



UNIVERSITE DE LILLE
FACULTE DE MEDECINE HENRI WAREMBOURG
Année : 2018

THESE POUR LE DIPLOME D'ETAT
DE DOCTEUR EN MEDECINE

**Proposition d'un outil intégrant l'environnement OHDSI
pour la pharmacoépidémiologie**

Présentée et soutenue publiquement le 17 septembre 2018 à 16h
au Pôle Formation
Par Nicolas DEPAS

JURY

Président :

Monsieur le Professeur Philippe Amouyel

Asseseurs :

Monsieur le Professeur Alain Duhamel

Monsieur le Professeur Emmanuel Chazard

Directeurs de Thèse :

Monsieur le Docteur Grégoire Ficheur

Monsieur le Docteur Alexandre Caron

Avertissement

La Faculté n'entend donner aucune approbation aux opinions émises dans les thèses : celles-ci sont propres à leurs auteurs.

Liste des abréviations

<i>ARS</i>	<i>Agence Régionale de Santé</i>
<i>CCAM</i>	<i>Classification Commune des Actes Médicaux</i>
<i>CCO</i>	<i>Case-crossover</i>
<i>CDM</i>	<i>Common Data Model</i>
<i>CIM-10</i>	<i>Classification Internationale des Maladies, 10^e révision</i>
<i>EIG</i>	<i>Effets indésirables graves</i>
<i>EIM</i>	<i>Effets indésirables médicamenteux</i>
<i>EU-ADR</i>	<i>Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge</i>
<i>FDA</i>	<i>Food and Drug Administration</i>
<i>GHM</i>	<i>Groupe homogène de malades</i>
<i>GHS</i>	<i>Groupe homogène de séjour</i>
<i>HAD</i>	<i>Hospitalisation à Domicile</i>
<i>MCO</i>	<i>Médecine Chirurgie Obstétrique</i>
<i>NIR</i>	<i>Numéro d'Inscription au Répertoire</i>
<i>OHDSI</i>	<i>Observational Health Data Sciences and Informatics</i>
<i>OMOP</i>	<i>Observational Medical Outcomes Partnership</i>
<i>OMS</i>	<i>Organisation Mondiale de la Santé</i>
<i>PAERPA</i>	<i>Parcours des personnes âgées en risque de perte d'autonomie</i>
<i>PMSI</i>	<i>Programme de médicalisation des systèmes d'information</i>
<i>PSY</i>	<i>Psychiatrie</i>
<i>SCC</i>	<i>Self-controlled cohort</i>
<i>SCCS</i>	<i>Self-controlled case series</i>
<i>SNDS</i>	<i>Système National des Données de Santé</i>
<i>SNIIRAM</i>	<i>Système National d'Information Inter-Régimes de l'Assurance Maladie</i>
<i>SNOMED</i>	<i>Systematized Nomenclature of Medicine</i>
<i>SSR</i>	<i>Soins de Suites et de Réadaptations</i>
<i>SynPUF</i>	<i>Synthetic Public Use File</i>

Table des matières

Résumé	21
Introduction générale	23
I. Les effets indésirables médicamenteux.....	23
A. Définitions	24
1. Les effets indésirables de type A.....	24
2. Les effets indésirables de type B.....	25
3. Les effets indésirables de type C	25
II. Détection des effets indésirables médicamenteux	26
A. Les essais cliniques randomisés	26
B. La pharmacovigilance	26
C. La pharmacoépidémiologie	27
III. Les bases médico-administratives.....	28
A. Le Système National des Données de Santé (SNDS)	28
1. Le Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM) :	28
2. Le Programme de Médicalisation des Systèmes d'Information (PMSI).....	29
3. Le Centre d'épidémiologie sur les causes médicales de décès (CépiDC)	30
4. Autres sources de données en cours d'ajout.....	30
IV. Projets de références en pharmacoépidémiologie	30
V. Observational Health Data Sciences and Informatics (OHDSI).....	32
A. Le projet Observational Medical Outcomes Partnership (OMOP).....	32
1. Le modèle de données commun (<i>Common Data Model</i> , CDM)	33
2. Corpus de référence	35
3. Evaluation des méthodes	36
a. Rationnel de la mesure de la performance des méthodes	36
b. Méthodes évaluées.....	36
c. Résultats et conclusions	37
B. Observational Health Data Sciences and Informatics.....	37
VI. Visualisation des données	38
VII. Objectifs	39
Matériel et Méthodes	41
I. Modèle de données commun OMOP.....	41
II. Les méthodes autocontrôlées.....	43
1. <i>Self-controlled cohort</i> (SCC)	43

2.	<i>Case-crossover (CCO)</i>	43
3.	<i>Self-controlled case series (SCCS)</i>	44
III.	L'application	44
A.	Paramètres de l'interface utilisateur	44
1.	Page 1 : Définition de la population d'intérêt	45
2.	Page 2 : Analyses descriptives	46
3.	Page 3 : Analyses autocontrôlées	46
B.	Spécifications techniques.....	48
IV.	Données test	49
1.	CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use File (SynPUF) 49	
2.	Les données du dispositif PAERPA	49
	Résultats	53
I.	Description des données mises en base	53
II.	Présentation de l'application.....	54
A.	Sélection de la population d'intérêt.....	55
B.	Analyses descriptives.....	55
C.	Analyses autocontrôlées	58
	Discussion	59
I.	La réutilisation des données	59
II.	Le choix du CDM OMOP	60
III.	L'utilisation des méthodes autocontrôlées	61
IV.	L'utilisation de l'outil.....	62
V.	Perspectives.....	63
VI.	Conclusion	64
	Références bibliographiques	65

Résumé

Introduction : En Europe, un rapport de la commission européenne de 2008 estimait que 3 à 10% des hospitalisations avait pour cause un effet indésirable médicamenteux (EIM). La pharmacoépidémiologie vient compléter les essais cliniques randomisés et les reports de cas en pharmacovigilance pour explorer le lien entre des couples médicament et EIM potentiel. Le projet *Observational Medical Outcomes Partnership* (OMOP) puis le consortium *Observational Health Data Sciences and Informatics* (OHDSI) ont étudié empiriquement la performance des designs d'étude en pharmacoépidémiologie. Sur les bases de données médico-administratives, ils ont montré l'intérêt des designs autocontrôlés. L'objectif de ce travail était de développer une application interactive permettant l'utilisation des méthodes autocontrôlées validées par OMOP/OHDSI en pharmacoépidémiologie.

Méthodes : L'application, développée sous R à partir du package RShiny, reposait sur le modèle de données commun OMOP pour son utilisation. Elle implémentait trois méthodes autocontrôlées : le *self-controlled case series*, le *self-controlled cohort* et le *case-crossover*. Deux bases de données ont servi au développement et au test de l'application : SynPUF, une base de données simulée à partir des données de Medicare, et les données de l'assurance maladie pour les patients de plus de 75 ans habitant dans le Valenciennois-Quercitain rentrant dans le cadre du projet PAERPA.

Résultats : Les données SynPUF et les données de l'assurance maladie ont été mises au format du modèle de données commun OMOP. A partir de ce format de données, l'application propose une analyse pharmacoépidémiologique en 3 étapes : (1) définition de la population d'intérêt, (2) description de la population sélectionnée et des diagnostics, traitement, actes enregistrés pour celle-ci et (3) réalisation d'analyses à la volée basées sur les méthodes autocontrôlées.

Discussion : Cette application permet de réaliser de manière exploratoire des analyses de pharmacoépidémiologie. L'environnement OMOP/OHDSI ne se limite pas aux méthodes autocontrôlées et est en constant développement, laissant la porte ouverte à des mises à jour de l'application pour d'autres méthodes. Au moment où l'accès au Système National des Données de Santé se démocratise, la transcription de ces méthodes dans le contexte français pourrait s'avérer intéressante.

Introduction générale

I. Les effets indésirables médicamenteux

De par leurs fréquences et leurs conséquences, les effets indésirables médicamenteux (EIM) constituent un fardeau pour la société en termes de morbidité et de mortalité. En 1994, aux Etats-Unis, plus de 100 000 décès étaient attribués à des EIM, et ils pouvaient représenter la 4^e cause de décès (1). Chez les patients hospitalisés, l'incidence des effets indésirables médicamenteux graves (EIG) était estimée à 6,7% dont 0,32% fatals (1). En 2000 en Finlande, les EIM étaient impliqués dans 5% des décès de patients hospitalisés (2). Cette étude, mise à jour en 2012, retrouvait encore 3% des décès de patients hospitalisés imputables aux médicaments (3). En Europe, un rapport de la commission européenne de 2008 estimait que 3 à 10% des hospitalisations avait pour cause un EIM et que 0,12% à 0,22% des hospitalisations se compliquaient d'un EIM fatal. Ce rapport estimait le coût des EIM à 79 milliard d'euros (4). En France, l'Enquête Nationale sur les Evénements Indésirables graves associés aux Soins (ENEIS) a recensé en moyenne 6,2 EIG pour 1 000 jours d'hospitalisation dont la moitié serait évitable. En 2009, dans les services de médecine et de chirurgie, 4,5% des séjours ont été causés par un EIG soit 1 séjour sur 20 et, c'est entre 330 000 et 490 000 séjours qui seraient évitables chaque année (5).

A. Définitions

En 1972, l'Organisation mondiale de la Santé (OMS) définit l'effet indésirable médicamenteux comme une « réaction nocive et non voulue, se produisant aux posologies normalement utilisées chez l'homme pour la prophylaxie, le diagnostic ou le traitement d'une maladie ou la modification d'une fonction physiologique ». En 2000, cette définition est complétée par la notion de « mésusage du produit de santé, constituant un syndrome de sevrage lors de l'arrêt du produit ou d'un syndrome de dépendance ainsi que toute réaction découlant d'un usage abusif. Il comprend également toute réaction nuisible pouvant découler d'une mauvaise qualité du produit de santé. » (6). De cette définition découle la notion d'effet indésirable grave quand l'effet indésirable est à l'origine d'une des situations suivantes :

- Mise en jeu du pronostic vital ou décès ;
- Hospitalisation ou prolongation d'hospitalisation ;
- Incapacité persistante ou invalidité ;
- Mésusage ou réaction de dépendance.

Les EIM sont nombreux, hétérogènes dans leurs caractéristiques et dans la manière de les détecter. Cependant, trois grandes catégories sont reprises par l'OMS (6,7).

1. Les effets indésirables de type A

Il s'agit d'effets indésirables liés à un effet pharmacologique (exagéré). Ils sont fréquents, dose-dépendants et peuvent être évités par l'utilisation de doses appropriées pour chaque patient (exemple : toux et inhibiteurs de l'enzyme de conversion (8)). Ces effets peuvent être reproduits et étudiés expérimentalement et ont souvent déjà été identifiés avant la commercialisation du médicament.

2. Les effets indésirables de type B

Il s'agit d'effets indésirables qui ne surviennent que chez une minorité de patients et indépendamment de la dose du médicament administrée. Ils sont en général rares, imprévisibles, graves et très difficiles à étudier (exemple : hépatite sous isoniazide).

3. Les effets indésirables de type C

Il s'agit d'effets indésirables qui réfèrent aux situations où l'utilisation d'un médicament, pour une raison indéterminée, va augmenter la fréquence d'apparition d'une maladie 'spontanée'. Les effets type C peuvent être bénins ou graves (incluant les tumeurs malignes) et peuvent avoir des répercussions en santé publique. Les effets type C résultent généralement d'une prise chronique du médicament. Souvent la chronologie d'apparition de l'événement par rapport à la prise médicamenteuse n'est pas suggestive et les liens de causalité sont difficiles à établir (exemple : complications thromboemboliques et contraception orale).

II. Détection des effets indésirables médicamenteux

A. Les essais cliniques randomisés

Lors d'un essai clinique randomisé, en plus de l'efficacité du médicament, sa tolérance est systématiquement évaluée. La déclaration des EIM est d'ailleurs encadrée en France et en Europe depuis le début des années 2000 (9,10). Cependant, leur validité externe reste limitée (11,12) et il existe un « gouffre, largement sous-estimé entre les mesures scientifiques basées sur les essais cliniques randomisés et le bénéfice mesuré en population » (Cochrane, 1972) (13). Enfin, la fréquence d'un EIM est souvent faible et un essai clinique n'a pas la puissance nécessaire pour les mettre en évidence. Une grande partie des EIM n'est donc pas repérée lors de la mise sur le marché.

B. La pharmacovigilance

Après la mise sur le marché, la détection des EIM s'appuie sur les systèmes de pharmacovigilance. Son objectif est la surveillance des médicaments et la prévention du risque d'EIM résultant de leur utilisation, que ce risque soit potentiel ou avéré (14). La pharmacovigilance se base sur la notification spontanée des EIM par les professionnels de santé, les patients ou les industriels et est de ce fait sujette à de nombreux biais, en particulier de sélection (seulement 5% des EIG déclarés en France en 1998)(15). Des méthodes de « disproportionnalités » ont été développées spécifiquement pour les bases de pharmacovigilance et permettent de mettre en évidence des associations médicament/événement surreprésentées dans ces bases.

C. La pharmacoépidémiologie

La pharmacoépidémiologie est une « discipline mettant en application les méthodes et/ou raisonnements épidémiologiques pour évaluer, généralement sur de grandes populations, l'efficacité, le risque et l'usage des médicaments » (16). Son objectif n'est pas de remplacer la pharmacovigilance, mais bien d'être complémentaire, la pharmacovigilance permettant de détecter et caractériser cliniquement un risque potentiel et la pharmacoépidémiologie apportant un meilleur niveau de preuve quant à la conclusion d'un risque avéré (17). Par certains aspects, elle élargit les champs d'application de la pharmacovigilance, en s'intéressant aux conditions d'usage (ou de mésusage) des médicaments, à l'analyse des facteurs médicaux, économique ou rationnels de leur utilisation, la vérification à large échelle de leur efficacité à long terme, la surveillance de l'automédication : la vie réelle du médicament après sa mise sur le marché (18).

Cette discipline n'est pas nouvelle. Déjà en 1974, le *Boston Collaborative Drug Surveillance Group* mettait en évidence l'effet cardio-protecteur potentiel de l'aspirine avant tout essai clinique (19,20). Plus récemment, la pharmacoépidémiologie a permis de montrer l'excès de prescription des médicaments aux propriétés atropiniques chez la personne âgée (21). En France, elle est maintenant intégrée dans le processus de décision, notamment après à son apport dans la mise en évidence de la relation entre le Médiator® et le risque de valvulopathie (20,22,23). Son utilisation dans le cadre de la réutilisation des données se voit favorisée par la récente création du Système National des Données de Santé et son ouverture pour la recherche (24).

III. Les bases médico-administratives

La pharmacoépidémiologie peut notamment s'appuyer sur des bases médico-administratives. Elles permettent d'avoir accès aux informations nécessaires à l'analyse pharmacoépidémiologique : les prescriptions médicamenteuses, les diagnostics médicaux, un suivi longitudinal pour chaque patient, une exhaustivité de la population cible, une attrition marginale,... (25). Non développées dans un but de recherche, elles proposent un défi en termes de sélection de données pertinentes et d'analyse. Elles sont cependant sujettes à d'autres limites comme le biais de recueil, le manque de précision sur les données cliniques et biologiques, ou encore l'évolution des classifications compliquant les analyses longitudinales sur de longues périodes (26).

A. Le Système National des Données de Santé (SNDS)

En France, le SNDS est accessible à la recherche depuis avril 2017 (27). Il regroupe les principales bases de données médico-administratives françaises qui sont chaînées à travers un identifiant unique (Numéro d'Inscription au Répertoire (NIR)) et permettent de suivre la consommation de soins d'un patient et d'analyser son parcours dans le système de santé (27,28). Le SNDS comprend :

1. Le Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM) :

Mis en place en 2003, il couvre 98,8% de la population française et contient des données :

- Démographiques ;
- Sur les pathologiques chroniques à travers le renseignement des affections longues durées (ALD) ;

- Sur les consommations de soins en ville comme les consultations médicales ou paramédicales et les médicaments délivrés en pharmacie.

Le SNIIRAM couvre la quasi-totalité de la population française de la naissance au décès quelle que soit son âge et sa situation socioéconomique. En 2016, 600 millions de patients-années y étaient stockés (29).

2. Le Programme de Médicalisation des Systèmes d'Information (PMSI)

En 1991, la loi portant sur la réforme hospitalière instaure que « les établissements de santé, publics et privés, doivent procéder à l'évaluation et à l'analyse de leur activité » (30). Le PMSI, déjà en cours de développement et d'expérimentation dans certains hôpitaux publics et privés s'impose comme le standard de recueil de l'information médicale. Initialement prévu pour la description simple de l'activité hospitalière, il devient en 2004 un outil d'allocation de ressource dans le cadre de la tarification à l'activité (T2A).

Les informations médicales principalement récupérées pour chaque hospitalisation sont les diagnostics codés en CIM-10 (Classification Internationale des Maladies, 10e révision) et les actes codés en CCAM (Classification Commune des Actes Médicaux).

SNIIRAM et PMSI ont déjà montré leurs utilité en pharmacoépidémiologie, notamment en mettant en évidence l'association entre le cancer de la vessie et la pioglitazone (31).

3. Le Centre d'épidémiologie sur les causes médicales de décès (CépiDC)

Il s'agit d'un laboratoire de l'Institut National de la Santé et de la Recherche Médicale (INSERM) ayant pour mission le recueil puis l'analyse statistique descriptive des causes médicales de décès en France. Ces données proviennent de la documentation en routine des certificats de décès (32).

4. Autres sources de données en cours d'ajout

- La Caisse Nationale de Solidarité pour l'Autonomie (CNSA) permettant de récupérer les données relatives au handicap, disponible en 2018 ;
- Un échantillon de données en provenance des organismes d'Assurance Maladie complémentaire disponible à partir de 2019.

IV. Projets de références en pharmacoépidémiologie

A partir de la fin des années 2000, plusieurs consortiums se créent dans le but de développer la surveillance des médicaments à partir des bases médico-administratives.

En 2007 aux Etats-Unis, le Congrès mandate la *Food and Drug Administration* (FDA) afin d'avoir une identification et une analyse active des risques après la mise sur le marché des médicaments (*Active Postmarket Risk Identification and Analysis, ARIA*). En Mai 2008, la FDA lance l'initiative *Sentinel* dans le but de créer un système d'information national pour la surveillance de la sécurité des médicaments avec comme projets principaux (33) :

- *Mini Sentinel* qui a pour but de développer les méthodes nécessaires au projet *Sentinel* par :

- Un laboratoire pour développer et évaluer les méthodologies scientifiques qui pourraient être utilisées dans le projet *Sentinel* ;
 - L'évaluation des questions de sécurité sur les bases de données médicales déjà existantes et la connaissance des challenges auxquels ils seront confrontés.
- Le projet *Observational Medical Outcomes Partnership* (OMOP) : partenariat public-privé qui a pour but d'informer sur l'utilisation appropriée des bases de données de santé (bases médico-administrative et dossiers patients informatisés) pour l'étude des effets des produits de santé (risques et bénéfiques). Il est poursuivi par la création du consortium *Observational Health Data Sciences and Informatics* (OHDSI). Les projets OMOP et OHDSI sont détaillés dans le chapitre suivant.

A l'international, d'autres projets se forment comme le réseau CNODES (*Canadian Network for Observational Drug Effect Studies*) (34) au Canada ainsi que le réseau AsPEN (*Asian Pharmacoepidemiology Network*) en Asie (35).

En Europe, trois projets sont mis en place :

- Le projet EU-ADR (*Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge*) a été lancé en 2008 par la Commission Européenne (36);
- Le projet PROTECT (*Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsortIum*), partenariat public privé coordonné par l'agence européenne des médicaments entre 2009 et 2015 (37,38) ;
- Le projet PSIP (*Patient Safety through Intelligent Procedures in medication*), porté par le Centre Hospitalo-Universitaire de Lille entre 2008 et 2011 (39).

V. Observational Health Data Sciences and Informatics (OHDSI)

Débuté en 2013, OHDSI reprend les résultats du projet OMOP et développe le caractère collaboratif et ouvert dans l'utilisation des méthodes développées. Afin d'expliquer les activités d'OHDSI, nous détaillerons d'abord le projet OMOP avant de développer l'apport d'OHDSI depuis 2013.

A. Le projet Observational Medical Outcomes Partnership (OMOP)

Le projet OMOP a été créé en 2008 pour une durée de 5 ans avec pour objectif d'évaluer de manière empirique la performance de différentes méthodes en pharmacoépidémiologie.

Ce projet impliquait de nombreux acteurs et l'évaluation des méthodes a été réalisée dans des structures telles que *Columbia University, Uppsala Monitoring Centre, University of Utah...* Chaque méthode était testée sur des bases de données observationnelles privées qui pouvaient être des bases médico-administratives (données *Medicare, Medicaid* et commerciales provenant de *Thomson Reuters*) ou des dossiers patients informatisés (*GE Healthcare*) regroupant les informations pour 86 millions de personnes. En parallèle, des données simulées étaient aussi utilisées pour évaluer la performance des méthodes sur une base où les risques relatifs des effets indésirables étaient connus. Dans ce contexte, l'utilisation d'un modèle de données commun standardisant des données hétérogènes devenait indispensable afin de satisfaire aux exigences de reproductibilité au sein du consortium (40).

L'objectif de ce programme de recherche était de clarifier les limites des méthodes et des données utilisées nécessaires pour une surveillance active de la sécurité des médicaments incluant (40)

- La validation des données : peut-on confirmer que les données natives sont bien transformées dans le modèle de données commun ?
- La faisabilité : est-ce que toutes les méthodes ont pu être exécutées sur toutes les bases de données ?
- La performance : quelles ont été les performances des méthodes et est-ce que les résultats obtenus étaient cohérents avec ceux attendus ?

1. Le modèle de données commun (*Common Data Model, CDM*)

Le CDM a été conçu de manière à récupérer toutes les informations nécessaires pour une surveillance active des médicaments et incluant les deux principaux types de bases de données en santé : bases médico-administratives et dossiers patients informatisés. Le CDM OMOP propose une structuration unique des données afin de pouvoir développer et comparer des méthodes sur les différentes sources (40).

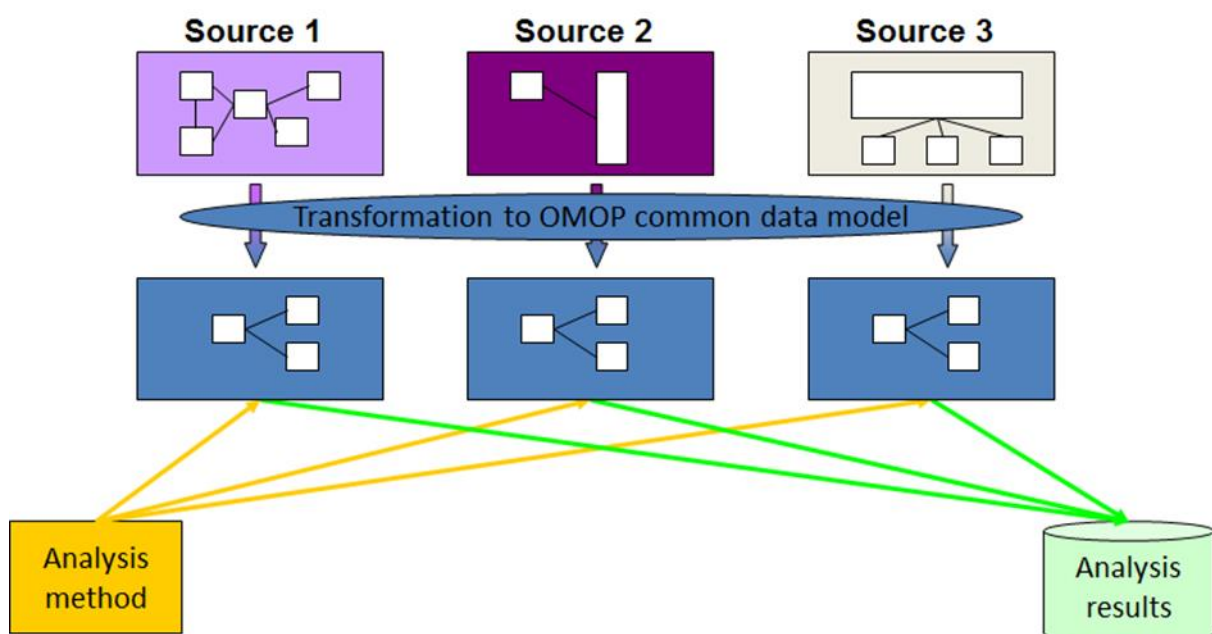


Figure 1 : Modèle de données commun (41)

Il a été publié en 2009 et testé sur une cohorte de 43 millions de patients provenant de 2 bases différentes (41) :

- Une base médico-administrative : *PHARMetrics choices, IMS Health* ;
- Une base de dossiers patients informatisés : *GE Commercial Data Set*.

Afin d'améliorer l'homogénéité entre les bases, OMOP a défini des classifications standards pour son CDM. Pour les diagnostics il s'agissait de la SNOMED (*Systematized Nomenclature of Medicine*) et pour les médicaments de RxNorm. OHDSI a par la suite fourni des outils pour réaliser la correspondances entre les classifications sources et les classifications standardisées par l'intermédiaire de leur site Athena (42).

Un travail de validation a été effectué sur 10 autres bases de données afin de quantifier la perte d'information associée à la transformation dans le CDM (terminologies différentes, précisions de l'information différente...) (43) : plus de 90% des diagnostics et des médicaments étaient correctement retrouvés dans les bases au format du CDM OMOP. Les limites soulevées lors de ce travail étaient l'évolution des terminologies dans le temps qui nécessitaient un investissement important dans le suivi des versions, une perte de granularité pour certains diagnostics (la terminologie d'origine était plus précise que la terminologie du CDM OMOP) et un temps nécessaire non négligeable pour la transformation des données avec l'implication et la coordination de différentes personnes et compétences.

2. Corpus de référence

La performance de chacune des méthodes était évaluée à partir d'un corpus de référence contenant des contrôles positifs (associations entre un traitement et un effet indésirable démontré) et des contrôles négatifs (associations entre un traitement et un effet indésirable non démontré) (44). Quatre effets indésirables considérés comme des priorités pour la pharmacovigilance ont été retenus : l'infarctus du myocarde, l'insuffisance rénale aiguë, l'insuffisance hépatique aiguë et les saignements gastro-intestinaux haut (45). La performance de chacune des méthodes a été testée sur les 165 contrôles positifs et les 234 contrôles négatifs identifiés lors d'une revue systématique à partir des sources suivantes (Figure 2) :

- Les effets indésirables mentionnés dans la notice des médicaments au 22 janvier 2011 (46) ;
- La revue systématique de Tisdale (47) ;
- Une revue de la littérature sur Pubmed recherchant les essais cliniques et les études observationnelles entre 1990 et 2011.

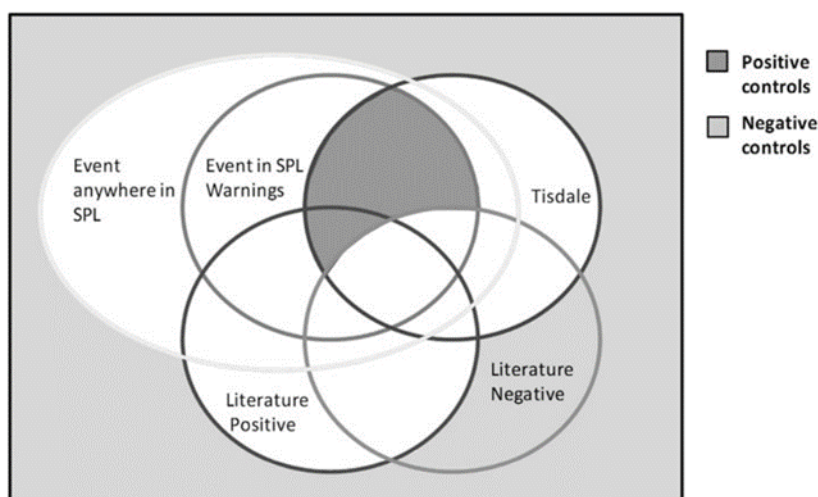


Figure 2 : Critères d'inclusions et d'exclusions pour les contrôles positifs et négatifs (44)

Les contrôles positifs retenus sont à l'intersection de la revue Tisdale et des effets mentionnés dans les notices à l'exclusion des médicaments ayant une ou plusieurs études négatives sur le sujet. Les contrôles négatifs ne doivent pas être présent dans la revue Tisdale, les notices des médicaments et les études positives.

3. Evaluation des méthodes

a. Rationnel de la mesure de la performance des méthodes

Lors de la construction d'une étude, les choix de designs reposent sur des décisions subjectives d'experts influencés par leurs connaissances, leurs habitudes d'analyse et leurs expériences passées (48). L'étude EDDIE (*Epidemiology Design Decision Inventory and Evaluation*) a interrogé plusieurs experts afin de connaître les designs qu'ils emploieraient pour réaliser l'étude de l'association entre un traitement et un effet indésirable. Alors que le design de cohorte était le plus souvent cité, les paramètres choisis pour sa réalisation étaient très hétérogènes (49). Le choix d'un design doit donc non seulement se baser sur les avis d'experts mais doit également fournir des preuves empiriques de sa performance dans le contexte de l'analyse proposée (50).

b. Méthodes évaluées

Le processus pour l'analyse de la performance des méthodes était commun afin d'assurer la comparabilité des résultats obtenus :

1. Définition des paramètres modifiables comme les périodes à risque après la prise d'un médicament, la prise en compte d'un seul ou de plusieurs effets indésirables, la prise en compte des effets indésirables le jour de l'introduction d'un médicament, le nombre de témoins par cas pour les méthodes cas-témoins...
2. Evaluation de chaque combinaison de paramètres possibles sur les 399 paires médicament / effets indésirable.
3. Calcul d'une aire sous la courbe roc, stratifiée sur les quatre effets indésirables.

Les différentes méthodes évaluées par OMOP étaient les suivantes :

- Méthodes de disproportionnalité issues de la pharmacovigilance (51) ;
- Méthodes épidémiologiques classiques : cas-témoins et cohortes (52,53) ;
- Méthodes autocontrôlées: *self-controlled case series*, *self-controlled cohort* (54,55). Il s'agit de méthodes où le patient est son propre témoin et qui de fait permettent de prendre en compte les facteurs de confusion indépendants du temps.

c. Résultats et conclusions

Les méthodes de disproportionnalité, développées pour des bases de pharmacovigilance, étaient les moins performantes sur les bases de données médico-administratives. La cohorte était plus performante que le cas-témoins et les méthodes autocontrôlées étaient les plus performantes (50). Des résultats similaires étaient retrouvés dans le projet européen EU-ADR lors de la réplique des analyses d'OMOP (56).

B. Observational Health Data Sciences and Informatics

OHDSI est un consortium créé en 2013 par plusieurs chercheurs ayant appartenus au consortium OMOP (57,58). Ils encouragent un changement de paradigme dans la manière de produire et diffuser les résultats de recherche : passer de résultats produits par des études uniques avec des méthodologies non reproductibles et non comparables à la production d'études à grande échelle, utilisant des méthodes standardisées sur des données enregistrées dans un même format permettant l'évaluation et la calibration des résultats (59). Les principaux objectifs du consortium, actuellement toujours actif, sont de :

- Continuer à développer des méthodes de pharmacoépidémiologie reproductibles. Les méthodes développées par OMOP sont mises à jour régulièrement et de nouvelles méthodes sont développées, utilisant notamment des algorithmes de *machine learning* ;
- Collaborer et former une communauté active : différentes branches d'OHDSI se sont développées en Europe, en Corée et en Chine et des publications collaboratives ont déjà vues le jour (60) ;
- Rendre accessible leurs méthodes à travers une politique open source, l'ensemble de leur travail étant disponible sur GitHub (61).

VI. Visualisation des données

Depuis le début des années 2000, le nombre, la taille et la complexité des bases de données, médico-administratives entre autres, n'ont cessé d'augmenter (62). Les données sont recueillies pour des besoins variés et l'augmentation de leurs taille entraîne une augmentation des associations potentielles à étudier (63). Il devient donc important de les résumer et de les simplifier de manière à pouvoir répondre aux questions de recherche posées (64,65). La visualisation des données permet de représenter les connaissances de manière intuitive et efficace à partir d'outils graphiques ainsi que donner l'accès à des méthodes statistiques avancées. (66,67).

Des travaux sont déjà menés dans ce sens, notamment en épidémiologie infectieuse (62), et récemment les formes d'applications web sont très utilisées (68). Le développement de ces outils permet l'utilisation de méthodes plus complexes : IT-CARES, développé par l'Université de Lille, permet par exemple de réaliser des analyses en case-crossover sur la base nationale PMSI (69,70).

VII. Objectifs

Les objectifs de ce travail étaient de développer une application permettant à un pharmacoépidémiologiste d'explorer la relation entre un médicament et un effet indésirable potentiel en appliquant l'un des designs autocontrôlés proposé par le consortium OHDSI.

Matériel et Méthodes

L'application développée repose sur l'utilisation de données médico-administratives stockées sous le format du modèle de données commun OMOP afin d'utiliser des méthodes autocontrôlées pour une étude de pharmacoépidémiologie. Nous détaillerons tout d'abord le modèle de données et les méthodes implémentées dans l'outil puis les données utilisées pour le développement et le test de l'application.

I. Modèle de données commun OMOP

Le CDM OMOP (figure 3) est le format d'entrée des données pour l'outil développé. Il est composé de 39 tables réparties en 6 catégories (71).

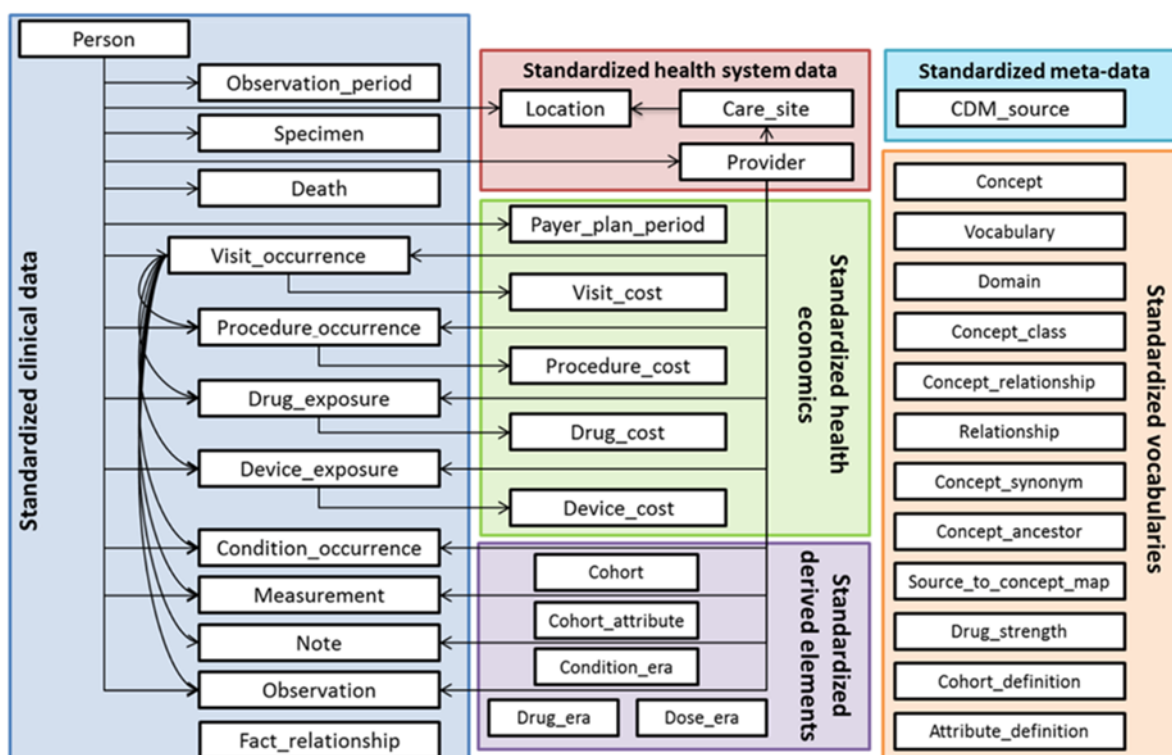


Figure 3 : Le modèle de données commun OMOP

Il s'agit d'un CDM centré sur le patient (*patient-centric*) : chaque information médicale des tables de la catégorie *Standardized clinical data* est reliée au patient. Tout contact avec le système de santé est enregistré dans la table *Visit_occurrence* qu'il s'agisse d'une hospitalisation, d'une consultation médicale ou paramédicale, d'une délivrance de médicaments en pharmacie, etc. Toutes les données médicales sont ensuite ajoutées dans les tables correspondantes (ex : actes dans *Procedure_occurrence*, diagnostics dans *Condition_occurrence*).

Chaque information est stockée sous la forme d'un code. L'ensemble de ces codes et leurs significations sont présents dans la table *Concept*. Ces concepts sont reliés entre eux grâce aux tables *Concept_ancestor* et *Concept_relationship*. Ces relations permettent d'identifier les correspondances entre nomenclatures, les composants d'un médicament ou encore les hiérarchies entre diagnostics. Tous les concepts et les relations sont mis à disposition par OHDSI sur leur site Athena (42).

Dans la catégorie *Standardized derivated elements*, les tables sont construites à posteriori du chargement des données. Les tables *Drug_era* et *Condition_era* permettent de créer des périodes de temps durant lesquelles un patient est atteint d'une pathologie ou prend un médicament, à partir d'événements ponctuels enregistrés dans la base de données. Pour un médicament délivré tous les mois, les données enregistrées correspondent aux dates de délivrance en pharmacie. La table *Drug_era* permet de construire la période durant laquelle le patient est exposé à ce médicament (Figure 4).

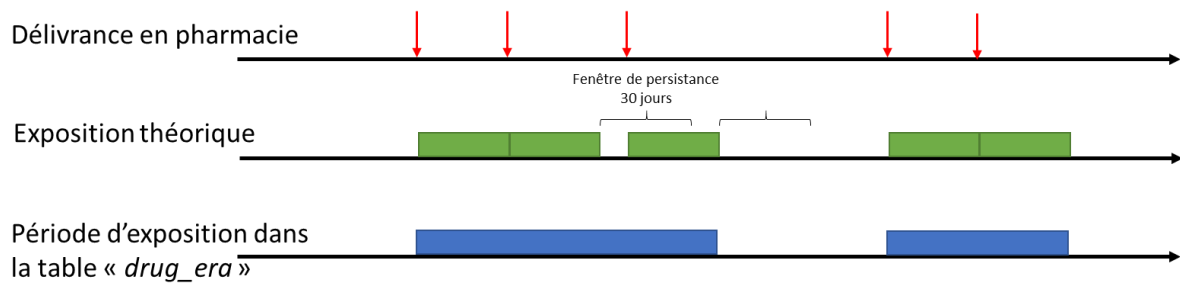


Figure 4 : Construction des périodes d'expositions dans la table « drug_era »

Une fenêtre de persistance est définie comme une période durant laquelle si une nouvelle délivrance est effectuée, le patient sera toujours considéré comme exposé. L'intervalle entre la fin d'une exposition théorique et une nouvelle délivrance peut être expliqué, par exemple, par un défaut d'observance de la part du patient conduisant à une délivrance tardive du médicament.

II. Les méthodes autocontrôlées

Le principe des méthodes autocontrôlées est d'utiliser le patient comme son propre témoin. Il n'est donc plus comparé à un autre patient, apparié selon divers critères (âge, sexe, etc.). Ce procédé présente l'avantage de contrôler l'ensemble des facteurs de confusion constants dans le temps (sexe, niveau socio-économique, génétique, etc.). Trois méthodes autocontrôlées sont proposées dans l'application.

1. *Self-controlled cohort (SCC)*

Seuls les patients avec au moins une exposition (médicamenteuse) sont inclus. Chez ces patients, on compare l'incidence d'un effet indésirable durant la période à risque post-exposition à l'incidence d'un effet indésirable durant la période préexposition.

2. *Case-crossover (CCO)*

Seuls sont inclus les cas, c'est-à-dire les patients avec au moins un effet indésirable médicamenteux. Chez ces patients on compare l'incidence de l'exposition dans la période cas précédant immédiatement l'effet indésirable à l'incidence de l'exposition dans une période contrôle située à distance de la période cas et que l'on considère comme indépendante de la survenue de l'effet indésirable.

3. *Self-controlled case series (SCCS)*

A l'intersection des méthodes précédentes, le *self-controlled case series* n'inclut que les patients avec au moins un effet indésirable et une exposition. Chez ces patients, on compare la probabilité d'effets indésirables durant les périodes exposées au traitement à la probabilité d'effets indésirables durant les périodes non exposées au traitement. Contrairement aux méthodes précédentes, le *self-controlled case series* permet de prendre en compte la durée d'exposition.

III. L'application

En format d'entrée, l'application s'appuie sur les données au format du CDM OMOP décrit dans le premier chapitre des méthodes.

A. Paramètres de l'interface utilisateur

Les trois étapes de réalisation d'une analyse de pharmacoépidémiologie sont réparties sur les trois pages constituant l'application.

1. Page 1 : Définition de la population d'intérêt

Les paramètres utilisateurs à définir sont détaillés dans le tableau 1.

Tableau 1 : Paramètres utilisateurs lors de la sélection de la population d'intérêt

Paramètre	Valeur	Utilisation
<i>Parameters to select cohort</i>	<i>Age, Sex or Chronic conditions</i>	Définit le ou les paramètres à utiliser pour la sélection de la population d'intérêt
<i>Age</i>	Age minimum et maximum	Inclut les patients en fonction de leurs âge au premier enregistrement dans la base de données
<i>Sex</i>	<i>Male / Female</i>	Inclut les patients en fonction du sexe
<i>Chronic conditions</i>	Liste de codes diagnostics	Inclut les patients avec au moins un des diagnostics chroniques sélectionnés au cours de leurs suivi

En sortie, un tableau récapitule les paramètres sélectionnés ainsi que le nombre de patients effectivement inclus. Le détail de ces patients (année de naissance, âge à l'inclusion, sexe, temps de présence enregistré dans la base en jours, décès) peut être affiché en sélectionnant le tableau.

2. Page 2 : Analyses descriptives

Les analyses descriptives sont divisées en 2 parties. La première partie restitue les données démographiques (âge et sexe), les données de mortalité ainsi que la durée de présence dans la base de données des patients sélectionnés.

La 2^e partie permet une exploration plus fine des diagnostics, actes et médicaments et propose les paramètres utilisateurs détaillés dans le tableau 2.

Tableau 2 : Paramètres utilisateurs de l'analyse descriptive

Paramètre	Valeur	Utilisation
<i>Select conditions table</i>	<i>Condition Era or Condition occurrence</i>	Sélectionne la table sur laquelle sera effectuée la recherche des diagnostics
<i>Select drugs table</i>	<i>Drug Era or Drug Exposure</i>	Sélectionne la table sur laquelle sera effectuée la recherche des médicaments
<i>Additional filters</i>	<i>Condition, Procedures or Drugs</i>	Définit des critères de sélections supplémentaires sur les diagnostics, actes et médicaments.
<i>Select condition</i>	Liste de diagnostics	Inclut les patients avec au moins un des diagnostics sélectionnés
<i>Select procedure</i>	Liste d'actes	Inclut les patients avec au moins un des actes sélectionnés
<i>Select drug</i>	Liste de médicaments	Inclut les patients avec au moins un des médicaments sélectionnés
<i>Show top associated</i>	<i>TRUE/FALSE</i>	Affiche les principaux diagnostics, actes ou médicaments associés
<i>How many associated elements to display</i>	Nombre	Définit le nombre de diagnostics associés à afficher si la case <i>Show top associated</i> est cochée

3. Page 3 : Analyses autocontrôlées

La page « Analyses » permet de réaliser plusieurs analyses autocontrôlées les unes à la suite des autres pour un ou plusieurs couples médicament / effet indésirable. Les paramètres utilisateurs sont détaillés dans le tableau 3. Les résultats seront résumés dans un même tableau ou graphique.

Tableau 3 : Paramètres utilisateurs des analyses autocontrôlées

Paramètre	Valeur	Utilisation
<i>Outcome of interest</i>	Liste de diagnostics	Sélectionne l'événement
<i>Drugs of interest</i>	Liste de médicaments	Sélectionne l'exposition
<i>Select analyses</i>	<i>Self-controlled case series,</i> <i>Self-controlled cohort or</i> <i>Case crossover</i>	Sélectionne parmi les trois méthodes autocontrôlées
<i>Select design</i>	Liste de designs	Sélectionne un des designs préconfiguré pour l'analyse autocontrôlée
<i>Exposure start</i>	Nombre	Période durant laquelle le médicament peut être suspecté d'être responsable d'un effet indésirable (Figure 5).
<i>Add exposed days to start</i>	TRUE/FALSE	Les périodes classiquement utilisées par OMOP et OHDSI sont :
<i>Exposure end</i>	Nombre	- La durée de l'exposition
<i>Add exposed days to end</i>	TRUE/FALSE	- 30 jours à partir du début de l'exposition - La durée de l'exposition + 30 jours - l'ensemble du suivi post exposition
<i>Naïve period</i>	Nombre	Uniquement dans le SCCS, elle est située au début de la période d'observation de la cohorte et permet de déterminer le statut du patient (antécédents, exposition actuelle...). Elle ne sera pas prise en compte pour l'analyse
<i>Washout period</i>	Nombre	Uniquement dans le SCC et CCO, elle se situe entre les périodes cas et témoins ou exposées et non exposées. Elle est là pour assurer un retour du risque relatif à un niveau basal.
<i>First outcome only</i>	TRUE/FALSE	Ne retenir que le premier événement pour réaliser l'analyse
<i>First exposure only</i>	TRUE/FALSE	Sauf dans le SCCS, n'utiliser que la première exposition pour l'analyse

Les designs préconfigurés dans l'application sont des designs utilisés par OHDSI et OMOP pour valider la performance des méthodes.

- Pour le *case-crossover*, il peut être réalisé sur l'ensemble de la population d'intérêt ou niché dans une cohorte ;
- Pour le *self-controlled cohort*, il s'agira de définir si la date index du début de l'exposition doit être prise en compte ou non lors de l'analyse et si le patient doit être exposé durant l'entièreté de la période exposée pour être inclus ;
- Pour le *self-controlled case series*, il peut être réalisé avec différents ajustements par l'âge, la saisonnalité ou encore les autres médicaments

présents chez ce patient. Il est aussi possible de définir une période de préexposition durant laquelle la probabilité de faire un événement est plus faible et qui si elle n'est pas définie tend à faire baisser l'estimation globale.

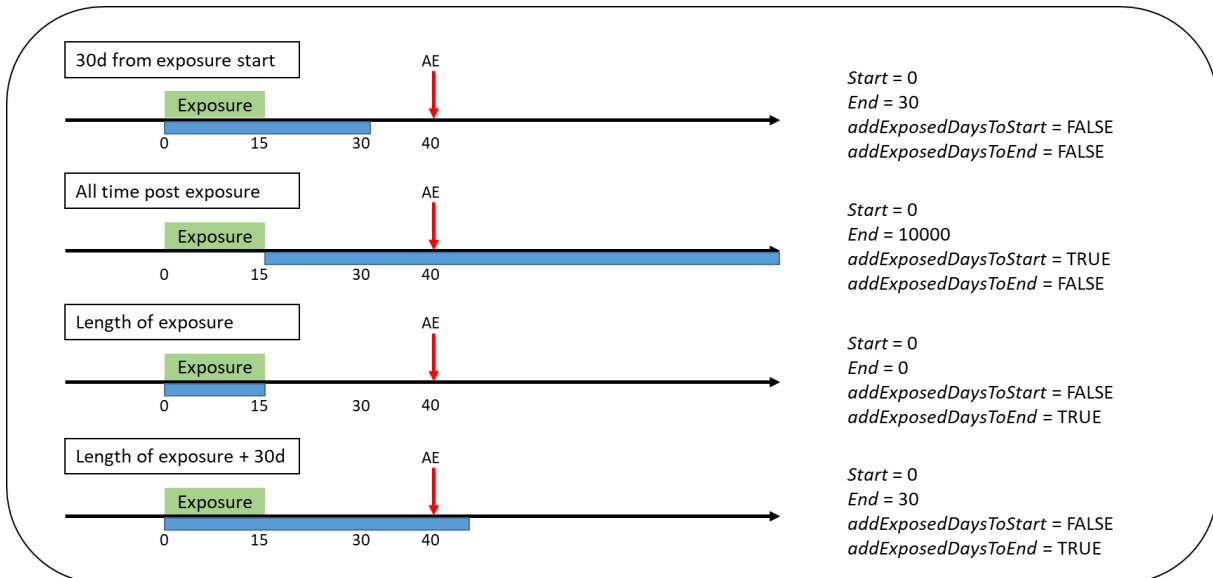


Figure 5 : Définition d'une période à risque pour le self-controlled case series

La flèche rouge correspond à l'événement indésirable. En fonction des arguments choisis, la période à risque (surface bleue) sera définie par rapport à la période d'exposition (surface verte)

B. Spécifications techniques

L'application a été développée sur le logiciel R version 3.4.4 (72) et RShiny qui permet le développement d'applications web interactives (73). L'environnement OHDSI fournissait les packages (présents sur GitHub) pour :

- La connexion et la réalisation des requêtes sur les bases de données grâce aux packages *DatabaseConnector* (74), *SQLRender* (75) ;
- La réalisation des analyses autocontrôlées : *self-controlled case series* (76), *self-controlled cohort* (77), *case-crossover* (78).

Enfin pour la restitution, les tableaux et graphiques étaient créés grâce aux packages *tidyverse* (79), *DT* (80) et *ggplot2* (81).

IV. Données test

Deux bases de données ont été utilisées pour développer et tester l'application.

1. CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use File (SynPUF)

SynPUF est une base de données simulée à partir de *Medicare* (82), créée par le *Centers for Medicare & Medicaid Services* avec pour objectifs de :

- Permettre aux entrepreneurs de développer et créer des logiciels et applications qui puissent être utilisées sur des données réelles *Medicare* ;
- Permettre aux chercheurs de s'entraîner sur des méthodes statistiques en prévision d'une analyse sur les données réelles ;
- Développer le *datamining*.

Cette base de données contient près de 2,5 millions de patients. La communauté d'OHDSI a réalisé les transformations nécessaires pour que cette base soit au format du CDM OMOP et offre donc la possibilité aux chercheurs de tester l'environnement OHDSI sur une base test. Dans le cadre de ce travail, un échantillon des 115 000 premiers patients a été utilisé pour développer l'application.

2. Les données du dispositif PAERPA

Le parcours des personnes âgées en risque de perte d'autonomie (PAERPA) est un dispositif du Ministère des Solidarités et de la Santé déployé depuis 2014 dans des territoires pilotes (83). PAERPA s'adresse aux personnes, âgées de 75 ans et plus, dont l'autonomie est susceptible de se dégrader pour des raisons d'ordre médical ou social. Il se concentre sur 5 actions : Renforcer le maintien à domicile, améliorer la coordination des intervenants et des interventions, sécuriser la sortie d'hôpital, éviter les hospitalisations inutiles et mieux utiliser les médicaments.

PAERPA est actuellement en cours d'évaluation dans les différents territoires pilotes. Dans le territoire du Valenciennois-Quercitain, cette évaluation passe par l'analyse des données SNIIRAM-PMSI pour les patients éligibles.

Pour les données PAERPA, la mise au format CDM OMOP a été réalisée et est présentée dans la partie résultats.



Figure 6 : PAERPA (83)

L'extraction des données SNIIRAM-PMSI PAERPA a été effectuée par l'ARS pour les années 2014 à 2016 et contient un ensemble de données réparties dans 27 tables :

- Les extractions PMSI pour les domaines Médecine Chirurgie Obstétrique (MCO), Soins de Suites et de Réadaptations (SSR), Hospitalisation à Domicile (HAD) et Psychiatrie (PSY) ;

- Les extractions SNIIRAM sur les données de consultations de ville médicales, paramédicales et de délivrance des médicaments.

La principale étape pour la mise en base des données était la répartition des informations des données sources PAERPA vers le modèle de données OMOP. Pour les données MCO, elles étaient réparties sur 5 tables : informations générales sur les séjours (table MCO), détails des diagnostics (table MCO_DIAG), des actes (table MCO_ACTE), des dispositifs médicaux implantables (table MCO_DMI) et des molécules onéreuses (table MCO_MED). Le tableau 4 détaille les informations présentes dans les tables ainsi que les tables de destination OMOP correspondantes. Une table source pouvait donc être répartie dans plusieurs tables du CDM OMOP.

A l'inverse, la table *visit_occurrence* était alimentée par les informations sur les dates d'hospitalisations des tables en MCO, SSR, HAD et PSY mais aussi par tous les autres contacts avec le système de soins : délivrance de médicaments en pharmacie, passage infirmier, séance de kinésithérapie, examen de radiologie et consultation chez un médecin.

Tableau 4 : Distribution des données MCO dans les différentes tables OMOP

Table Source	Informations sources	Tables OMOP
MCO	Patient	<i>Visit_occurrence</i>
	Etablissement	<i>Visit_occurrence</i>
	Année et mois d'entrée	<i>Visit_occurrence</i>
	Mode d'entrée et de sortie	<i>Visit_occurrence</i>
	Diagnostic principal et relié	<i>Condition_occurrence</i>
	Nombre de diagnostics associés	<i>Observation</i>
	Nombre d'actes CCAM	<i>Observation</i>
	Groupe homogène de malade (GHM) Groupe homogène de séjour (GHS)	<i>Observation</i>
MCO_DIAG	Diagnostics associés	<i>Condition_occurrence</i>
MCO_ACTE	Actes CCAM	<i>Procedure_occurrence</i>
MCO_DMI	Dispositifs médicaux implantables LPP	<i>Device_exposure</i>
MCO_MED	Molécules onéreuses UCD	<i>Drug_exposure</i>

Résultats

I. Description des données mises en base

Les dimensions des principales tables chargées dans le CDM OMOP sont détaillées dans le tableau 5 pour les données SynPUF et PAERPA. La table *visit_occurrence* de SynPUF ne contenait que les hospitalisations ou les consultations médicales. Dans PAERPA, elle contenait aussi les passages infirmiers, kiné et les passages en pharmacie pour délivrance ce qui explique son volume plus important. La table *care_site* contient les informations sur l'ensemble des établissements de santé de France. Enfin, certains concepts ont été ajoutés dans la table *concept*, notamment pour les classifications françaises comme la CCAM, la NGAP (Nomenclature générale des actes professionnels), CIP13, etc. Cela explique la différence du volume de concepts présents entre les deux tables.

Tableau 5 : Dimensions des principales tables dans le CDM OMOP en nombre de lignes

Table	SynPUF	OMOP
<i>Person</i>	116 352	38 382
<i>Visit occurrence</i>	5 579 542	10 745 675
<i>Condition occurrence</i>	14 455 993	1 212 301
<i>Drug exposure</i>	6 303 388	6 418 243
<i>Procedure occurrence</i>	13 926 771	16 012 480
<i>Care site</i>	239 158	94 266
<i>Concept</i>	2 933 389	4 410 368

Concernant les données PAERPA, il y avait donc 38 382 patients enregistrés dans la base. Les données enregistrées étaient comprises entre le 1^{er} janvier 2014 et le 31 décembre 2016. Le tableau 6 présente quelques indicateurs descriptifs généraux relatifs aux données PAERPA.

Tableau 6 : Analyse descriptive PAERPA

Variabes	Effectifs	Nombre de patients	Médiane [Q1 ; Q3]
Hospitalisations			
Hospitalisations en MCO	114 297	24 172	2 [1 ; 4]
Diagnostics CIM10 en MCO	805 975	24 435	3 [2 ; 10]
Molécules onéreuses en MCO	9 633	785	2 [1 ; 9]
Actes CCAM en MCO	395 109	22 973	9 [5 ; 18]
Médecine de ville			
Consultations de spécialité	174 125	23 704	5 [2 ; 10]
Consultations de médecine générale	958 339	29 680	30 [17 ; 43]
Passage en pharmacie	815 545	29 341	33 [22 ; 36]
Médicaments délivrés en pharmacie	6 408 610	29 598	202 [109 ; 302]
Kinésithérapie	1 451 753	16 568	40 [16 ; 107]
Passage de soins infirmiers	7 018 538	26 965	28 [8 ; 108]

II. Présentation de l'application

L'application a pu être développée et testée entièrement sur la base SynPUF et est présentée à travers l'exemple de l'association connue entre hypokaliémie et diurétiques thiazidique (hydrochlorothiazide) (84). Tous les résultats présents dans captures d'écrans proviennent d'une base simulée et ne reflètent donc pas la réalité.

A. Sélection de la population d'intérêt

Pour l'exemple, l'ensemble de la population est sélectionné sans restriction sur l'âge, le sexe ou les pathologies chroniques. Au total 116 352 patients sont inclus pour la suite de l'analyse.

The screenshot shows the OHDSI analyses interface with the following elements:

- 1** Parameters to select cohort: A list of checkboxes for 'Age', 'Sex', and 'Chronic conditions', all of which are checked.
- 2** Age: A range selector set from 0 to 112. Sex: Radio buttons for 'Female' and 'Male', both of which are selected.
- 3** Run Selection: A button to execute the selection.
- Sortie**: A label indicating the output of the selection process.
- Summary selection**: A table summarizing the selection criteria and the resulting patient count.

Variables	Selection
Age	From 0 to 112
Sex	Female; Male
Conditions	No selection on conditions
Result	Inclusion of 116352 patients (select row for details)

Figure 7 : Sélection de la population d'intérêt

B. Analyses descriptives

La figure 8 est issue de la première partie des analyses descriptives et représente les données démographiques de la population sélectionnée. La base SynPUF provenant de Medicare, on retrouve les mêmes distributions pour l'âge et le sexe : majoritairement des patients âgés de plus de 65 ans et des patients en situation de handicap pour les moins de 65 ans (85).

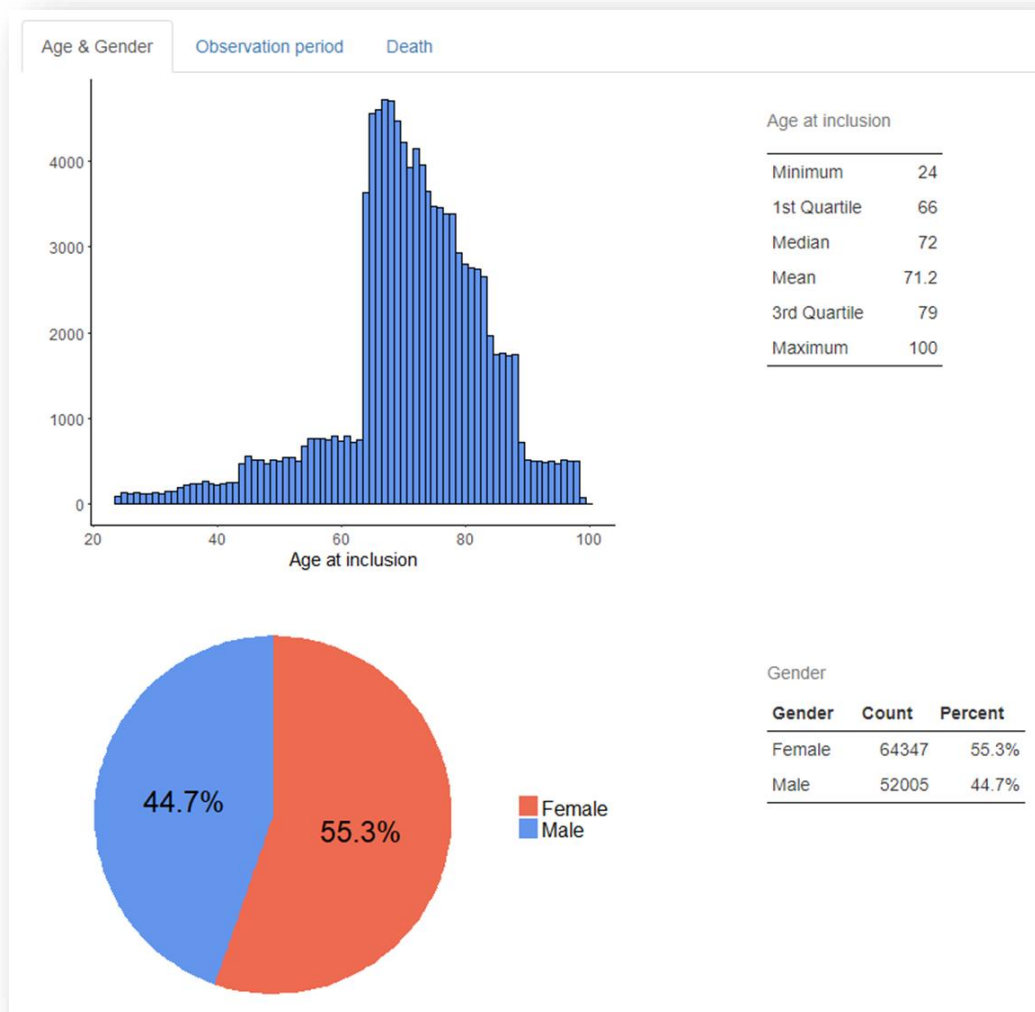


Figure 8 : Analyse descriptive des données démographiques

Sur la partie permettant d'explorer les diagnostics, actes et médicaments, dans la base SynPUF, on observe que le diagnostic d'hypokaliémie a été posé 24 435 fois pour un total de 17 645 patients. Chez ces patients, la comorbidité retrouvée la plus fréquemment était le diabète de type 2, chez 16872 patients. Chez ces patients, on peut observer que 9 982 patients ont eu une exposition à l'hydrochlorothiazide (Figures 9 et 10).

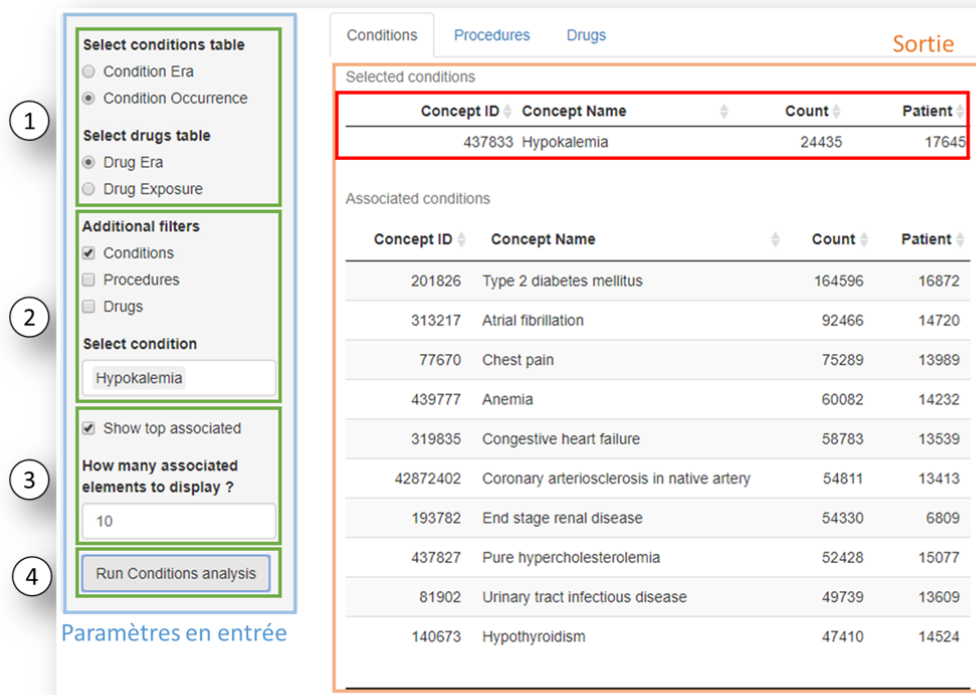


Figure 9 : Analyse descriptive des diagnostics

Recherche dans la table *condition_occurrence* des patients avec un diagnostic d'hypokaliémie (base SynPUF) ainsi que les 10 diagnostics les plus fréquemment associés chez ces patients.

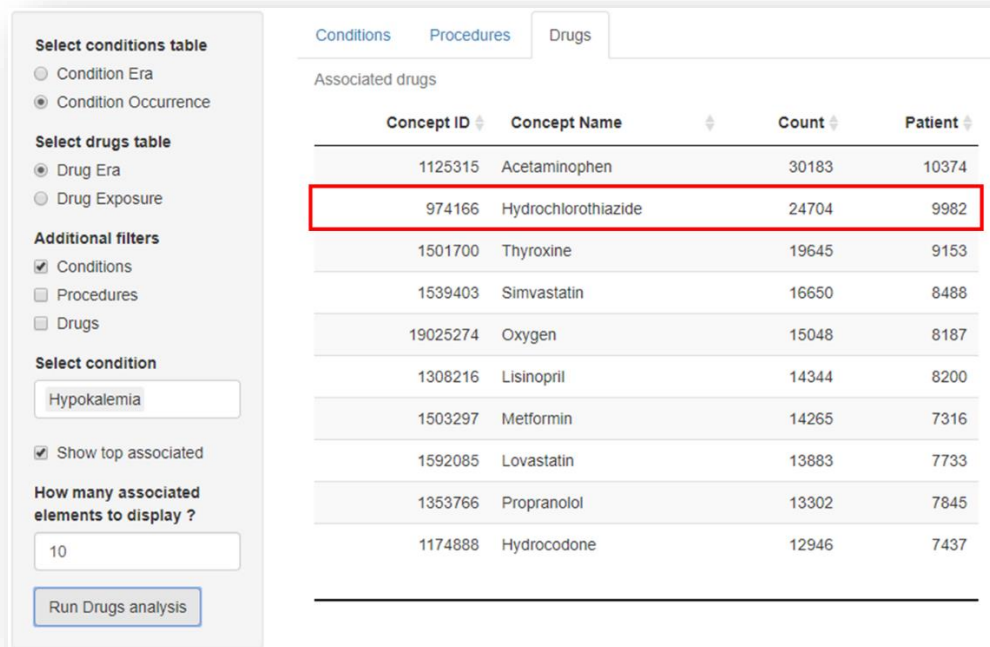


Figure 10 : Analyse descriptive des médicaments

Présentation des 10 médicaments les plus fréquemment retrouvés chez les patients ayant eu un diagnostic d'hypokaliémie (base SynPUF).

C. Analyses autocontrôlées

La figure 12 présente la page permettant de réaliser les analyses. Ici, les trois types analyses autocontrôlées ont été réalisées pour étudier le couple hypokaliémie et hydrochlorothiazide. Les paramètres présentés dans la figure 11 correspondent à une analyse par le design simple de *self-controlled case serie*. La période à risque correspond à la durée d'exposition au médicament, la période naïve est fixée à 365 jours et seul le premier événement d'hypokaliémie est utilisé pour l'analyse. Au total, l'association entre l'hypokaliémie et l'hydrochlorothiazide dans la base SynPUF, n'était pas significative (Incidence Rate Ratio = 1,06 [0,88 ; 1,26])

Table

Plot

Show 10 entries

Analysis ID	Method	Design	Outcome	Exposure	Estimate	SE	CI95% lower bound	CI95% upper bound
1	CCO	Simple case-crossover	Hypokalemia	Hydrochlorothiazide	1.19	1.14	0.92	1.54
2	SCC	Length of exposure, index date in exposure window	Hypokalemia	Hydrochlorothiazide	0.94	1.14	0.73	1.21
3	SCCS	Simple SCCS	Hypokalemia	Hydrochlorothiazide	1.06	1.10	0.88	1.26

Sortie

Previous 1 Next

Figure 11 : Analyses autocontrôlées

En sortie, 3 analyses ont été effectuées sur le couple hypokaliémie et hydrochlorothiazide (résultats sur la base simulée SynPUF, donc non interprétables) : CCO, SCC et SCCS. Dans les paramètres d'entrées sont présents les paramètres spécifiques au SCCS.

Discussion

L'application développée propose un outil interactif exploratoire permettant l'utilisation des méthodes autocontrôlées de pharmacoépidémiologie fournies par OHDSI. A partir de données au format du CDM OMOP, cet outil permet l'évaluation exploratoire de couples médicament / effet indésirable potentiel à partir de trois méthodes autocontrôlées : *self-controlled case series*, *self-controlled cohort* et *case-crossover*.

I. La réutilisation des données

La réutilisation des bases médico-administratives pour la recherche permet de réaliser des analyses sur des échantillons de grande taille tout en limitant considérablement les coûts de mise en œuvre (86). La puissance statistique doit néanmoins être mise en perspective avec la qualité des bases qui peut être limitée. On citera notamment la difficulté et le temps que peut prendre le renseignement exhaustif des informations liées à un patient donné (87,88). La prise en compte de cette limite peut s'effectuer de deux manières distinctes. Tout d'abord le biais de renseignement est variable selon le type d'information : certaines informations administratives (NIR, âge, etc.) et celles ayant attrait à la facturation sont de meilleure qualité. Des revues systématiques ont par ailleurs recensé les codes à utiliser pour étudier certaines pathologies (89,90). Ensuite, d'un point de vue statistique, les méthodes inférentielles sont moins influencées par la qualité du codage car il est possible de faire l'hypothèse qu'un certain nombre des codes manquants sont répartis de manière aléatoire. Le biais de codage associé à ces codes est donc non différentiel et ne se traduit que par une perte de puissance statistique.

II. Le choix du CDM OMOP

Il existe autant de formats de données que de bases médico-administratives, ce qui entraîne une interopérabilité faible entre ces bases (91). Partant de ce constat, l'utilisation d'un CDM est indispensable afin de pouvoir regrouper différentes sources et être reproductible entre études. Le CDM OMOP semble se dégager comme l'un des principaux acteurs dans le domaine de la pharmacoépidémiologie. Ainsi, il est déjà utilisé en Angleterre (92), en Corée (93), ou encore en Australie (94), toujours dans une perspective de recherche.

Les données du SNIIRAM-PMSI obtenues dans le cadre du projet PAERPA ont pu être formatées au CDM OMOP. Certaines données n'étaient pas prévues pour être nativement dans le modèle (par exemple, les modes d'entrées et de sortie du PMSI MCO ou les délimitations par semaines du PMSI SSR) et ont fait l'objet de décisions par avis d'expert pour décider de la table où les enregistrer. En plus des compétences informatiques, une connaissance approfondie des bases à transformer est nécessaire afin de relier les informations sources aux différentes tables du CDM OMOP (43). Afin d'améliorer la reproductibilité sur les bases de données françaises, il serait intéressant de pouvoir partager les décisions prises lors de ces transformations. Il faut cependant noter qu'un travail de maintenance est nécessaire afin de rester à jour à plusieurs niveaux :

- Sur les données du PMSI et du SNIIRAM : gérer les modifications annuelles des nomenclatures CIM-10, CCAM, GHM et GHS, etc. ;
- L'extraction réalisée à l'ARS reste une requête « artisanale » et est susceptible d'être modifiée dans sa restitution en fonction de la personne qui la réalise ;

- Les évolutions du CDM OMOP : OHDSI forme une communauté active améliorant régulièrement le CDM ainsi que les packages utilisés pour réaliser les analyses.

III. L'utilisation des méthodes autocontrôlées

Avant l'utilisation de ces méthodes pour de nouvelles questions de recherche, OHDSI recommande de mesurer leurs performances sur la base de données. Chaque base de données possède ses propres limites susceptibles d'impacter la performance des méthodes d'analyses. Le choix d'une méthode ne devrait donc pas être seulement basé sur un avis d'expert mais aussi sur des preuves montrant que la méthode choisie est la plus adaptée à l'objectif poursuivi (50). Dans ce but, OHDSI propose un benchmark permettant cette validation (95). Cette dernière n'a pour le moment pas été réalisée sur les données PAERPA et ce pour plusieurs raisons :

- La taille de l'échantillon semble insuffisante pour traiter l'ensemble des couples : l'extraction de données fournie par l'ARS contenait les informations pour 38 000 patients âgés de plus de 75 ans. Même s'il s'agit d'une population particulièrement à risque d'EIM, cet échantillon reste limité pour mettre en évidence des EIM rares avec une puissance suffisante.
- Pour effectuer cette validation, le transcodage des données en CIM-10 et ATC vers les nomenclatures standardisées SNOMED et RxNorm est indispensable. Lors d'un transcodage, la correspondance n'est pas toujours exacte entre les termes et une partie de l'information peut être perdue. Les classifications ATC et RxNorm par exemple, n'ont pas eu la même approche lors de leur création et la correspondance entre les deux n'est pas toujours à un ratio 1 :1.

- Les nomenclatures SNOMED et RxNorm sont assez peu connues en France et ne favoriseraient pas l'utilisation de l'outil.

Par ailleurs, OHDSI recommande de présenter les résultats d'une étude après calibration, c'est à dire après correction d'un potentiel biais présent dans les données (96). L'augmentation de la taille des échantillons analysés entraîne une augmentation de la précision des estimations mais n'a aucun effet sur les biais. Pour rechercher ces erreurs systématiques, ils proposent l'utilisation de contrôles négatifs sur lesquels les analyses seront réalisées en parallèle.

IV. L'utilisation de l'outil

D'un point de vue technique, les packages R développés par OHDSI pour utiliser leurs méthodes, sont compatibles avec des bases de grandes dimensions. La requête ciblée de la population d'intérêt directement dans la base de données limite le volume de données chargées en mémoire et permet une exécution fluide des calculs. De plus les méthodes autocontrôlées n'ont, par définition, pas besoin d'étapes d'appariement et ne sélectionnent donc que les patients d'intérêt. En pratique, les méthodes autocontrôlées présentes dans l'application nécessitent un entraînement du clinicien ou l'aide d'un méthodologiste avant d'être utilisées (63). Par exemple, l'étape de calibration décrite dans le paragraphe précédent illustre la nécessité de former l'utilisateur.

Actuellement, aucune évaluation de performance et d'utilisabilité de l'application n'ont été effectuées. Cette étape sera nécessaire dans la perspective d'une diffusion de l'outil.

V. Perspectives

Ce premier travail montre la compatibilité du SNDS avec le CDM OMOP, modèle plus orienté vers la recherche. Toutes les données PAERPA ont pu être importées dans le CDM OMOP. Les étapes suivantes de ce travail consistent en la création des tables *condition_era* et *drug_era* nécessaires pour l'utilisation des méthodes autocontrôlées dans l'application, la validation des données importées par des contrôles qualités de la base et l'intégration des données 2017 récupérées récemment.

L'entièreté du CDM OMOP n'est pas exploitée et dans le cadre de PAERPA, des données supplémentaires pourraient alimenter le modèle. Tout d'abord, les doses de médicaments pourraient être ajoutées et apporter de la finesse sur l'étude des effets indésirables. Ensuite, des données hospitalières telles que la biologie, les prescriptions ou les dossiers patients informatisés pourraient permettre d'être plus précis sur la chronologie des événements à l'intérieur d'une hospitalisation. Il existe aussi toute une partie médico-économique du modèle qui n'a pour l'instant pas été exploitée alors qu'elle permettrait par exemple l'intégration des données de remboursements de l'Assurance Maladie.

L'environnement OHDSI ne se limite pas aux méthodes autocontrôlées. Dans les approches populationnelles, les méthodes de cohorte et de cas-témoins sont aussi développées sous forme de packages R. Sur une approche individuelle, des méthodes prédictions par *machine learning* sont en développement (97). Des méthodes de *Natural Language Processing* (98), sont aussi prévues et pourraient s'interfacer avec les documents des dossiers patients informatisés comme les courriers médicaux ou les comptes rendus opératoires. Toutes ces méthodes présentes dans l'environnement OHDSI reposent sur le CDM OMOP et pourraient être ajoutées à l'application. Celle-ci doit par ailleurs être évaluée en pratique par les utilisateurs (pharmacologues,

épidémiologistes...) afin de s'améliorer et de proposer une interface au plus proche des besoins.

Pour finir, la communauté OHDSI est très active avec beaucoup de groupes de travail (99). En mars 2018, la branche européenne a organisé son premier symposium à Rotterdam (100). En France, le modèle se démocratise progressivement. L'équipe de Bordeaux, dans le cadre du projet ALCAPONE, s'inspire d'OHDSI pour travailler sur la génération de signaux de pharmacovigilance à partir du SNIIRAM (101). L'Assistance Publique des Hôpitaux de Paris a organisé en janvier un datathon sur des données de soins intensifs au format OMOP (102). A la suite de ce datathon, un groupe de travail autour du CDM OMOP appliqué aux données françaises et en cours de création avec les villes de Paris, Toulouse, Rennes et Lille.

VI. Conclusion

Nous avons développé une application permettant de rechercher des associations entre médicaments et effets indésirables potentiels en utilisant des designs autocontrôlés. Ce travail fait partie des premiers s'intéressant à la mise en base des données françaises au format OMOP (101). Au moment où l'accès au SNDS se démocratise et où des entrepôts de données se développent dans les CHU, (103,104) il semble indispensable de pouvoir transcrire dans le contexte français les méthodes de référence en pharmacoépidémiologie et cet outil est une façon de rendre accessibles ces méthodes.

Références bibliographiques

1. Lazarou J, Pomeranz BH, Corey PN. Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-analysis of Prospective Studies. *JAMA*. 1998 Apr 15;279(15):1200–5.
2. Juntti-Patinen L, Neuvonen PJ. Drug-related deaths in a university central hospital. *Eur J Clin Pharmacol*. 2002 Oct;58(7):479–82.
3. Lapatto-Reiniluoto O, Patinen L, Niemi M, Backman JT, Neuvonen PJ. Drug-Related Inadvertent Deaths in a University Hospital – A Declining Trend. *Basic Clin Pharmacol Toxicol*. 117(6):421–6.
4. DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending, as regards pharmacovigilance, Directive 2001/83/EC on the Community code relating to medicinal products for human use [Internet]. [cited 2018 Jun 30]. Available from: https://ec.europa.eu/health/sites/health/files/files/pharmacos/pharmpack_12_2008/pharmacovigilance-ia-vol1_en.pdf
5. Michel P, Lathelize M, Bru-Sonnet R, Domecq S, Kret M, Quenon JL. Enquête Nationale sur les Evénements Indésirables graves liés aux Soins 2009 (ENEIS2): description des résultats 2009. Rapp Final À DREES Ministère Trav L'emploi Santé–Février. 2011;3–3.
6. WHO Collaborating Centre for International Drug Monitoring. Surveillance de la sécurité d'emploi des médicaments: guide pour la création et le fonctionnement d'un centre de pharmacovigilance. Uppsala: The Uppsala Monitoring Centre (UMC); 2000.
7. Meyboom RHB, Egberts ACG, Edwards IR, Hekster YA, Koning FHP de, Gribnau FWJ. Principles of Signal Detection in Pharmacovigilance. *Drug Saf*. 1997 Jun 1;16(6):355–65.
8. Bangalore S, Kumar S, Messerli FH. Angiotensin-Converting Enzyme Inhibitor Associated Cough: Deceptive Information from the Physicians' Desk Reference. *Am J Med*. 2010 Nov 1;123(11):1016–30.
9. Directive 2001/20/CE du parlement européen et du conseil du 4 avril 2001 concernant le rapprochement des dispositions législatives, réglementaires et administratives des États membres relatives à l'application de bonnes pratiques cliniques dans la conduite d'essais cliniques de médicaments à usage humain [Internet]. [cited 2018 Jun 30]. Available from: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2001:121:0034:0044:fr:PDF>
10. Loi n° 2004-806 du 9 août 2004 relative à la politique de santé publique.
11. Martin K, Bégaud B, Latry P, Miremont-Salamé G, Fourrier A, Moore N. Differences between clinical trials and postmarketing use. *Br J Clin Pharmacol*.

2004 Jan;57(1):86–92.

12. Rothwell PM. External validity of randomised controlled trials: “To whom do the results of this trial apply?” *The Lancet*. 2005 Jan 1;365(9453):82–93.
13. Cochrane AL. Archie Cochrane in his own words: Selections arranged from his 1972 introduction to “effectiveness and efficiency: Random reflections on the health services.” *Control Clin Trials*. 1989 Dec 1;10(4):428–33.
14. Organisation de la pharmacovigilance nationale - ANSM : Agence nationale de sécurité du médicament et des produits de santé [Internet]. [cited 2018 Jun 30]. Available from: [http://ansm.sante.fr/Declarer-un-effet-indesirable/Pharmacovigilance/Organisation-de-la-pharmacovigilance-nationale/\(offset\)/0](http://ansm.sante.fr/Declarer-un-effet-indesirable/Pharmacovigilance/Organisation-de-la-pharmacovigilance-nationale/(offset)/0)
15. Bégaud B, Martin K, Haramburu F, Moore N. Rates of Spontaneous Reporting of Adverse Drug Reactions in France. *JAMA*. 2002 Oct 2;288(13):1588–1588.
16. Bégaud B. Dictionnaire de pharmaco-épidémiologie. 3e éd. Bordeaux: ARME-pharmacovigilance éd; 1998. 239 p.
17. Faillie J-L, Montastruc F, Montastruc J-L, Pariente A. L’apport de la pharmaco-épidémiologie à la pharmacovigilance. /data/revues/00405957/v71i2/S0040595716000342/ [Internet]. 2016 May 4 [cited 2018 Jun 30]; Available from: <http://www.em-consulte.com/en/article/1052398>
18. MONTASTRUC J-L, SOMMET A, MONTAS F, MOULIS G, BAGHERI H, DAMASE C. Pharmacoepidemiology : definition, methods and applications. 2015;12.
19. Boston Collaborative Drug Surveillance Group (1974) [Internet]. The James Lind Library. 2010 [cited 2018 Jun 30]. Available from: <http://www.jameslindlibrary.org/boston-collaborative-drug-surveillance-group-1974/>
20. La pharmaco-épidémiologie : une autre surveillance du médicament [Internet]. [cited 2018 Jun 30]. Available from: <http://www.academie-medecine.fr/wp-content/uploads/2016/06/PAGE-275-280.pdf>
21. Moulis F, Moulis G, Balardy L, Gérard S, Montastruc F, Sourdet S, et al. Exposure to atropinic drugs and frailty status. *J Am Med Dir Assoc*. 2015 Mar;16(3):253–7.
22. Weill A, Païta M, Tuppin P, Fagot J-P, Neumann A, Simon D, et al. Benfluorex and valvular heart disease: a cohort study of a million people with diabetes mellitus. *Pharmacoepidemiol Drug Saf*. 2010 Dec;19(12):1256–62.
23. Zureik M. La pharmaco-épidémiologie au service des décisions publiques. *Médecine*. 2016;12(1):29–31.
24. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public

- decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Epidemiol Sante Publique*. 2017 Oct;65 Suppl 4:S149–67.
25. Arnaud M. Développement d'un système automatisé de détection et de priorisation des signaux de sécurité adapté aux bases de données médico-administratives [PhD Thesis]. Université de Bordeaux; 2017.
 26. Schneeweiss S. Understanding secondary databases: a commentary on "Sources of bias for health state characteristics in secondary databases." *J Clin Epidemiol*. 2007 Jul;60(7):648–50.
 27. LOI n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé. 2016-41 Jan 26, 2016.
 28. Accueil | SNDS [Internet]. [cited 2018 Jul 1]. Available from: <https://www.snds.gouv.fr/SNDS/Accueil>
 29. Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 26(8):954–62.
 30. LOI n° 91-748 du 31 juillet 1991 portant réforme hospitalière. 91-748 Jul 31, 1991.
 31. Neumann A, Weill A, Ricordeau P, Fagot JP, Alla F, Allemand H. Pioglitazone and risk of bladder cancer among diabetic patients in France: a population-based cohort study. *Diabetologia*. 2012 Jul 1;55(7):1953–62.
 32. CégiDc - causes médicales de décès [Internet]. [cited 2018 Jul 1]. Available from: <http://www.cepidc.inserm.fr/>
 33. Commissioner O of the. Sentinel Initiative - Transforming How We Monitor Product Safety - FDA's Sentinel Initiative - Background [Internet]. [cited 2018 Jul 1]. Available from: <https://www.fda.gov/Safety/FDAsSentinelInitiative/ucm149340.htm>
 34. Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med*. 2012 Oct 30;6(4):e134–40.
 35. Andersen M, Bergman U, Choi N-K, Gerhard T, Huang C, Jalbert J, et al. The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf*. 22(7):700–4.
 36. Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf*. 20(1):1–11.
 37. Wisniewski AFZ, Bate A, Bousquet C, Brueckner A, Candore G, Juhlin K, et al.

- Good Signal Detection Practices: Evidence from IMI PROTECT. *Drug Saf.* 2016;39:469–90.
38. Reynolds RF, Kurz X, de Groot MCH, Schlienger RG, Grimaldi-Bensouda L, Tcherny-Lessenot S, et al. The IMI PROTECT project: purpose, organizational structure, and procedures. *Pharmacoepidemiol Drug Saf.* 2016 Mar;25 Suppl 1:5–10.
 39. Beuscart R, McNair P, Brender J, PSIP consortium. Patient safety through intelligent procedures in medication: the PSIP project. *Stud Health Technol Inform.* 2009;148:6–13.
 40. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the Science for Active Surveillance: Rationale and Design for the Observational Medical Outcomes Partnership. *Ann Intern Med.* 2010 Nov 2;153(9):600.
 41. Reisinger SJ, Ryan PB, O'Hara DJ, Powell GE, Painter JL, Pattishall EN, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. *J Am Med Inform Assoc JAMIA.* 2010;17(6):652–62.
 42. Athena [Internet]. [cited 2018 Aug 16]. Available from: <http://athena.ohdsi.org/search-terms/terms>
 43. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc JAMIA.* 2012;19(1):54–60.
 44. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a Reference Set to Support Methodological Research in Drug Safety. *Drug Saf.* 2013 Oct 1;36(1):33–47.
 45. Trifirò G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salamé G, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf.* 18(12):1176–84.
 46. DailyMed [Internet]. [cited 2018 Jul 9]. Available from: <https://dailymed.nlm.nih.gov/dailymed/>
 47. Tisdale JE, Miller DA. *Drug-induced Diseases: Prevention, Detection, and Management.* ASHP; 2010. 1134 p.
 48. Ioannidis JPA. Why Most Published Research Findings Are False. *PLOS Med.* 2005 Aug 30;2(8):e124.
 49. Stang PE, Ryan PB, Overhage JM, Schuemie MJ, Hartzema AG, Welebob E. Variation in Choice of Study Design: Findings from the Epidemiology Design Decision Inventory and Evaluation (EDDIE) Survey. *Drug Saf.* 2013 Oct 1;36(1):15–25.

50. Ryan PB, Stang PE, Overhage JM, Suchard MA, Hartzema AG, DuMouchel W, et al. A Comparison of the Empirical Performance of Methods for a Risk Identification System. *Drug Saf.* 2013 Oct 1;36(1):143–58.
51. DuMouchel W, Ryan PB, Schuemie MJ, Madigan D. Evaluation of Disproportionality Safety Signaling Applied to Healthcare Databases. *Drug Saf.* 2013 Oct 1;36(1):123–32.
52. Madigan D, Schuemie MJ, Ryan PB. Empirical Performance of the Case–Control Method: Lessons for Developing a Risk Identification and Analysis System. *Drug Saf.* 2013 Oct 1;36(1):73–82.
53. Ryan PB, Schuemie MJ, Gruber S, Zorych I, Madigan D. Empirical Performance of a New User Cohort Method: Lessons for Developing a Risk Identification and Analysis System. *Drug Saf.* 2013 Oct 1;36(1):59–72.
54. Ryan PB, Schuemie MJ, Madigan D. Empirical Performance of a Self-Controlled Cohort Method: Lessons for Developing a Risk Identification and Analysis System. *Drug Saf.* 2013 Oct 1;36(1):95–106.
55. Suchard MA, Zorych I, Simpson SE, Schuemie MJ, Ryan PB, Madigan D. Empirical Performance of the Self-Controlled Case Series Design: Lessons for Developing a Risk Identification and Analysis System. *Drug Saf.* 2013 Oct 1;36(1):83–93.
56. Schuemie MJ, Gini R, Coloma PM, Straatman H, Herings RMC, Pedersen L, et al. Replication of the OMOP Experiment in Europe: Evaluating Methods for Risk Identification in Electronic Health Record Databases. *Drug Saf.* 2013 Oct 1;36(1):159–69.
57. OHDSI – Observational Health Data Sciences and Informatics [Internet]. [cited 2018 Jul 9]. Available from: <https://www.ohdsi.org/>
58. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform.* 2015;216:574–8.
59. Schuemie MJ, Ryan PB, Hripcsak G, Madigan D, Suchard MA. A systematic approach to improving the reliability and scale of evidence from health care data. :52.
60. Duke JD, Ryan PB, Suchard MA, Hripcsak G, Jin P, Reich C, et al. Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network. *Epilepsia.* 2017;58(8):e101–6.
61. Observational Health Data Sciences and Informatics [Internet]. GitHub. [cited 2018 Jul 9]. Available from: <https://github.com/OHDSI>
62. Carroll LN, Au AP, Detwiler LT, Fu T, Painter IS, Abernethy NF. Visualization and Analytics Tools for Infectious Disease Epidemiology: A Systematic Review. *J Biomed Inform.* 2014 Oct;51:287–98.

63. Thomas J, Kielman J. Challenges for Visual Analytics
,
Challenges for Visual Analytics. *Inf Vis.*
2009 Jan 1;8(4):309–14.
64. Margolis R, Derr L, Dunn M, Huerta M, Larkin J, Sheehan J, et al. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. *J Am Med Inform Assoc JAMIA.* 2014 Nov;21(6):957–8.
65. Keim D, Qu H, Ma K. Big-Data Visualization. *IEEE Comput Graph Appl.* 2013 Jul;33(4):20–1.
66. Philip Chen CL, Zhang C-Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf Sci.* 2014 Aug 10;275:314–47.
67. Shortliffe EH, Cimino JJ, editors. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine [Internet].* 4th ed. London: Springer-Verlag; 2014 [cited 2018 Aug 18]. Available from:
[//www.springer.com/gp/book/9781447144731](http://www.springer.com/gp/book/9781447144731)
68. Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. *J Med Internet Res [Internet].* 2009 Mar 27;11(1). Available from:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2762766/>
69. Caban JJ, Gotz D. Visual analytics in healthcare – opportunities and research challenges. *J Am Med Inform Assoc.* 2015 Mar 1;22(2):260–2.
70. Caron A, Chazard E, Muller J, Perichon R, Ferret L, Koutkias V, et al. IT-CARES: an interactive tool for case-crossover analyses of electronic medical records for patient safety. *J Am Med Inform Assoc JAMIA.* 2017 Mar;24(2):323–30.
71. *CommonDataModel: Definition and DDLs for the OMOP Common Data Model (CDM) [Internet].* *Observational Health Data Sciences and Informatics;* 2018. Available from: <https://github.com/OHDSI/CommonDataModel>
72. R Core Team. *R: A Language and Environment for Statistical Computing [Internet].* Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: <https://www.R-project.org>
73. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J, RStudio, et al. shiny: Web Application Framework for R [Internet]. 2018. Available from: <https://CRAN.R-project.org/package=shiny>
74. Schuemie MJ, Suchard MA. *DatabaseConnector: A Package for Connecting to Various DBMSs.* 2018.
75. Schuemie M, Suchard M. *SqlRender: Rendering Parameterized SQL and Translation to Dialects [Internet].* 2018. Available from: <https://CRAN.R-project.org/package=SqlRender>

76. Schuemie M, Ryan P, Shaddox T, Suchard MA. SelfControlledCaseSeries: Self-Controlled Case Series. 2018.
77. Ryan P, Schuemie M. SelfControlledCohort: Population-level estimation method that estimates incidence rate comparison of exposed/unexposed time within an exposed cohort. 2018.
78. Schuemie M. CaseCrossover: Case-Crossover. 2018.
79. Wickham H, RStudio. tidyverse: Easily Install and Load the “Tidyverse” [Internet]. 2017. Available from: <https://CRAN.R-project.org/package=tidyverse>
80. Xie Y, Cheng J, Allaire JJ, Tan X, Girlich M, htmlwidgets/lib) jQuery contributors (jQuery in, et al. DT: A Wrapper of the JavaScript Library “DataTables” [Internet]. 2018. Available from: <https://CRAN.R-project.org/package=DT>
81. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2009. Available from: <http://ggplot2.org>
82. Medicare C for, Baltimore MS 7500 SB, Usa M. DE_Syn_PUF [Internet]. 2014 [cited 2018 Jul 11]. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF.html
83. Le dispositif Paerpa [Internet]. Ministère des Solidarités et de la Santé. 2017 [cited 2018 Jul 11]. Available from: <http://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/parcours-des-patients-et-des-usagers/le-parcours-sante-des-aines-paerpa/article/le-dispositif-paerpa>
84. Agarwal R. Hypertension, hypokalemia, and thiazide-induced diabetes: a 3-way connection. *Hypertens Dallas Tex* 1979. 2008 Dec;52(6):1012–3.
85. Medicare C for, Baltimore MS 7500 SB, Usa M. CMS Statistics Reference Booklet Overview [Internet]. 2015 [cited 2018 Aug 23]. Available from: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Statistics-Reference-Booklet/index.html>
86. Zhan C, Miller M. Administrative data based patient safety research: a critical review. *Qual Saf Health Care*. 2003 Dec;12(Suppl 2):ii58–63.
87. Payne TH, Corley S, Cullen TA, Gandhi TK, Harrington L, Kuperman GJ, et al. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. *J Am Med Inform Assoc JAMIA*. 2015 Sep;22(5):1102–10.
88. Oxentenko AS, West CP, Popkave C, Weinberger SE, Kolars JC. Time spent on clinical documentation: a survey of internal medicine residents and program directors. *Arch Intern Med*. 2010 Feb 22;170(4):377–80.
89. Singh JA, Kundukulam JA, Bhandari M. A systematic review of validated methods for identifying orthopedic implant removal and revision using administrative data. *Pharmacoepidemiol Drug Saf*. 2012 Jan;21 Suppl 1:265–73.
90. Tamariz L, Harkins T, Nair V. A systematic review of validated methods for

- identifying venous thromboembolism using administrative and claims data. *Pharmacoepidemiol Drug Saf.* 2012 Jan;21 Suppl 1:154–62.
91. Vest JR, Zhao H, Jasperson 'Jon, Gamm LD, Ohsfeldt RL. Factors motivating and affecting health information exchange usage. *J Am Med Inform Assoc JAMIA.* 2011;18(2):143–9.
 92. Zhou X, Murugesan S, Bhullar H, Liu Q, Cai B, Wentworth C, et al. An evaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance. *Drug Saf.* 2013 Feb;36(2):119–34.
 93. You SC, Lee S, Cho S-Y, Park H, Jung S, Cho J, et al. Conversion of National Health Insurance Service-National Sample Cohort (NHIS-NSC) Database into Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM). *Stud Health Technol Inform.* 2017;245:467–70.
 94. Rinner C, Gezgin D, Wendl C, Gall W. A Clinical Data Warehouse Based on OMOP and i2b2 for Austrian Health Claims Data. *Stud Health Technol Inform.* 2018;248:94–9.
 95. Martin J.Shuemie, M. Soledad Cepeda, Marc A.Suchard, Yuxi Tian, Alejandro Schuler, Patrick B.Ryan George Hripcsak. A benchmark for population-level estimation methods [Internet]. 2017 Oct [cited 2018 Jul 26]. Available from: http://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=resources:ohdsi_symposium_benchmark_v08.pdf
 96. Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci.* 2018 Mar 9;201708282.
 97. PatientLevelPrediction: An R package for performing patient level prediction [Internet]. *Observational Health Data Sciences and Informatics*; 2018. Available from: <https://github.com/OHDSI/PatientLevelPrediction>
 98. NLPTools: [in development] Tools to support Natural Language Processing of freetext to create structured data elements for analysis [Internet]. *Observational Health Data Sciences and Informatics*; 2018. Available from: <https://github.com/OHDSI/NLPTools>
 99. Project Workgroups [Observational Health Data Sciences and Informatics] [Internet]. [cited 2018 Jul 11]. Available from: <http://www.ohdsi.org/web/wiki/doku.php?id=projects:overview>
 100. OHDSI Symposium [Internet]. [cited 2018 Jul 11]. Available from: <http://www.ohdsi-europe.org/symposium>
 101. Thurin N, Lassalle R, Blin P, Pénichon M, Schuemie M, Gagne J-J, et al. Validation et calibration de l'approche cas–population dans le Sniiram en vue de la génération de signaux de pharmacovigilance : projet ALCAPONE. *Rev DÉpidémiologie Santé Publique.* 2017 Jun 1;65:S135.
 102. DAT-ICU Event 2018 [Internet]. dat-icu 2018. [cited 2018 Jul 11]. Available

from: <http://blogs.aphp.fr/dat-icu/>

103. La création de l'entrepôt des données de santé de l'AP-HP validée par la Cnil [Internet]. [cited 2018 Jul 24]. Available from: https://www.ticsante.com/la-creation-de-l-entrepot-des-donnees-de-sante-de-l-AP-HP-validee-par-la-Cnil-NS_3437.html
104. France D université L-. Partenariat d'envergure pour la recherche en santé sur le territoire ! [Internet]. [cited 2018 Jul 24]. Available from: https://www.univ-lille.fr/actualites/detail-actualite/?tx_news_pi1%5Bnews%5D=1193&tx_news_pi1%5Bcontroller%5D=News&tx_news_pi1%5Baction%5D=detail&cHash=1d2740023c17c52b51fb9e2e4bad30f3

AUTEUR : DEPAS Nicolas

Date de Soutenance : 17 septembre 2018

Titre de la Thèse : Proposition d'un outil intégrant l'environnement OHDSI pour la pharmacoépidémiologie

Thèse - Médecine - Lille 2018

Cadre de classement : Médecine

DES + spécialité : Santé Publique et Médecine Sociale

Mots-clés : pharmacoépidémiologie, OHDSI, bases médico-administratives, méthodes autocontrôlées, application web

Introduction : En Europe, un rapport de la commission européenne de 2008 estimait que 3 à 10% des hospitalisations avait pour cause un effet indésirable médicamenteux (EIM). La pharmacoépidémiologie vient compléter les essais cliniques randomisés et les reports de cas en pharmacovigilance pour explorer le lien entre des couples médicament et EIM potentiel. Le projet Observational Medical Outcomes Partnership (OMOP) puis le consortium Observational Health Data Sciences and Informatics (OHDSI) ont étudié empiriquement la performance des designs d'étude en pharmacoépidémiologie. Sur les bases de données médico-administratives, ils ont montré l'intérêt des designs autocontrôlés. L'objectif de ce travail était de développer une application interactive permettant l'utilisation des méthodes autocontrôlées validées par OMOP/OHDSI en pharmacoépidémiologie.

Méthodes : L'application, développée sous R à partir du package RShiny, reposait sur le modèle de données commun OMOP pour son utilisation. Elle implémentait trois méthodes autocontrôlées : le self-controlled case series, le self-controlled cohort et le case-crossover. Deux bases de données ont servi au développement et au test de l'application : SynPUF, une base de données simulée à partir des données de Medicare, et les données de l'assurance maladie pour les patients de plus de 75 ans habitant dans le Valenciennois-Quercitain rentrant dans le cadre du projet PAERPA.

Résultats : Les données SynPUF et les données de l'assurance maladie ont été mises au format du modèle de données commun OMOP. A partir de ce format de données, l'application propose une analyse pharmacoépidémiologique en 3 étapes : (1) définition de la population d'intérêt, (2) description de la population sélectionnée et des diagnostics, traitement, actes enregistrés pour celle-ci et (3) réalisation d'analyses à la volée basées sur les méthodes autocontrôlées.

Discussion : Cette application permet de réaliser de manière exploratoire des analyses de pharmacoépidémiologie. L'environnement OMOP/OHDSI ne se limite pas aux méthodes autocontrôlées et est en constant développement, laissant la porte ouverte à des mises à jour de l'application pour d'autres méthodes. Au moment où l'accès au Système National des Données de Santé se démocratise, la transcription de ces méthodes dans le contexte français pourrait s'avérer intéressante.

Composition du Jury :

Président :	Monsieur de Professeur Philippe Amouyel
Asseseurs :	Monsieur le Professeur Alain Duhamel Monsieur le Professeur Emmanuel Chazard
Directeurs de thèse :	Monsieur le Docteur Grégoire Ficheur Monsieur le Docteur Alexandre Caron