Fourrier-Reglat, and Association Française des Centres Régionaux de Pharmacovigilance (CRPV). Effect of Competition Bias in Safety Signal Generation. *Drug Safety*, 35(10):855–864, 2012.

[14] Ola Caster, G. Niklas Norén, David Madigan, and Andrew Bate. Large-scale regression-based pattern discovery: The example of screening the WHO global drug safety database. *Statistical Analysis and Data Mining*, 3(4):197–208, 2010.

[15] Ismaïl Ahmed, Antoine Pariente, and Pascale Tubert-Bitter. Class-imbalanced subsampling lasso algorithm for discovering adverse drug reactions. *Statistical Methods in Medical Research*, 27(3):785–797, 2018.

[16] Nicholas P. Tatonetti, Patrick P. Ye, Roxana Daneshjou, and Russ B. Altman. Data-Driven Prediction of Drug Effects and Interactions. *Science Translational Medicine*, 4(125), 2012.

[17] Émeline Courtois, Antoine Pariente, Francesco Salvo, Étienne Volatier, Pascale Tubert-Bitter, and Ismaïl Ahmed. Propensity Score-Based Approaches in High Dimension for Pharmacovigilance Signal Detection: an Empirical Comparison on the French Spontaneous Reporting Database. *Frontiers in Pharmacology*, 9(September):1010, 2018.

[18] Xueying Wang, Lang Li, Lei Wang, Weixing Feng, and Pengyue Zhang. Propensity score-adjusted three-component mixture model for drug-drug interaction data mining in FDA Adverse Event Reporting System. *Statistics in Medicine*, 2019.

[19] Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[20] Jianqing Fan and Runze Li. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[21] Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 72(4):417–473, 2010.

[22] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.

[23] Peter Bühlmann and Sara vand de Geer. Lasso for linear models. In *Statistics for High-Dimensional Data*, chapter 2, pages 25–33. Springer, 2011.

[24] Jian Huang, Shuangge Ma, and Cun-Hui Zhang. The Iterated Lasso for High-Dimensional Logistic Regression. *The University of Iowa, Department of Statistics and Actuarial Sciences*, pages 1–20, 2008.

[25] Jian Huang, Shuangge Ma, and Cun-Hui Zhang. Adaptive Lasso for Sparse High-Dimensional Regression Models. *Statistica Sinica*, pages 1603–1618, 2008.

[26] Kristin L Ayers and Heather J Cordell. SNP Selection in Genome-Wide and Candidate Gene Studies via Penalized Logistic Regression. *Genetic Epidemiology*, 34:879–891, 2010.

[27] Jeremy A. Sabourin, William Valdar, and Andrew B. Nobel. A permutation approach for selecting the penalty parameter in penalized model selection. *Biometrics*, 71(4):1185–1194, 2015.

[28] Minjun Chen, Vikrant Vijay, Qiang Shi, Zhichao Liu, Hong Fang, and Weida Tong. FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discovery Today*, 16(15-16):697–703, 2011.

[29] Minjun Chen, Ayako Suzuki, Shraddha Thakkar, Ke Yu, Chuchu Hu, and Weida Tong. DILIrank: The largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discovery Today*, 21(4):648–653, 2016.

[30] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrilca*, 70(1):41–55, 1983.

[31] Sebastian Schneeweiss, Jeremy A. Rassen, Robert J. Glynn, Jerry Avorn, Helen Mogun, and M. Alan Brookhart. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*, 20:512–522, 2009.

[32] Peter C. Austin. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Beahavioral Research*, 46(3):399–424, 2011.

[33] Liang Li and Tom Greene. A weighting analogue to pair matching in propensity score analysis. *International Journal of Biostatistics*, 9(2):215–234, 2013.

[34] Jessica M. Franklin, Wesley Eddings, Peter C. Austin, Elizabeth A. Stuart, and Sebastian Schneeweiss. Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. *Statistics in Medicine*, 36(12):1946–1963, 2017.

[35] Yoav Benjamini and Daniel Yekuteli. The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.

[36] Nicolai Meinshausen, Lukas Meier, and Peter Bühlmann. P-Values for High-Dimensional Regression. *Journal of the American Statistical Association*, 104(488):1671–1681, 2009.

[37] Kazuki Yoshida, Sonia Hernández-Díaz, Daniel H. Solomon, John W. Jackson, Joshua J. Gagne, Robert J. Glynn, and Jessica M. Franklin. Matching weights to simultaneously

compare three treatment groups: Comparison to three-way matching. *Epidemiology*, 28:387–395, 2017.

[38] Hansheng Wang, Bo Li, and Chenlei Leng. Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71(3):671–683, 2009.

[39] Tao Wang and Lixing Zhu. Consistent tuning parameter selection in high dimensional sparse linear regression. *Journal of Multivariate Analysis*, 102(7):1141–1151, 2011.

[40] Hui Zou, Trevor Hastie, and Robert Tibshirani. On the Degrees of Freedom of the Lasso. 2012.

[41] Yingying Fan and Cheng Yong Tang. Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 75(3):531–552, 2013.

[42] Jiahua Chen and Zehua Chen. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771, 2008.

[43] F. K. Hui, D. I. Warton, and S. D. Foster. Tuning parameter selection for the adaptive lasso using ERIC. *Journal of the American Statistical Association*, 110(509):262–269, 2015.

[44] Linn Cecilie Bergersen, Ingrid K. Glad, and Heidi Lyng. Weighted lasso with data integration. *Statistical Applications in Genetics and Molecular Biology*, 10(1), 2011.

[45] Tonje G. Lien, Ørnulf Borgan, Sjur Reppe, Kaare Gautvik, and Ingrid Kristine Glad. Integrated analysis of DNA-methylation and gene expression using high-dimensional penalized regression: A cohort study on bone mineral density in postmenopausal women. *BMC Medical Genomics*, 11(1):1–11, 2018.

6 Supplementary Materials

Scenarios	β_0	n_{TP}	β_{TP}	Methods				
				adapt-cv	adapt-univ	adapt-univ-bic	adapt-bic	adapt-cisl
No true predictors								
1	-2	0	0	0.48	9.64	0.07	0.12	0.13
2	-4	0	0	0.97	6.67	0.06	0.11	0.12
3	-6	0	0	0.51	1.04	0.05	0.08	0.09
True predictors reported more than 100 times								
4	-2	5	1	9.81	25.90	4.96	4.95	4.25
5	-2	5	2	5.65	20.56	5.09	5.12	5.00
6	-2	20	1	41.69	48.69	19.65	19.42	18.61
7	-2	20	2	34.48	38.63	20.14	20.11	20.01
8	-4	5	1	11.42	30.06	2.53	2.47	1.89
9	-4	5	2	11.23	19.81	5.12	5.04	4.75
10	-4	20	1	40.52	58.81	11.65	11.00	10.77
11	-4	20	2	39.41	42.69	20.43	19.75	19.75
12	-6	5	1	2.88	4.63	0.18	0.34	0.47
13	-6	5	2	13.63	20.08	2.99	2.98	2.76
14	-6	20	1	17.99	27.04	1.23	1.92	2.28
15	-6	20	2	41.06	49.55	14.85	14.19	14.74
True predictors reported between 20 and 100 times								
16	-2	5	1	3.65	19.01	1.45	1.57	1.09
17	-2	5	2	5.96	21.14	4.78	4.82	3.92
18	-2	20	1	28.24	40.92	5.69	5.87	2.68
19	-2	20	2	33.47	40.15	18.78	17.62	18.35
20	-4	5	1	2.14	9.59	0.21	0.27	0.27
21	-4	5	2	9.54	19.77	2.47	2.53	1.79
22	-4	20	1	5.85	19.18	0.57	0.74	0.71
23	-4	20	2	35.26	46.08	9.97	9.56	7.89
24	-6	5	1	0.89	1.54	0.08	0.11	0.12
25	-6	5	2	1.78	2.99	0.29	0.33	0.34
26	-6	20	1	0.85	1.90	0.11	0.17	0.17
27	-6	20	2	8.20	12.36	1.13	1.25	1.19

Table 1: Average number of signals generated by adaptive lasso-based approaches across all simulated scenarios

Scenarios	β_0	n_{TP}	β_{TP}	Methods				
				RFET	lasso-cv	lasso-bic	lasso-perm	CISL
No true predictors								
1	-2	0	0	0.00	1.41	0.13	0.55	0.06
2	-4	0	0	0.01	1.37	0.12	0.63	0.04
3	-6	0	0	0.00	0.70	0.08	0.55	0.20
True predictors reported more than 100 times								
4	-2	5	1	5.91	14.70	4.97	5.65	4.74
5	-2	5	2	11.81	12.53	5.16	5.70	5.05
6	-2	20	1	30.51	45.02	19.58	20.74	18.61
7	-2	20	2	57.84	43.17	20.15	20.85	20.07
8	-4	5	1	2.09	13.50	2.64	3.31	2.10
9	-4	5	2	7.87	13.57	5.07	5.86	4.90
10	-4	20	1	13.77	44.39	11.33	13.02	9.00
11	-4	20	2	45.58	42.61	20.17	21.34	19.64
12	-6	5	1	0.17	3.49	0.38	0.86	0.65
13	-6	5	2	2.83	15.82	3.01	3.82	2.99
14	-6	20	1	1.74	21.17	2.05	2.53	2.26
15	-6	20	2	22.73	49.24	14.89	15.99	12.34
True predictors reported between 20 and 100 times								
16	-2	5	1	0.63	8.51	1.69	2.81	1.15
17	-2	5	2	5.13	12.64	4.93	5.66	4.63
18	-2	20	1	3.68	32.46	6.03	9.25	3.75
19	-2	20	2	21.89	40.44	19.16	20.68	17.71
20	-4	5	1	0.02	2.82	0.28	1.00	0.08
21	-4	5	2	1.63	11.70	2.55	3.92	1.68
22	-4	20	1	0.17	7.29	0.75	2.18	0.22
23	-4	20	2	9.28	38.93	10.14	13.84	6.13
24	-6	5	1	0.00	1.19	0.11	0.70	0.25
25	-6	5	2	0.06	2.34	0.33	1.21	0.42
26	-6	20	1	0.01	1.13	0.17	0.88	0.23
27	-6	20	2	0.45	10.29	1.28	3.32	0.94

Table 2: Average number of signals generated by RFET and lasso-based approaches across all simulated scenarios

Scenarios	β_0	n_{TP}	β_{TP}	Methods			
				ps-adjust	ps-mw	ps-iptw	ps-iptwT
No true predictors							
1	-2	0	0	0.02	0.00	4.73	0.26
2	-4	0	0	0.18	0.00	13.93	0.46
3	-6	0	0	0.16	0.00	14.07	0.38
True predictors reported more than 100 times							
4	-2	5	1	5.17	3.27	7.96	6.95
5	-2	5	2	6.84	5.48	13.67	13.55
6	-2	20	1	22.47	15.24	22.08	33.37
7	-2	20	2	32.59	22.64	47.43	62.30
8	-4	5	1	2.86	0.41	14.02	3.79
9	-4	5	2	6.65	3.14	13.84	10.54
10	-4	20	1	14.39	2.77	15.03	19.73
11	-4	20	2	31.90	16.46	16.29	55.09
12	-6	5	1	0.65	0.00	16.08	1.11
13	-6	5	2	3.94	0.25	18.88	5.41
14	-6	20	1	3.13	0.03	23.65	5.35
15	-6	20	2	21.18	3.24	30.62	32.86
True predictors reported between 20 and 100 times							
16	-2	5	1	1.21	0.00	5.46	1.58
17	-2	5	2	5.60	1.89	10.29	6.01
18	-2	20	1	6.67	0.04	7.87	7.24
19	-2	20	2	23.26	10.46	26.83	24.00
20	-4	5	1	0.43	0.00	13.88	0.73
21	-4	5	2	3.57	0.00	13.89	3.96
22	-4	20	1	1.44	0.00	13.84	1.84
23	-4	20	2	15.95	0.03	13.93	16.67
24	-6	5	1	0.24	0.00	13.64	0.50
25	-6	5	2	0.61	0.00	13.83	0.89
26	-6	20	1	0.32	0.00	13.89	0.56
27	-6	20	2	2.37	0.00	14.76	2.74

Table 3: Average number of signals generated by PS-based approaches across all simulated scenarios

D.3 Résumé long de la communication orale aux 51èmes Journées de Statistique de la Société Française de Statistique.

Exploitation conjointe de plusieurs bases de données dans la détection de signaux en pharmacovigilance via l'utilisation du lasso pondéré

Émeline Courtois [1] & Ismaïl Ahmed [2] & Pascale Tubert-Bitter [3]

[1] ***UMR 1181 - Inserm : Biostatistique, Biomathématique, Pharmaco-Épidémiologie et Maladies Infectieuses***
Hôpital Paul Brousse, 16 avenue Paul Vaillant-Couturier, Villejuif
emeline.courtois@inserm.fr
[2] ***UMR 1181 - Inserm : Biostatistique, Biomathématique, Pharmaco-Épidémiologie et Maladies Infectieuses***
Hôpital Paul Brousse, 16 avenue Paul Vaillant-Couturier, Villejuif
ismail.ahmed@inserm.fr
[3] ***UMR 1181 - Inserm : Biostatistique, Biomathématique, Pharmaco-Épidémiologie et Maladies Infectieuses***
Hôpital Paul Brousse, 16 avenue Paul Vaillant-Couturier, Villejuif
pascale.tubert@inserm.fr

Résumé. La pharmacovigilance a pour objectif de détecter le plus précocement possible les effets indésirables des médicaments commercialisés. Ce travail de détection s'apparente à une problématique de sélection de variables en grande dimension. Classiquement il s'effectue sur les bases de notifications spontanées, mais récemment un intérêt croissant s'est porté sur l'exploitation des bases médico-administratives. Nous proposons dans ce travail d'intégrer l'information issue d'une stratégie de détection réalisée à partir d'un référentiel de témoins fournis par les bases médico-administratives, dans les modèles d'analyses de la base des notifications spontanées. Cette intégration de l'information est effectuée via l'utilisation de pénalités différentiées pour chaque covariable médicament dans un lasso pondéré afin de guider la sélection de variables opérée. Ces pénalités différentiées sont obtenues à partir de deux types d'informations : des odds ratio et des p-valeurs corrigées pour les tests multiples. Les performances des méthodes basées sur le lasso pondéré sont comparées à celle d'un lasso classique et sont évaluées empiriquement grâce à un ensemble de signaux de référence concernant l'évènement indésirable "lésion hépatique aiguë".

Mots-clés. détection de signal, pharmacovigilance, intégration de données, grande dimension, lasso pondéré

Abstract. The objective of pharmacovigilance is to detect adverse reactions of marketed drugs as early as possible. This detection process is a variable selection issue. Usually it relies on spontaneous reports databases, but recently there has been a growing interest

in the use of medico-administrative databases. In this work, we propose to integrate information from a detection strategy based on controls provided by medico-administrative databases into the analysis models of the spontaneous reports database. This integration of information is achieved through the use of different penalties for each drug covariate in a weighted lasso to guide the covariates selection. These penalties are obtained from two types of information : odds ratios and p-values corrected for multiple testing. The performance of weighted lasso methods is compared to that of a conventional lasso and is empirically evaluated using a set of reference signals for the adverse event "acute liver injury".

Keywords. signal detection, pharmacovigilance, multiple data integration, high dimensional data, weighted lasso

1 Introduction

La pharmacovigilance a pour objectif de détecter le plus précocement possible les effets indésirables des médicaments mis sur le marché. Elle repose le plus souvent sur l'exploitation des bases de notifications spontanées, qui sont constituées de déclarations de la survenue d'évènements indésirables (EIs) par les praticiens de santé, et dont l'origine suspectée est médicamenteuse. A l'échelle nationale, ces notifications spontanées représentent de grands ensemble de données : sur la période 2000-2016 la base française de notifications spontanées comptabilise environ 380 000 notifications, qui comprennent 5 900 EI et 2 300 médicaments différents. Il a donc été proposé depuis une quinzaine d'années un certain nombre de méthodes statistiques de fouille de données visant à détecter des associations statistiques suspectes entre médicaments et EIs. On parle de détection automatisée de signaux, les signaux statistiques générés devant être évalués par des experts. Les méthodes classiques de détection de signaux sont les méthodes de disproportionnalités qui consistent, pour chaque couple (médicament p, EI j), à (1) déterminer le nombre attendu de notifications qui mentionnent le médicament p et l'EI j si la la mention de l'EI j est indépendante de la mention du médicament p ; (2) construire une mesure de "disproportionnalité" pour quantifier l'excès du nombre observé n_{pj} face au nombre attendu de notifications mentionnant le médicament p et l'EI j. Finalement, un signal est généré si cette mesure dépasse un certain seuil critique.

Plus récemment, des approches de détection de signaux basées sur des régressions logistiques multiples ont été proposées [1, 2]. Au lieu de considérer les données agrégées comme les méthodes de disproportionnalité, ces nouvelles approches sont directement appliquées aux notifications spontanées : l'observation devient la notification individuelle, la réponse est la présence ou l'absence d'un EI donné, et les covariables sont toutes des indicatrices de présence de médicaments. En raison du très grand nombre de covariables potentielles, ces méthodes reposent sur des versions pénalisées de régressions logistiques multiples.

Pour toutes ces approches, les expositions médicamenteuses des cas d'un EI donné sont comparées à celles d'un groupe témoin constitué par l'ensemble des individus notifiés pour un autre EI que celui d'intérêt. En effet, par construction, on ne dispose pas dans les notifications spontanées d'individus témoins représentatifs de la population générale.

En outre, depuis plusieurs années un intérêt croissant s'est porté sur l'exploitation des bases médico-administratives pour la détection de signaux en pharmacovigilance. En France, le SNDS (Système National des Données de Santé) regroupe les données des bases hospitalières et de remboursement de soins, et couvre la quasi-totalité de la population française. Il a également été créé à partir du SNDS l'Échantillon Généraliste des Bénéficiaires (EGB), échantillon au $1/97^{\text{ème}}$. La difficulté dans l'exploitation de ces grandes bases de données tient au fait qu'elles n'ont pas été conçues pour répondre à des questions de santé. Par ailleurs, les EIs pouvant être étudiés sont nécessairement graves puisque requérant une hospitalisation. De plus, malgré une taille très importante, l'exploitation de l'EGB (plus facilement accessible à la communauté des chercheurs) n'est réalisable que pour des EIs relativement fréquents.

L'objectif général de ce travail est de proposer et d'évaluer des stratégies de détection de signaux qui intègrent l'information sur l'exposition médicamenteuse de témoins issus des données de l'EGB. Cette intégration d'information est rendue possible par l'utilisation du lasso pondéré [3], dont le lasso [4] peut être vu comme un cas particulier, où l'on tient compte de pénalités individuelles pour chaque covariable. Ici, ces pénalités individuelles sont déterminées à partir des résultats d'un travail de détection de signal dans une base dite hybride, composée des cas des notifications spontanées et des témoins de l'EGB. Cette méthode d'intégration de données est comparée à une méthode basée sur un lasso classique dans la base des notifications spontanées seule, dénommée ci-après base principale. Une évaluation empirique avec une application à l'évènement indésirable "lésions hépatiques aiguës" (ALI) est réalisée, en utilisant l'ensemble de signaux de référence établi par l'Observational Medical Outcome Partnership (OMOP) [5].

2 Méthode

2.1 Lasso et lasso pondéré

En considérant un EI donné, on introduit les notations suivantes. Soit N le nombre de notifications (i.e. d'observations) et P le nombre de covariables médicaments. On note Y la variable réponse, indicatrice de la présence ou de l'absence de l'EI considéré et X_p la variable binaire de l'exposition au médicament p. Pour $i \in \{1, ..., N\}$, le modèle logistique pour l'EI considéré est défini par :

$$\text{logit}(\Pr(Y_i = 1)) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ip}. \tag{1}$$

Le nombre de covariables P étant très grand, on a recours à la régression pénalisée de type lasso, qui permet de forcer certains coefficients β_p à valoir exactement zéro. Ce type de régression consiste à maximiser la vraisemblance du modèle (1) moins un terme de pénalité défini par

$$pen(\lambda) = \lambda \sum_{p=1}^{P} |\beta_p|. \tag{2}$$

Le lasso pondéré est une généralisation de la pénalisation définie en (2) qui permet d'introduire pour chaque covariable une valeur de pénalité différente. Le terme de pénalité s'écrit dans ce cas

$$pen(\lambda) = \lambda \sum_{p=1}^{P} w_p |\beta_p|. \tag{3}$$

Au final, la pénalité appliquée à la covariable p est définie par $\lambda_p = \lambda w_p$. L'intérêt d'une telle différentiation de pénalité est de guider la sélection de variables opérée par le lasso : plus la valeur du coefficient w_p est grande, plus la variable p est pénalisée, moins la variable est susceptible d'être incluse dans le modèle. Dans la suite, on parlera de poids pour désigner le coefficient w_p et de pondération pour définir une stratégie visant à déterminer ces poids.

2.2 Pondérations proposées

Suivant l'idée de Bergensen *et al.* [3], nous proposons ici deux pondérations différentes qui permettent de prendre en compte l'information provenant d'une base de données externe. On suppose qu'à partir de cette source d'information externe, on dispose pour chaque covariable médicament p du résultat d'un test d'association entre cette covariable et la réponse. A l'issue de cet ensemble de tests, chaque covariable a une mesure d'association ainsi qu'une p-valeur associée. Comme on s'intéresse ici uniquement à des associations délétères, les test réalisés sont unilatéraux et la mesure d'association, ici l'odds ratio (OR), est jugée pertinente si elle excède 1. Pour un seuil de significativité α, seul un sous ensemble de ces tests sont significatifs. Les pondérations que nous proposons dans la suite prennent en compte cette sélection de variables.

Poids 1 : Odds-Ratio A partir des odds ratios des covariables significativement associées à la réponse au seuil α, nous proposons d'introduire les poids suivant dans le lasso pondéré.

$$w_p^{(1)} = \frac{1}{\log(\mathrm{OR}_p)}. \tag{4}$$

Ainsi, plus une covariable est associée fortement à la réponse, moins elle est pénalisée.

Poids 2 : P-valeurs Soit s le nombre de covariables qui ont une p-valeur associée inférieure à α. Nous proposons d'attribuer à ces covariables un poids dans le lasso pondéré déterminé à partir de la fonction de répartition empirique de ces p-valeurs

$$w_p^{(2)} = \frac{1}{s} \sum_{i=1}^{s} \mathbb{1}_{\text{p-valeur}_i \leq \text{p-valeur}_p}. \tag{5}$$

Avec cette transformation monotone, les covariables avec les p-valeurs les plus faibles sont très favorisées dans le processus de sélection.

Pour ces deux pondérations, les covariables non significativement associées à la réponse d'après la source externe sont exclues des covariables candidates à la sélection dans le lasso pondéré effectué dans la base principale. Les poids associés aux variables significatives sont normalisés, de telle sorte à ce que leur moyenne soit égale à 1 pour chacune des pondérations.

3 Application

3.1 Matériel

La première étape de ce travail a consisté à construire pour les cas de ALI des notifications spontanées un groupe de contrôles issus de l'EGB. Pour des raisons de disponibilité des données d'hospitalisation dans l'EGB, nous avons considéré la période 2006-2016.

L'évènement ALI est défini selon un ensemble de *Preferred Term*(PT) de la terminologie MedDRA (Medical Dictionary for Regulatory Activities) qui est utilisé pour coder les EIs dans les notifications spontanées. Un cas de ALI est un individu dont la notification fait mention d'au moins un des ces termes. Sur la période considérée, et pour les notifications où l'âge et le sexe sont renseignés, les notifications spontanées comprennent 11 618 cas et 240 500 non cas de ALI.

Les témoins de l'EGB ont été appariés aux cas des notifications spontanées sur le sexe et par classes d'âge de 10 ans. Sont considérés comme témoins les individus qui n'ont pas été hospitalisés pour un quelconque problème hépatique l'année de leur appariement ou les années antérieures. Nous avons cherché à apparier jusqu'à 100 témoins par cas, tout en interdisant le fait qu'un témoin soit apparié avec plusieurs cas. Au final, nous obtenons 525 506 témoins uniques de l'EGB. Les consommations médicamenteuses des témoins ont été regardées dans le mois précédant la date de l'EI du cas auquel ils sont appariés. En se restreignant aux expositions médicamenteuses communes aux cas et aux témoins des deux différentes bases, et qui sont notifiées plus de 10 fois dans la base hybride, nous considérons 1 020 covariables médicaments.

3.2 Calcul des poids et analyses

Sur la base hybride, nous avons implémenté une méthode de détection de signal proche des méthodes de disproportionnalité classiques : le Reporting Fisher Exact Test (RFET) [6]. Cette approche, de par l'utilisation du test exact de Fisher, est robuste aux situations où les effectifs observés de cas exposés au médicament p sont faibles. Ainsi pour chaque covariable médicament, on obtient avec cette procédure un OR ainsi qu'une p-valeur. Pour prendre en compte la multiplicité des comparaisons effectuées, nous corrigeons ces p-valeurs avec la procédure de correction de tests multiple proposée par Ahmed *et al.* [6], qui repose sur le contrôle du Fase Discovery Rate (FDR) et qui est adaptée aux tests d'hypothèses unilatérales effectués ici. Cette approche amène de fait à une sélection de variables, puisque sont considérés comme signaux dans la base hybride toutes les covariables avec une p-valeur corrigée inférieure à un seuil de significativité $\alpha = 0.05$.

A partir des deux types de pondération présentées dans la partie 2.2, nous avons appliqué dans la base principale les deux lasso pondérés correspondants ainsi qu'une régression lasso classique. Afin de s'affranchir du choix d'une pénalité optimale, nous avons comparé les résultats de ces trois régressions pénalisées à nombre de signaux générés équivalents, pour une grille de valeurs de pénalités. Avec ces méthodes, ce que l'on définit comme un signal est une covariable qui a un coefficient pénalisé strictement supérieur à zéro.

3.3 Paramétrage de la comparaison

Les performances des différentes approches de détection ont été évaluées grâce à un ensemble de signaux de référence concernant l'EI ALI défini par l'OMOP. Parmi les covariables médicament considérées, cet ensemble comprend 52 témoins positifs et 13 témoins négatifs.

4 Résultats

Le figure (1-(a)) présente la répartition des signaux générés par le lasso pondéré basé sur les OR issus de la base hybride et le lasso classique. Le détail de cette répartition est aussi présentée pour les témoins positifs (figure 1-(b)) et les témoins négatifs détectés (figure 2-(c)), et les performances de ces deux méthodes en termes de sensibilité et de proportion de fausses découvertes (FDP) sont représentées (figure 2-(d)). Les figures (3) et (4) présentent les mêmes résultats pour le lasso pondéré basé sur les p-valeurs issues de la base hybride, comparé au lasso classique.

Les performances entre les méthodes de détection basées sur le lasso pondéré et celle basée sur la lasso classique sont proches. Néanmoins les approches proposées ont une meilleure sensibilité, et tout particulièrement pour un nombre de signaux générés importants. Le lasso pondéré basé sur les OR génère son premier et unique faux positif à un

nombre de signaux plus grand que la lasso classique, et le lasso pondéré basé sur les p-valeurs ne génère pas de faux négatif. Les variables sélectionnées par les lasso pondérés sont différentes de celles sélectionnées avec le lasso classique.

5 Discussion

Les méthodes basées sur le lasso pondéré ont dans l'ensemble des performances comparables en terme de détection par rapport à une méthode basée sur une régression lasso classique. Néanmoins, les signaux générés par ces méthodes sont différents, ce qui laisse penser que l'ajout d'une information extérieure dans le processus de détection de signaux apporte une information différente voire complémentaire à celle issue d'un processus de détection plus classique.

Avec les pondérations que nous proposons, les covariables qui ne sont pas jugées significatives à l'issue d'un test basé sur une source d'information extérieure sont exclues de la sélection dans le lasso pondéré. Ce choix étant potentiellement assez fort, il convient de préciser que l'information externe est pensée comme étant une information pertinente Néanmoins, pour relâcher cette contrainte, on peut envisager de considérer une valeur de α plus grande.

Une limite avec l'utilisation du lasso pondéré dans un contexte de sélection de variable réside dans la choix optimal de la pénalité. Une solution serait d'avoir recours à la validation croisée pour déterminer le paramètre λ dans (3), bien que la pénalité définie ainsi ne soit pas la plus adaptée dans le cadre de la sélection de variable.

Références

[1] Ola Caster, G. Niklas Norén, David Madigan, and Andrew Bate. Large-scale regression-based pattern discovery : The example of screening the WHO global drug safety database. *Statistical Analysis and Data Mining*, 3(4) :197–208, 2010.

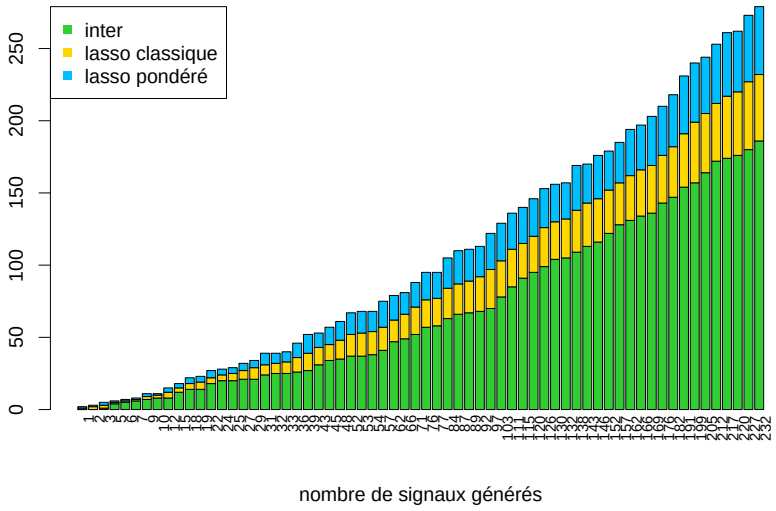
[2] Ismaïl Ahmed, Antoine Pariente, and Pascale Tubert-Bitter. Class-imbalanced subsampling lasso algorithm for discovering adverse drug reactions. *Statistical Methods in Medical Research*, 27(3) :785–797, 2018.

[3] Linn Cecilie Bergersen, Ingrid K. Glad, and Heidi Lyng. Weighted lasso with data integration. *Statistical Applications in Genetics and Molecular Biology*, 10(1), 2011.

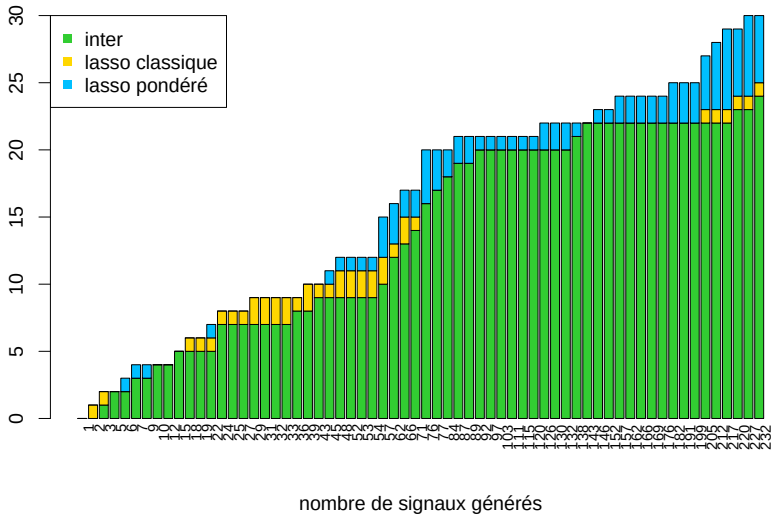
[4] Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1) :267–288, 1996.

[5] Patrick B. Ryan, Martijn J. Schuemie, Emily Welebob, Jon Duke, Sarah Valentine, and Abraham G. Hartzema. Defining a Reference Set to Support Methodological Research in Drug Safety. *Drug Safety*, 36(Suppl 1), 2013.

[6] I. Ahmed, C. Dalmasso, F. Haramburu, F. Thiessard, P. Broët, and P. Tubert-Bitter. False discovery rate estimation for frequentist pharmacovigilance signal detection methods. *Biometrics*, 2010.

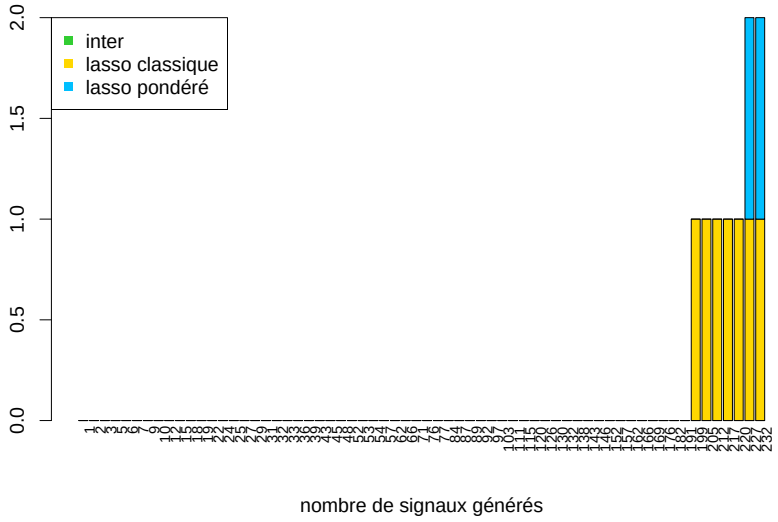


(a)

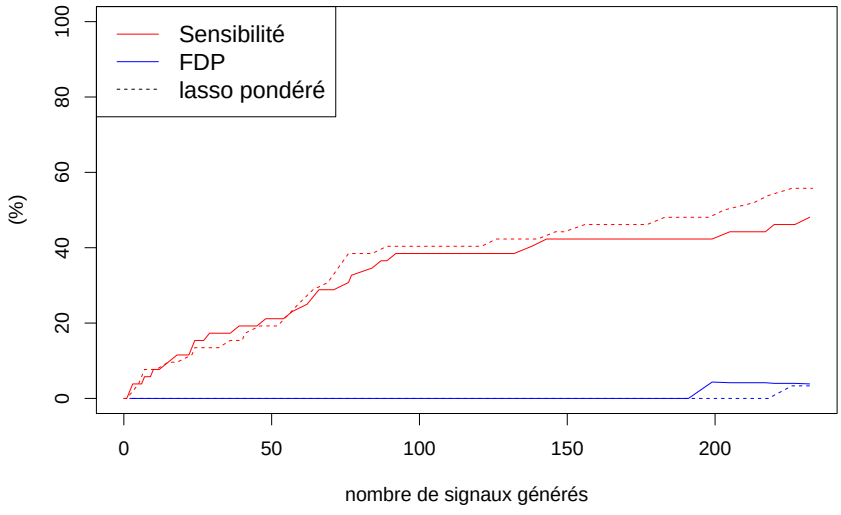


(b)

FIGURE 1 – Comparaison des performances entre le lasso et le lasso pondéré basé sur les **OR de la base hybride**. La figure (a) représente le nombre de signaux détectés par les deux méthodes (vert) ainsi que les signaux détectés uniquement par l'une des méthodes (jaune/bleu) en fonction du nombre de signaux générés. De manière analogue, la figure (b) présente le nombre de témoins positifs détectés par les deux méthodes.

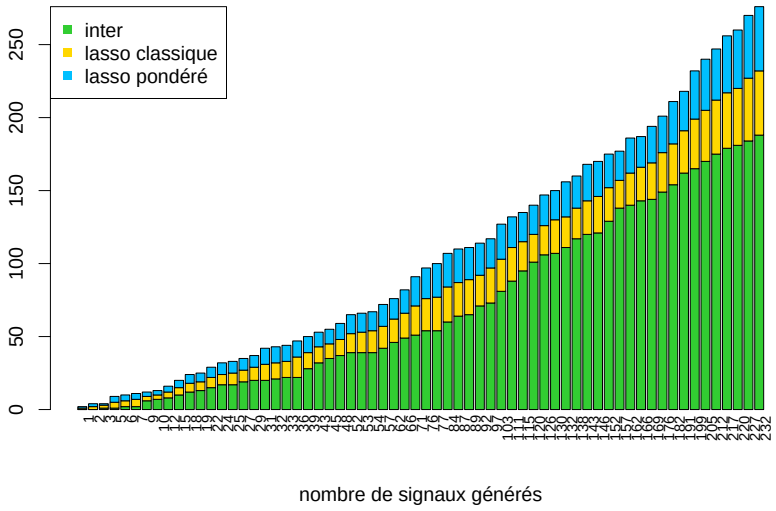


(c)

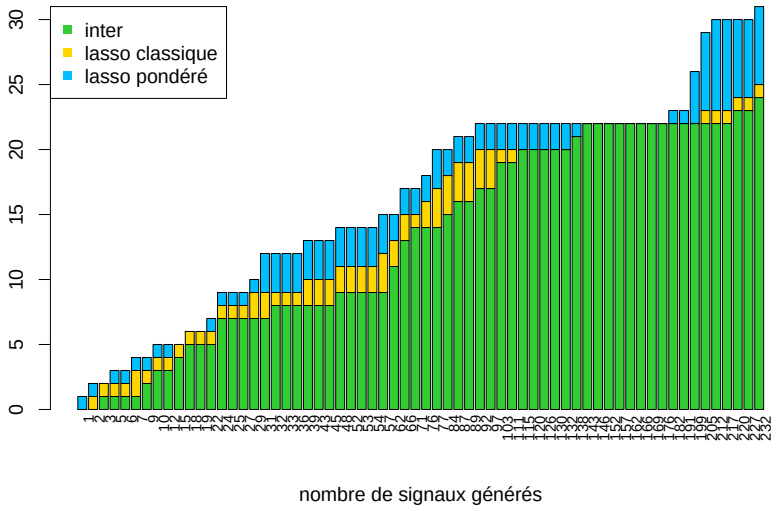


(d)

FIGURE 2 – Comparaison des performances entre le lasso et le lasso pondéré basé sur les **OR de la base hybride**. La figure (c) représente le nombre de témoins négatifs détectés par les deux méthodes (vert) ainsi que les témoins négatifs détectés uniquement par l'une des méthodes (jaune/bleu) en fonction du nombre de signaux générés. La figure (d) présente les résultats des deux approches en terme de sensibilité (%) et de proportion de fausses découvertes (FDP, %) pour les signaux à statut connu, en fonction du nombre de signaux générés.

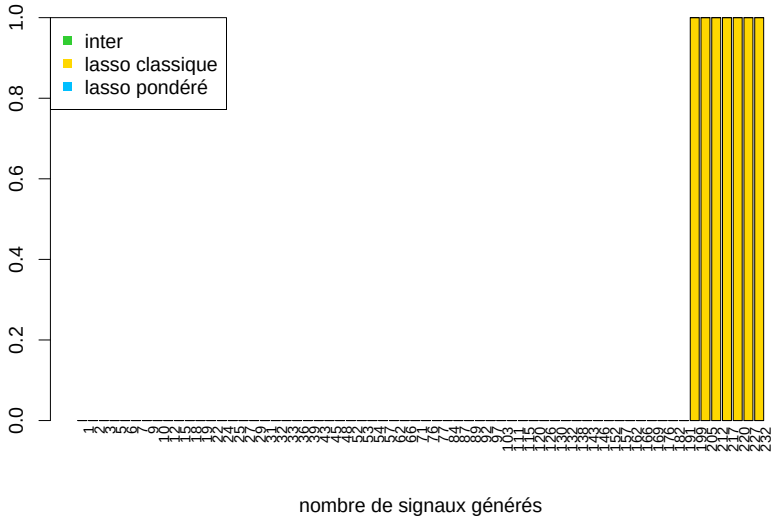


(a)

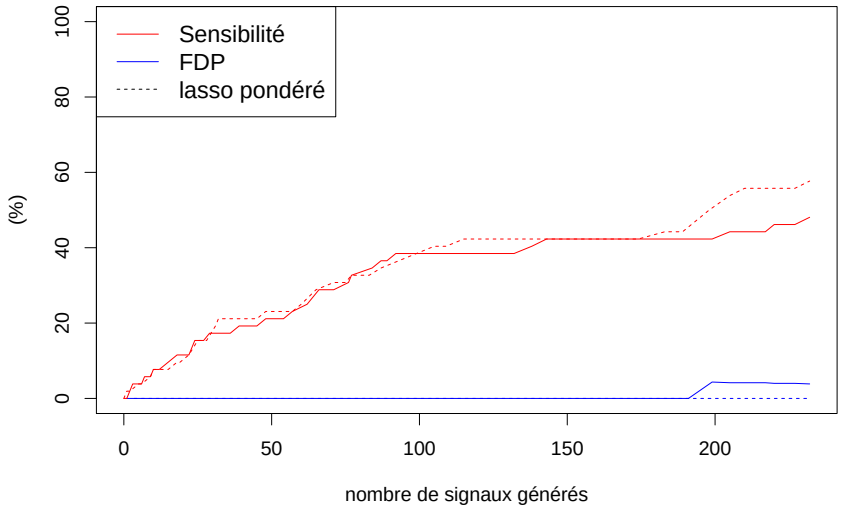


(b)

FIGURE 3 – Comparaison des performances entre le lasso et le lasso pondéré basé sur les **p-valeurs de la base hybride**. La figure (a) représente le nombre de signaux détectés par les deux méthodes (vert) ainsi que les signaux détectés uniquement par l'une des méthodes (jaune/bleu) en fonction du nombre de signaux générés. De manière analogue, la figure (b) présente le nombre de témoins positifs détectés par les deux méthodes.



(c)



(d)

FIGURE 4 – Comparaison des performances entre le lasso et le lasso pondéré basé sur les **p-valeurs de la base hybride**. La figure (c) représente le nombre de témoins négatifs détectés par les deux méthodes (vert) ainsi que les témoins négatifs détectés uniquement par l'une des méthodes (jaune/bleu) en fonction du nombre de signaux générés. La figure (d) présente les résultats des deux approches en terme de sensibilité (%) et de proportion de fausses découvertes (FDP, %) pour les signaux à statut connu, en fonction du nombre de signaux générés.

Titre : Score de propension en grande dimension et régression pénalisée pour la détection automatisée de signaux en pharmacovigilance

Mots clés : Détection de signaux en pharmacovigilance, Sélection de variables en grande dimension, Notifications spontanées, Régression pénalisée lasso, Score de propension en grande dimension, Bases médico-administratives

Résumé : La pharmacovigilance a pour but de détecter le plus précocement possible les effets indésirables des médicaments commercialisés. Elle repose sur l'exploitation de grandes bases de données de notifications spontanées, c'est-à-dire de cas rapportés par des professionnels de santé d'évènements indésirables soupçonnées d'être d'origine médicamenteuse. L'exploitation automatique de ces données pour l'identification de signaux statistiques repose classiquement sur des méthodes de disproportionnalité qui s'appuient sur la forme agrégée des données. Plus récemment, des méthodes basées sur des régressions multiples ont été proposées pour prendre en compte les poly-expositions médicamenteuses. Dans le chapitre 2, nous proposons une méthode basée sur le score de propension en grande dimension (HDPS). Une étude empirique, conduite sur la base de pharmacovigilance française et basée sur un ensemble de référence relatif aux atteintes hépatiques aigues (DILIrank), est réalisée pour comparer les performances de cette méthode (déclinée en 12 modalités) à des méthodes basées sur des régressions pénalisées lasso. Dans ce travail, l'influence de la méthode d'estimation des scores est minime, contrairement à la méthode d'intégration des scores. En particulier, la pondération sur l'HDPS avec des poids matching weights montre de bonnes performances, comparables à celles des méthodes basées sur le lasso. Dans le chapitre 3, nous proposons une méthode basée sur extension du lasso: le lasso adaptatif qui permet d'introduire des pénalités propres à chaque variable via des poids. Nous proposons deux nouveaux poids adaptés aux données de notifications, ainsi que l'utilisation du BIC pour le choix de la valeur de pénalité. Une vaste étude de simulations est réalisée pour comparer les performances de nos propositions à d'autres implémentations du lasso adaptatif, une méthode de disproportionnalité, des méthodes basées sur le lasso et sur l'HDPS. Les méthodes proposées montrent globalement de meilleurs résultats en termes de fausses découvertes et de sensibilité que les méthodes concurrentes. Une étude empirique analogue à celle du chapitre 2 vient compléter l'évaluation. Toutes les méthodes présentées sont implémentées dans le package R « adapt4pv » disponible sur le CRAN. En parallèle des développements méthodologiques sur les notifications spontanées, un intérêt croissant s'est porté autour de l'utilisation des bases médico-administratives pour la détection de signaux en pharmacovigilance. Les efforts de recherche méthodologique dans ce domaine en sont encore à leurs débuts. Dans le chapitre 4, nous explorons des stratégies de détection exploitant les notifications spontanées et l'Echantillon Généraliste des Bénéficiaires (EGB). Nous évaluons tout d'abord les performances d'une détection sur l'EGB à partir de DILIrank. Puis, nous considérons une détection conduite sur les notifications spontanées basée sur un lasso adaptatif intégrant, au travers de ses poids, l'information relative à l'exposition médicamenteuse d'individus contrôles mesurée dans l'EGB. Dans les deux cas, l'apport des données médico-administratives est difficile à évaluer du fait de la relative faible taille des données de l'EGB.

Title: High-dimensional propensity score and penalized regression for automated signal detection in pharmacovigilance

Keywords: Signal detection in pharmacovigilance, High-dimensional variable selection, Spontaneous reports, Lasso penalized regression, High-dimensional propensity score, Medico-administrative databases

Abstract: Post-marketing pharmacovigilance aims to detect as early as possible adverse effects of marketed drugs. It relies on large databases of individual case safety reports of adverse events suspected to be drug-induced. Several automated signal detection tools have been developed to mine these large amounts of data in order to highlight suspicious adverse event-drug combinations. Classical signal detection methods are based on disproportionality analyses of counts aggregating patients' reports. Recently, multiple regression-based methods have been proposed to account for multiple drug exposures. In chapter 2, we propose a signal detection method based on the high-dimensional propensity score (HDPS). An empirical study, conducted on the French pharmacovigilance database with a reference signal set pertaining to drug-induced liver injury (DILIrank), is carried out to compare the performance of this method (in 12 modalities) to methods based on lasso penalized regressions. In this work, the influence of the score estimation method is minimal, unlike the score integration method. In particular, HDPS weighting with matching weights shows good performances, comparable to those of lasso-based methods. In chapter 3, we propose a method based on a lasso extension: the adaptive lasso which allows to introduce specific penalties to each variable through adaptive weights. We propose two new weights adapted to spontaneous reports data, as well as the use of the BIC for the choice of the penalty term. An extensive simulation study is performed to compare the performances of our proposals with other implementations of the adaptive lasso, a disproportionality method, lasso-based methods and HDPS-based methods. The proposed methods show overall better results in terms of false discoveries and sensitivity than competing methods. An empirical study similar to the one conducted in chapter 2 completes the evaluation. All the evaluated methods are implemented in the R package "adapt4pv" available on the CRAN. Alongside to methodological developments in spontaneous reporting, there has been a growing interest in the use of medico-administrative databases for signal detection in pharmacovigilance. Methodological research efforts in this area are to be developed. In chapter 4, we explore detection strategies exploiting spontaneous reports and the national health insurance permanent sample (Echantillon Généraliste des bénéficiaires, EGB). We first evaluate the performance of a detection on the EGB using DILIrank. Then, we consider a detection conducted on spontaneous reports based on an adaptive lasso integrating, through weights, the information related to the drug exposure of a control group measured in the EGB. In both cases, the contribution of medico-administrative data is difficult to evaluate because of the relatively small size of the EGB.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France