



HAL
open science

Bases médico-administratives : de l'épidémiologie à la pharmacovigilance

Aurélie Bannay

► **To cite this version:**

Aurélie Bannay. Bases médico-administratives : de l'épidémiologie à la pharmacovigilance. Santé publique et épidémiologie. Université de Lorraine, 2021. Français. NNT: 2021LORR0183. tel-03482463

HAL Id: tel-03482463

<https://hal.science/tel-03482463v1>

Submitted on 4 Dec 2023

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

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



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



Ecole Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE DE LORRAINE

Mention : « Sciences de la Vie et de la Santé »

par Aurélie BANNAY

Bases médico-administratives : de l'épidémiologie à la pharmacovigilance

Le 5 novembre 2021

Membres du jury :

Rapporteurs :	Monsieur Rodolphe THIEBAUT	PU-PH, UMR 1219 Inserm Bordeaux Population Health, Université de Bordeaux, Bordeaux
	Madame Leslie GRAMMATICO- GUILLON	MCU-PH, UMR 1259 Inserm MAVIVH, Université de Tours, Tours
Examineurs :	Monsieur Marc CUGGIA	PUPH, UMR 1099 Inserm LTSI, Université de Rennes 1, Rennes
	Madame Catherine DUCLOS	PU-PH, UMRS 1142 Inserm LIMICS, Université Sorbonne Paris Nord, Bobigny
	Madame Ndeye Coumba NDIAYE	MCU-PH, UMR 1256 Inserm NGERE, Université de Lorraine, Nancy
	Monsieur Nicolas JAY	PU-PH, UMR 7503 LORIA, Université de Lorraine, Nancy, directeur de thèse
	Monsieur François ALLA	PU-PH, UMR 1219 Inserm Bordeaux Population Health, Université de Bordeaux, Bordeaux, co-directeur de thèse

Remerciements

A mes directeurs de thèse, les Pr Nicolas Jay et François Alla

Dès le début de ma formation vous m'avez intégrée dans vos équipes et fait découvrir l'épidémiologie, l'information et l'informatique médicales mais surtout le pont qu'il existe entre ces disciplines. Je vous remercie pour votre soutien et votre bienveillance.

Soyez assurés de ma profonde reconnaissance.

Aux membres de mon jury,

Au **Pr Rodolphe Thiebaut** et **Dr Leslie Grammatico-Guillon** pour avoir accepté d'être les rapporteurs de ce travail.

Au **Pr Marc Cuggia**, **Pr Catherine Duclos**, et **Dr Ndeye Coumba Ndiaye** pour avoir accepté de juger ce travail.

Votre regard critique permettra d'enrichir ce travail.

Aux membres de mon comité de suivi de thèse : le Pr Rodolphe Thiebaut et le Dr Ndeye Coumba Ndiaye

Pour votre accompagnement tout au long de mon parcours de thèse.

Au Professeur Marc Cuggia,

Pour m'avoir accueillie dans ton équipe et pour la confiance dont tu m'as témoigné sur le projet INSHARE.

A l'équipe de l'ex-DESP de la CNAM : Alain Weill, Philippe Ricordeau, Anke Neumann, Myriam Mezzarobba, Cécile Billionnet, Marie Dalichampt, Christophe Chaignot, Pierre-Olivier Blotière, Géric Maura, Mickaël Basson, Moussa Laanani

Pour m'avoir fait découvrir le SNDS, pour leurs précieux conseils pour l'utiliser, et pour les moments partagés

Aux membres de l'équipe K : Jean Lieber, Emmanuel Nauer, Nicolas Lasolle

Pour leur accueil dans l'équipe et leur soutien

A mes collègues et amis

A **Jeanne** pour m'avoir initié à la recherche en « périnat », pour son aide dans ce travail et son amitié

A **Sylvain** pour son expertise et sa sympathie

A **Thierry, Agnès, Emilie, Jonathan, Sigrid et Marie-Caroline** pour tous ses moments partagés, parfois malgré la distance

A ma belle-famille,

Pour m'avoir fait entrer dans leur famille si naturellement

A mes parents,

Pour leur soutien inconditionnel depuis toujours

A Guillaume,

Pour sa relecture attentive, et surtout pour sa présence à mes côtés

The pure and simple truth is rarely pure and never simple.

Oscar Wilde

Table des matières

Liste des abréviations	7
Liste des tableaux	9
Liste des figures	10
Valorisation scientifique.....	11
Articles publiés	11
Articles soumis	11
Article en préparation	11
1 Introduction.....	12
1.1 Généralités sur les bases de données médico-administratives	12
1.1.1 Définition des bases de données médico-administratives.....	12
1.2 Exemples internationaux de BMA	13
1.3 Les BMA en France :	15
1.3.1 Programme de Médicalisation des Systèmes d'Information (PMSI).....	15
1.3.2 Du Système National d'Information Inter-Régimes de l'Assurance Maladie : SNIIRAM à la base principale du Système National de Données de Santé : SNDS.....	19
1.4 Synthèse des principales données disponibles dans le SNDS	24
1.5 Ré-utilisation des BMA en recherche	25
2 Objectif de thèse et axes de recherche.....	27
3 Axe 1 : montrer l'apport de travaux méthodologiques pour prendre en compte les spécificités des données issues de bases médico-administratives	28
3.1 Adaptation et validation du score de comorbidités de Charlson aux données du SNDS: Etude 1	28
3.1.1 Contexte	28
3.1.2 Méthode.....	29
3.2 Résultats	30
3.2.1 Synthèse de l'étude 1	44
3.3 Construction et validation d'un algorithme d'identification des hospitalisations non programmées dans le PMSI chez les patients atteints de maladie inflammatoire chronique de l'intestin : Etude 2	46
3.3.1 Contexte	46
3.3.2 Méthode.....	46
3.3.3 Résultats.....	47
3.3.4 Synthèse de l'étude 2	67
3.4 Complémentarité des données issues des bases médico-administratives et d'un entrepôt hospitalier : illustration par une étude de cohorte sur les patients consommateurs de statine : Etude 3	68

3.4.1	Contexte	68
3.4.2	Méthode	68
3.4.3	Résultats	69
3.4.4	Synthèse de l'étude 3	97
3.5	Synthèse de l'axe 1	98
4	Axe 2 : illustrer, au travers de cas d'usage, l'apport de l'utilisation des données médico-administratives en épidémiologie et en pharmacovigilance	99
4.1	Impact des comorbidités maternelles sur l'issue de grossesse et le devenir de l'enfant : Etude 4	99
4.1.1	Contexte	99
4.1.2	Méthodes	100
4.1.3	Résultats	100
4.1.4	Synthèse de l'étude 4	121
4.2	Pronostic des patients atteints de maladie inflammatoire chronique après un syndrome coronarien aigu : Etude 5	122
4.2.1	Contexte	122
4.2.2	Méthodes	122
4.2.3	Résultats	123
4.2.4	Synthèse de l'étude 5	137
4.3	Effet de l'indice de masse corporelle sur la mortalité de patients atteints de COVID-19 pris en charge en réanimation : Etude 6	138
4.3.1	Contexte	138
4.3.2	Méthode	138
4.3.3	Résultats	139
4.3.4	Synthèse de l'étude 6	150
4.4	Synthèse de l'axe 2	151
5	Discussion générale	152
5.1	Principaux résultats	152
5.2	Place des données issues des BMA en recherche	153
6	Conclusion et perspectives	157
7	Références	159

Liste des abréviations

ALD	Affection longue durée
ANSM	Agence nationale de sécurité du médicament et des produits de santé
ARS	Agence régionale de santé
ATIH	Agence technique de l'information hospitalière
BMA	Base de données médico-administratives
CNAM	Caisse nationale d'assurance maladie
CNMSS	Caisse nationale militaire de sécurité sociale
CépiDc	Centre d'épidémiologie sur les causes médicales de décès
CHU	Centre hospitalier universitaire
CNRS	Centre national de la recherche scientifique
CCAM	Classification commune des actes médicaux
CIM10	Classification statistique internationale des maladies et des problèmes de santé connexes 10 ^{ème} révision
CPRD	Clinical Practice Research Datalink
CIP	Code identifiant de présentation
CESREES	Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé
CNIL	Commission nationale de l'informatique et des libertés
DRG	Diagnosis Related Group
GPRD	General Practice Research Database
GHM	Groupe homogène de malades
GHS	Groupe homogène de séjours
HAS	Haute autorité de santé
HDH	Health Data Hub
HMO	Health Maintenance Organization
HAD	Hospitalisation à domicile
IGS II	Indice de gravité simplifié II
IMC	Indice de masse corporelle
INSERM	Institut national de la santé et de la recherche médicale
MICI	Maladies inflammatoires chroniques de l'intestin

MCO	Médecine chirurgie obstétrique et odontologie
MSA	Mutualité sociale agricole
NIR	Numéro d'inscription au répertoire des personnes physiques
OMS	Organisation mondiale de la santé
PMSI	Programme de médicalisation des systèmes d'information
PMI	Protection maternelle infantile
RSI	Régime social des indépendants
ReDSiam	Réseau pour mieux utiliser les données du système national des données de santé
RUM	Résumé d'unité médicale
RSA	Résumé de sortie anonyme
RSS	Résumé de sortie standardisé
SSR	Soins de suite et de réadaptation
SNDS	Système national des données de santé
SNIIRAM	Système national d'information inter-régimes de l'assurance maladie
T2A	Tarifcation à l'activité
VPP	Valeur prédictive positive
VHA	Veterans Health Administration

Liste des tableaux

Tableau 1: Evolution des finalités du SNIIRAM/SNDS.....	21
Tableau 2: Disponibilité des données dans le SNDS	24

Liste des figures

Figure 1: Production de l'information médicale : du RUM au RSA	19
Figure 2: Produits du SNDS historique	23
Figure 3: Tables du DCIR	24

Valorisation scientifique

Articles publiés / acceptés

Bannay A, Chaignot C, Blotière PO, Basson M, Weill A, Ricordeau P, Alla F. **The Best Use of the Charlson Comorbidity Index With Electronic Health Care Database to Predict Mortality.** *Medical Care.* 2016; 54:188-94.

Dana R*, Bannay A*, Bourst P, Ziegler C, Losser MR, Gibot S, Levy B, Audibert G, Ziegler O. **Obesity and mortality in critically ill COVID-19 patients with respiratory failure.** *Int J Obes (Lond).* 2021; 45:2028-2037 (*co-premiers auteurs)

Bannay A, Bories M, Le Corre P, Riou C, Lemordant P, Van Hille P, Chazard E, Dode X, Cuggia M, Bouzillé G. **Leveraging national claim and hospital big data integration: a cohort study on a statin-drug interaction use case.** Accepté dans *Journal of medical Internet Research(JMIR) Medical Informatics*

Articles soumis

Bannay A, Guardiolle V, Baillet S, Jay N, Gauchotte E, Deneux-Tharoux C, Fresson J. **Impact of maternal Charlson index's comorbidities on pregnancy outcome: a population-based cohort study.** Soumis au *British Journal of Obstetrics and Gynaecology*

Person L, Dominique Y, Jay N, Bannay A. **Construction and evaluation of an algorithm to detect unplanned hospitalizations in patients with chronic inflammatory bowel disease in the French national discharge database.** Soumis au *BMC Medical informatics and decision making*

Article en préparation

Popovic B, Y Juillère Y, Jay N, Bannay A. **Outcomes after acute coronary syndrome in patients with inflammatory bowel disease**

1 Introduction

En santé, comme dans de nombreux domaines, le volume des données rendues disponibles grâce à la dématérialisation, constitue un nouvel enjeu quant à leur réutilisation, non seulement en recherche, mais également dans le pilotage des structures sanitaires, dans l'évaluation des politiques de santé, dans la sécurité et la veille sanitaire.

Les sources de données en santé sont multiples, elles peuvent être distinguées selon différentes caractéristiques telles le type de données (structurées ou non), leur finalité ou leur origine en fonction du producteur de données : le patient au travers des objets connectés, le professionnel/la structure de soins ou de prévention au travers du dossier médical, les états au travers des registres d'état civil, les registres de morbidités, les organismes payeurs de soins au travers des remboursements. Ces derniers sont à l'origine des bases de données médico-administratives (BMA).

1.1 Généralités sur les bases de données médico-administratives

1.1.1 Définition des bases de données médico-administratives

Les bases de données médico-administratives ne répondent pas à une définition consensuelle. Dans la littérature médicale, on retrouve le concept de « claims database », ou d' « administrative (medical) database » ou encore de « real-world database ». Leur caractérisation peut être large comme celle donnée par Gavriellov-Yusim (1) « administrative databases are massive repositories of data collected in healthcare for various purposes » ou plus ciblée telle que celle décrite par Peek (2) « administrative data [...] describe those that are not primarily clinical, but which may contain data that are related to the clinical enterprise. These data are generated as a result of patient encounters and consist primarily of billings which are based on diagnosis and procedure codes ». En France, un rapport du Haut Conseil de la Santé Publique publié en 2012 (3) « Pour une meilleure utilisation des bases de données administratives et médico-administratives nationales pour la santé publique et la recherche » les décrit comme des bases « centralisées, constituées et gérées par des organismes publics, couvrant de façon exhaustive et permanente l'ensemble de la population dans divers domaines stratégiques pour la

santé publique et la recherche : recours aux soins, hospitalisation, handicaps, prestations et situation professionnelle, sociale et économique des bases ». Ces définitions font apparaître les finalités et les champs d'utilisation de ces bases médico-administratives.

1.2 Exemples internationaux de BMA

Aux Etats-Unis, des bases médico-administratives existent tant dans le domaine public que privé.

Dans le domaine public, la Medicare claims database regroupe les données des sujets pris en charge par le programme Medicare qui couvre entre 95 et 98 % des sujets à partir de 65 ans et ceux avant 65 ans présentant un handicap et ceux avec une insuffisance rénale terminale, quel que soit leur âge(4,5). Cette base comporte 4 grandes sections : la section A concerne les établissements de santé avec les données d'hospitalisations, la section B concerne les professionnels de santé avec les données de soins hors hospitalisation, la section C concerne les Health Maintenance Organization (HMO) avec les données de soins dans ce type de prise en charge, la section D concerne les délivrances médicamenteuses en officine.

Il existe également le programme Medicaid, géré au niveau de chaque état qui couvre les personnes les moins favorisées sans condition d'âge (6,7). Son expansion dans le cadre de la loi Obamacare été adoptée de façon variable selon les états. Les données disponibles sont similaires à celles de Medicare, mais avec une couverture non représentative au niveau national, car les critères d'éligibilité et les taux de prise en charge diffèrent selon les états.

La Veterans Health Administration (VHA) couvre, quant à elle, les vétérans ayant servi durant au moins 24 mois de service actif. Elle regroupe un nombre important de base de données sur la prise en charge de cette population au niveau national, en particulier sur les prises en charge hospitalières et extra-hospitalières (diagnostics, actes médicaux), les examens complémentaires (examens radiologiques, résultats biologiques), les délivrances médicamenteuses et le statut vital (8–10).

La Kaiser Permanente est un système de soins privé à but non lucratif implanté dans 8 états américains.

Il a mis en place une base de données pilote au niveau de la Caroline du Nord (Kaiser Permanente

Northern Carolina) regroupant les données d'hospitalisation et de prise en charge extra-hospitalière enrichies de données cliniques issues des dossiers médicaux électroniques. Ce projet pilote a été suivi d'une extension au niveau national (National Research Database of Kaiser Permanente) (11,12) visant l'intégration de près de 30 millions de patients.

Au Canada, le système de soins est financé au niveau fédéral, mais géré par chaque province. Si les données concernant les hospitalisations et les prises en charge extra-hospitalières sont disponibles, la profondeur d'historique n'est pas similaire entre les provinces remontant de façon variable de 1983 à 1997. Les critères de couverture des délivrances médicamenteuses diffèrent également selon les provinces : 66 ans et plus, personnes à faible revenu (13).

Au Royaume Uni, la Clinical Practice Research Datalink (CPRD), anciennement General Practice Research Database (GPRD), regroupent les données de la patientèle (file active de 16 millions de patients) d'un échantillon de plus de 1 800 médecins généralistes britanniques en 2020 (14,15). Ces données peuvent être chaînées, sur demande, aux données d'hospitalisation et de prise en charge extra-hospitalière, aux données de mortalité, et à celles de registres de pathologies spécifiques (cancers, santé mentale) pour environ 88 % des sujets suivis (16).

Ce panorama anglo-saxon fait ressortir des données non exhaustives en matière de population couverte ou de remboursements. Ces modèles contrastent fortement avec ceux proposés dans les pays scandinaves.

Ainsi, le Danemark a une longue histoire de collecte de données dans des registres d'état civil, d'immigration et de pathologies. Actuellement, un identifiant unique attribué à chaque habitant (et à chaque entreprise) permet de faire le lien entre les données issues de multiples registres spécifiques de chaque pathologie (diabète, psychiatrie, cancer colorectal, sclérose en plaque...), les données de soins primaires mais également des hospitalisations (17).

Le système français se situe entre ces 2 types de modèles.

1.3 Les BMA en France :

1.3.1 Programme de médicalisation des systèmes d'information

1.3.1.1 *Historique*

Le programme de médicalisation des systèmes d'information (PMSI) a été introduit en France au début des années 80 avec pour fonction de décrire et analyser l'activité médico-économique des établissements de santé (18). L'obligation de cette analyse médicalisée pour les établissements sanitaires a été inscrite dans la loi en 1991 afin de « tenir compte notamment des pathologies et des modes de prise en charge en vue d'améliorer la connaissance et l'évaluation de l'activité et des coûts et de favoriser l'optimisation de l'offre de soins » (19). Après des expérimentations régionales, le PMSI a été généralisé pour les établissements publics en 1989, puis pour les établissements privés en 1996 (20). Initialement présenté comme étant à visée épidémiologique, cet outil est actuellement utilisé pour le financement des établissements de court séjour (médecine, chirurgie, obstétrique et odontologie (MCO)) dans le cadre de la tarification à l'activité (T2A). Initialement limitée à hauteur de 10 % du financement en 2004 (21), la part de la T2A a progressivement augmenté jusqu'à atteindre 100 % en 2008 (22). Parmi les autres champs du PMSI (psychiatrie, soins de suite et réadaptation (SSR) et hospitalisation à domicile (HAD)), seul le SSR a commencé à introduire une part de financement basée sur l'activité mesurée par le PMSI dans le cadre de la « dotation modulée à l'activité ». Dans la suite de cette thèse, seul le PMSI situé dans le champ MCO sera développé.

1.3.1.2 *Principe*

Décliné à partir du modèle des Diagnosis Related Groups (DRGs) développé aux Etats-Unis par le Pr Fetter dans les années 80 (23), le principe du PMSI est de classer chaque séjour hospitalier dans l'un des groupes homogènes de malades (GHM) prédéfinis. Au sein d'un même GHM, les séjours présentent des similarités à la fois en termes de prise en charge médicale, mais également de coût des ressources mobilisées. A chaque GHM correspond un ou plusieurs groupes homogènes de séjour (GHS) permettant de lui affecter un tarif.

Cette étape de classement d'un séjour donné dans une catégorie de séjours se fait à partir d'un recueil standardisé de données administratives et médicales (24) puis d'un algorithme (fonction de groupage) mis à jour chaque année.

1.3.1.3 Contenu

Actuellement les données du PMSI sont accessibles au travers de 2 points d'accès : les bases nationales accessibles par l'Agence technique de l'information hospitalière (ATIH) (25), les bases PMSI intégrées au Système National des Données de Santé (SNDS) accessibles par le portail du SNDS de la Caisse nationale d'assurance maladie (CNAM) (26) ou via des extractions ad hoc. Dans ces 2 points d'accès, le modèle physique de données, les noms des tables et des variables ne sont pas identiques. Toutefois le contenu est similaire, à l'exception des dates exactes d'hospitalisation disponibles dans le SNDS, alors que seuls des délais par rapport à une date fixe pérenne spécifique à chaque individu sont calculés dans les bases de l'ATIH. La description des données, dans la suite de ce travail, se focalise sur les données disponibles dans le SNDS.

Au sein de chaque établissement, les données administratives et médicales sont rassemblées dans une synthèse produite à l'issue de chaque séjour hospitalier : le résumé de sortie standardisé (RSS) constitué d'un ou plusieurs résumés d'unité médicale (RUM) produits pour chaque passage dans une unité médicale. Après un processus de pseudonymisation, les établissements transmettent mensuellement les résumés de sortie anonymes (RSA) des séjours sortis depuis le 1^{er} janvier de l'année en cours à l'ATIH.

Le processus de pseudonymisation retire les identifiants de l'unité médicale sauf les autorisations, remplace l'identifiant administratif du séjour par un identifiant indépendant de l'établissement, la date de naissance par l'âge à l'entrée du séjour, le code postal par le code géographique. Depuis 2001, les différentes hospitalisations d'un même patient peuvent être chaînées grâce à un numéro anonyme d'hospitalisation obtenu à partir du numéro d'assuré social de l'ouvrant droit, du sexe et de la date de naissance du bénéficiaire (27). Depuis le 1^{er} mars 2011 (28), un numéro de chaînage entre le séjour d'accouchement de la mère et le séjour de naissance de l'enfant est disponible dans le PMSI.

On peut distinguer les principales données en fonction des entités qu'elles décrivent :

- ✓ patient : identifiant anonyme, âge, sexe, code géographique de résidence, poids pour les nouveau-nés, délai de la date des dernières règles par rapport à la date d'entrée / âge gestationnel pour les femmes enceintes
- ✓ séjour hospitalier (RSA) : numéro de séjour anonyme, numéro FINESS de l'établissement, date d'entrée et de sortie, mode d'entrée/sortie, destination/provenance, GHM, GHS, nombre de suppléments (réanimation, soins continus, soins intensifs, dialyse, caisson hyperbare, antepartum, néonatalogie, radiothérapie, défibrillateur cardiaque) diagnostic principal, diagnostic relié, diagnostic(s) associé(s), nombre de RUM, actes
- ✓ passage dans l'unité médicale (RUM) : numéro séquentiel du RUM dans le RSS, autorisation de l'unité médicale, durée du RUM, indice de gravité simplifié II (IGS II) pour les patients pris en charge en réanimation, diagnostic principal du RUM, diagnostic relié du RUM.

Le recueil standardisé de ces données repose sur des règles publiées et mises à jour chaque année dans un fascicule spécial du bulletin officiel intitulé « Guide méthodologique de production des informations relatives à l'activité médicale et à sa facturation en médecine, chirurgie, obstétrique et odontologie ». Ce guide spécifie également les nomenclatures utilisées. Concernant les diagnostics, le codage repose sur la classification statistique internationale des maladies et des problèmes de santé connexes dans sa 10^{ème} révision (CIM10) à usage du PMSI (29) qui correspond à une version modifiée de la CIM10 de l'Organisation mondiale de la santé (OMS) notamment par l'ajout de subdivisions spécifiques, de notes et des modifications de libellés . Une entité nosologique est identifiée par un code unique défini par, au minimum, 3 caractères (une lettre et 2 chiffres). Cette classification mono-axiale regroupe les pathologies en 22 chapitres (tumeurs, maladie de l'appareil respiratoire, etc.). A l'intérieur de chaque chapitre (correspondant au 1^{er} caractère, la lettre), l'arborescence descend jusqu'à un niveau de détail (extension en chiffres) qui dépend de la fréquence de la maladie.

Les actes médicaux sont codés à l'aide d'une nomenclature française : la classification commune des actes médicaux (CCAM) gérée par l'assurance maladie. A chaque acte correspond un code unique sur 7 caractères : les 2 premiers caractères en lettres sont relatifs à la topographie, le 3^{ème} caractère en lettre à l'action, le 4^{ème} caractère en lettres à la technique et les 3 derniers chiffres sont un compteur.

Ces codes sont organisés en 19 chapitres correspondant aux appareils ou structures anatomiques et en sous chapitres selon leur visée diagnostique et thérapeutique.

Des extensions documentaires sur 3 caractères (telles que la latéralité d'un acte) ont été mises en place dans la CCAM descriptive à usage du PMSI depuis 2015.

Certains concepts spécifiques du PMSI nécessitent de s'attarder sur leurs définitions issues du guide méthodologique :

- ✓ *Diagnostic principal (DP)* : il est « le problème de santé qui a motivé l'admission du patient dans l'unité médicale (UM), déterminé à la sortie de l'UM » depuis 2009
- ✓ *Diagnostic relié (DR)* : il a « pour rôle, en association avec le DP et lorsque celui-ci n'y suffit pas, de rendre compte de la prise en charge du patient en termes médico-économiques. Trois modalités régissent son codage dans le RUM : i) « le DP relève du chapitre XXI de la CIM10 (codes en Z) », ii) le DR correspond à une « maladie chronique ou de longue durée ou un état permanent, présent au moment du séjour objet du résumé », iii) le DR précise « la maladie ou état [de] la prise en charge enregistrée comme DP »
- ✓ *Diagnostic associé significatif (DAS)* : il est une « affection, un symptôme ou tout autre motif de recours aux soins coexistant avec le DP-DR, et constituant : un problème de santé distinct supplémentaire ou une complication de la morbidité principale ; ou de [son] traitement »

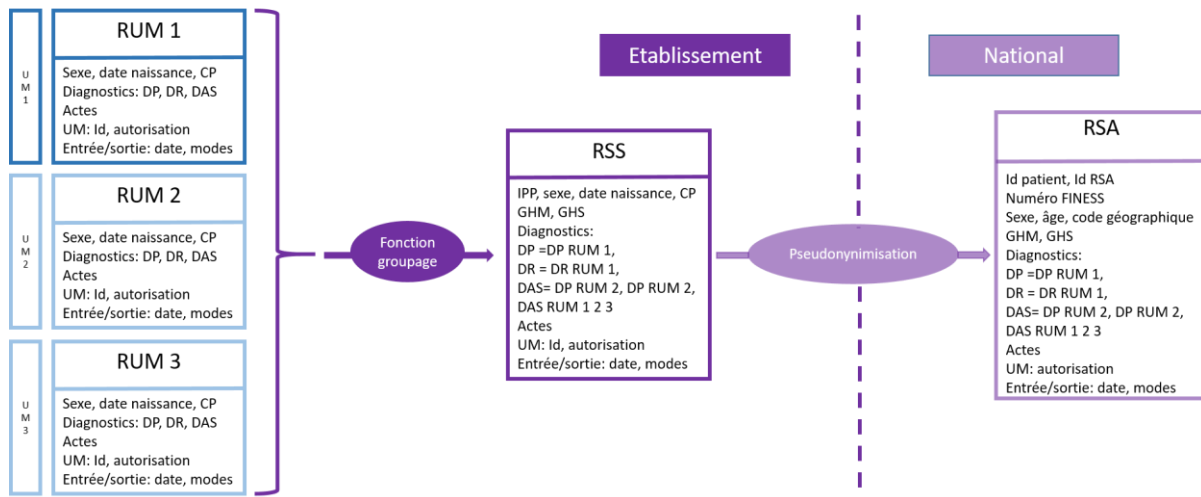


Figure 1: Production de l'information médicale : du RUM au RSA

La figure 1 résume le principe du groupage d'un séjour hospitalier dans un GHM et les principales données contenues dans les documents produits du RUM jusqu'au RSA. Il est à noter que le modèle relationnel de données du PMSI dans le SNDS ne permet pas de rattacher les DAS à un RUM contrairement au modèle relationnel de données du PMSI disponible sur la plateforme de l'ATIH. En conséquence, lors d'une étude se focalisant, par exemple, sur les passages en réanimation, il ne sera pas possible d'identifier les complications survenues spécifiquement dans ce service de soins critiques.

1.3.2 Du Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM) à la base principale du Système National de Données de Santé (SNDS)

1.3.2.1 Historique

Le SNIIRAM a été créé dans la loi (30) en 1998. En 2001, la Commission nationale de l'informatique et des libertés (CNIL) a rendu un avis favorable à sa mise en œuvre (31). Ses finalités ont été étendues dès 2002 (32) puis en 2005 (33) pour faire passer le SNIIRAM d'un outil de gestion et de planification des dépenses de l'assurance maladie à un outil de gestion des politiques de santé. Enfin, en 2017, il est devenu un outil de recherche épidémiologique et de surveillance sanitaire (Tableau 1) (34). En parallèle, l'accès des utilisateurs initialement limité à l'assurance maladie, s'est élargi entre 1998 et 2017, aux agents des agences sanitaires (Institut national de Veille Sanitaire/Santé publique France, Agence nationale de sécurité du médicament et des produits de santé (ANSM), Agence régionale de santé (ARS), Haute autorité de santé (HAS), ATIH...) et des acteurs de la recherche (Institut national de

la santé et de la recherche médicale (INSERM), Centre national de la recherche scientifique (CNRS), Centres hospitaliers universitaires (CHU)) en fonction des produits et de la profondeur d'historique. En devenant en 2017 le SNDS, son périmètre s'est agrandi.

Tableau 1: Evolution des finalités du SNIIRAM/SNDS

Finalités	1998	2002	2005	2017
Gestion des dépenses de santé	Contribuer à la connaissance des dépenses de l'ensemble des régimes d'assurance maladie par circonscription géographique, par nature de dépenses, par catégorie de professionnels responsables de ces dépenses et par professionnel ou établissement	Contribuer à une meilleure gestion de l'assurance maladie, notamment par la connaissance des dépenses de l'ensemble des régimes d'assurance maladie par circonscription géographique, par nature de dépense, par catégorie de professionnels responsables de ces dépenses et par professionnel ou établissement, l'évaluation des transferts entre enveloppes correspondant aux objectifs sectoriels de dépenses fixés, en fonction de l'objectif national de dépenses d'assurance maladie, dans le cadre de la loi annuelle de financement de la sécurité sociale	Contribuer à une meilleure gestion de l'assurance maladie, notamment par la connaissance des dépenses de l'ensemble des régimes d'assurance maladie par circonscription géographique, par nature de dépense, par catégorie de professionnels responsables de ces dépenses et par professionnel ou établissement, l'évaluation des transferts entre enveloppes correspondant aux objectifs sectoriels de dépenses fixés, en fonction de l'objectif national de dépenses d'assurance maladie, dans le cadre de la loi de financement de la sécurité sociale, l'analyse quantitative des déterminants de l'offre de soins et la mesure de leurs impacts sur l'évolution des dépenses d'assurance maladie	Contribuer à la connaissance des dépenses de santé, des dépenses de l'assurance maladie et des dépenses médico-sociales, en permettant d'analyser les dépenses des régimes d'assurance maladie par circonscription géographique, par nature de dépense, par catégorie de professionnels ou de prescripteurs et par professionnel ou établissement, les dépenses d'assurance maladie au regard des objectifs sectoriels de dépenses fixés, dans le cadre de l'ONDAM, l'analyse quantitative des déterminants de l'offre de soins et la mesure de leurs impacts sur l'évolution des dépenses d'assurance maladie
Feedback aux praticiens/établissements	Contribuer à la transmission en retour aux prestataires de soins d'informations pertinentes relatives à leur activité et leurs recettes, et s'il y a lieu à leur prescription	Transmettre aux prestataires de soins les informations pertinentes relatives à leur activité, à leurs recettes et, s'il y a lieu, à leurs prescriptions	Transmettre aux prestataires de soins les informations pertinentes relatives à leur activité, à leurs recettes et, s'il y a lieu, à leurs prescriptions	Contribuer à l'information des professionnels de santé, des structures et des établissements de santé ou médico-sociaux sur leur activité, en permettant la transmission aux prestataires de soins des informations pertinentes relatives à leur activité, à leurs recettes et, s'il y a lieu, à leurs prescriptions et la mise à la disposition de leurs représentants de données ne faisant pas apparaître l'identité des professionnels de santé
Qualité des soins		Améliorer la qualité des soins, notamment par la comparaison des pratiques aux référentiels, accords de bons usages ou contrats de bonne pratique(...) et moyenne professionnelles ,l'évaluation des comportements de consommation de soins, l'analyse des caractéristiques et des déterminants de la qualité des soins	Améliorer la qualité des soins, notamment par la comparaison des pratiques aux référentiels, accords de bons usages ou contrats de bonne pratique(...) et moyenne professionnelles ,l'évaluation des comportements de consommation de soins, l'analyse des caractéristiques et des déterminants de la qualité des soins	Contribuer à l'information sur la santé ainsi que sur l'offre de soins, la prise en charge médico-sociale et leur qualité, l'orientation des usagers dans le système de santé, en permettant la comparaison des pratiques de soins, des équipements et des tarifs des établissements et des professionnels de santé
Gestion politique santé			Contribuer à une meilleure gestion des politiques de santé, notamment par l'identification des parcours de soins des patients, le suivi et l'évaluation de l'état de santé des patients et leurs conséquences sur la consommation de soins, l'analyse de la couverture sociale des patients, la surveillance de la consommation de soins en fonction de différents indicateurs de santé publique ou de risque	Contribuer à la définition, à la mise en œuvre et à l'évaluation des politiques de santé et de protection sociale, en favorisant l'identification des parcours de soins des patients, le suivi et l'évaluation de leur état de santé et de leur consommation de soins et de services d'accompagnement social, l'analyse de la couverture sociale des patients, la surveillance de la consommation de soins en fonction d'indicateurs de santé publique ou de risques sanitaires ;
Veille, sécurité sanitaire				Contribuer à la surveillance, à la veille et à la sécurité sanitaires, en développant l'observation de l'état de santé des populations, l'évaluation et la production d'indicateurs relatifs à l'état de santé de la population et l'analyse de leur variation dans le temps et dans l'espace, la détection d'événements de santé inhabituels pouvant représenter une menace pour la santé publique et l'évaluation de leurs liens éventuels avec des facteurs d'exposition et l'évaluation d'actions de santé publique ;
Recherche, épidémiologie				Contribuer à la recherche, aux études, à l'évaluation et à l'innovation dans les domaines de la santé et de la prise en charge médico-sociale

1.3.2.2 Périmètre

Le SNIIRAM est un entrepôt de données médico-administratives. Opérationnel depuis 2003, il comporte initialement uniquement les données de consommations de soins ayant donné lieu à un remboursement par le régime général de l'assurance maladie. A partir de 2009, les principaux régimes (Mutualité sociale agricole (MSA) et Régime social des indépendants (RSI)) puis les autres régimes (Caisse nationale militaire de sécurité sociale (CNMSS)) alimentent également le SNIIRAM qui devient alors réellement inter-régimes. La date de décès n'est alors disponible que pour les bénéficiaires du régime général. En 2010, le chaînage entre les données du SNIIRAM et du PMSI, mis en place de façon rétrospective et complète jusqu'aux données de 2007, permet de suivre les consommations de soins en ville et hospitalières. Depuis l'instauration du SNDS, de nouvelles sources de données ont été prévues dans la loi. Les causes de mortalité issues du Centre d'épidémiologie sur les causes médicales de décès (CépiDC INSERM) sont actuellement disponibles pour les années 2013 et 2014, et sont en cours de rattrapage pour les années précédentes et ultérieures. Les données issues des assurances maladies complémentaires et celles relatives au handicap, prévues par la loi de 2017, ne sont, à ce jour, pas disponibles.

En 2019, la loi (35) étend la durée de conservation des données du SNDS à 20 ans, de même que leur périmètre. Le SNDS couvre, aujourd'hui, les données recueillies lors des visites médicales et de dépistage obligatoire, lors des prises en charge par la protection maternelle infantile (PMI) ou lors du recours à la santé au travail, ainsi que les données des études ou des données relatives à la perte d'autonomie en cas d'appariement aux données du SNDS tel que défini avant cette loi. On distingue ainsi 2 catégories de données dans ce SNDS élargi :

- la **base principale** encore appelée « SNDS historique » (SNIIRAM, PMSI, causes de décès) qui couvre l'ensemble de la population vivant en France
- les données inscrites au **catalogue** de la Plateforme des données de Santé, restreintes à une partie de la population

1.3.2.3 Produits

Etant donné que les données inscrites au catalogue ne sont pas encore disponibles, cette section se focalise sur les produits du « **SNDS historique** ». La figure 2 présente les différents produits de restitution (datamarts) selon le niveau de granularité (données agrégées ou individuelles) et le centre d'intérêt (dépenses, offre de soins, professionnels/établissement, bénéficiaires).

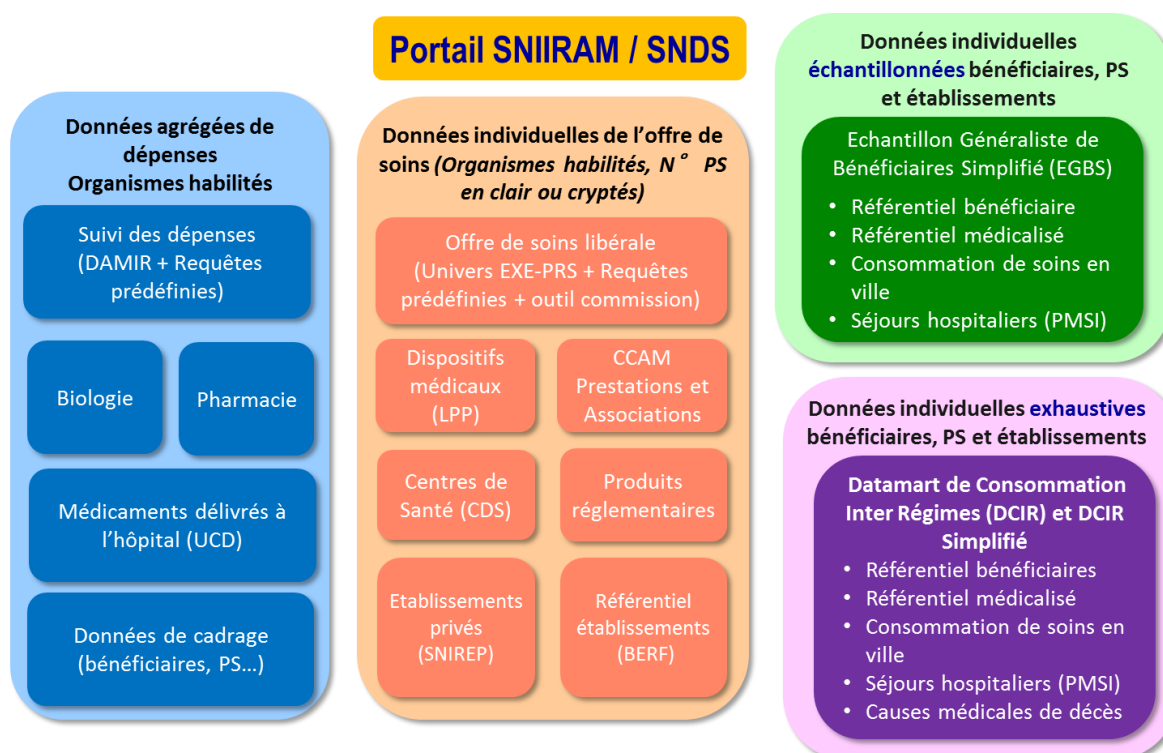


Figure 2: Produits du SNDS historique

Source : Site du SNDS

1.3.2.4 Focus sur le Datamart de Consommation Inter-Régimes (DCIR)

L'univers DCIR est organisé selon un modèle de donnée en étoile autour de la table prestation qui contient l'ensemble des remboursements de consommations de soins. Seuls les bénéficiaires ayant au moins une consommation de soins (consommants) y apparaissent. Un codage détaillé des prestations de biologie, de pharmacie, des actes médicaux (CCAM), des dispositifs médicaux et des médicaments rétrocedés est disponible (Figure 3).

La table prestations peut être jointe au référentiel des bénéficiaires qui lui-même peut être joint au PMSI, au référentiel médicalisé qui comporte les données sur les affections de longue durée (ALD) et aux causes médicales de décès. Une documentation détaillée des tables et des datamarts est

accessible sur le site de documentation de la Plateforme des Données de Santé (Health Data Hub (HDH)) (36).

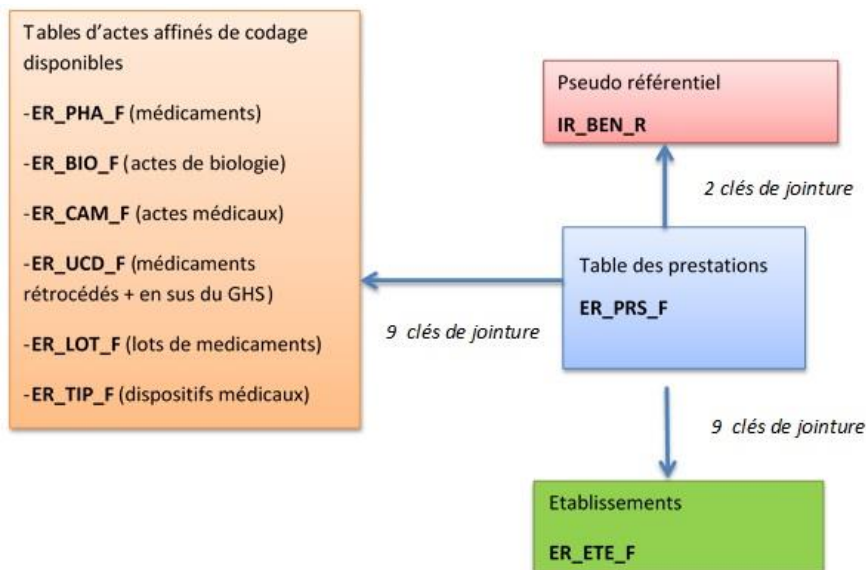


Figure 3: Tables du DCIR

Source : Site de documentation du SNDS par le Health Data Hub

1.4 Synthèse des principales données disponibles dans le SNDS

Le tableau 2 résume les principales données disponibles dans le SNDS en fonction du type de prise en charge (hospitalisation ou hors hospitalisation).

Tableau 2: Disponibilité des données dans le SNDS

	Hors hospitalisation	Hospitalisation
diagnostics	absent sauf ALD	
actes médicaux		non exhaustifs
médicaments		uniquement avec financement spécifique(en sus du GHM)
prescripteurs		pas de spécialité pour les salariés
date de décès	exhaustive pour régime général	
causes de décès	exhaustives pour régime général	exhaustives pour régime général

	présence
	partiellement présent
	présence en fonction des régimes
	présence très limitée

1.5 Réutilisation des BMA en recherche

Aux Etats-Unis , dès le début des années 90, la réutilisation de données de santé en recherche a soulevé de nombreuses questions et donné lieu aux premières recommandations pour traiter ces données (37,38). Depuis, on constate une réutilisation croissante de ces données en recherche en raison de leur exhaustivité sur un périmètre donné, d'un moindre coût d'accès et d'un suivi prolongé des sujets (39,40).

En France, à partir de 1998, l'exhaustivité et la disponibilité des données PMSI a attiré les chercheurs afin de les réutiliser en épidémiologie. Mais un certain nombre de critiques sur les données issues des bases médico-administratives (41,42) ont été décrites, parmi lesquelles : la qualité des données, l'absence de données d'intérêt, ou encore la difficulté de reconstituer l'intégralité des parcours de soins.

La qualité des données du PMSI s'est améliorée depuis la mise place de ce recueil standardisé au cours des années. Les processus qualité mis en œuvre au sein des établissements y ont contribué (43). Plusieurs études de validation par retour au dossier ont permis de montrer une qualité satisfaisante des données dans différentes pathologies (44–46), même si des améliorations peuvent encore être apportées notamment à l'aide de technique de fouille de données (47,48). Une autre critique fréquemment portée à l'encontre des bases médico-administratives concerne leur finalité première. En effet, d'une part, les données présentes peuvent être influencées par leur impact sur le financement (40). D'autre part, certaines données utiles en recherche telles que des données cliniques (poids, taille, statut tabagique, etc) ou de prise en charge peuvent être absentes (49).

Depuis 2007, les données du PMSI sont chaînées aux données du SNIIRAM au travers d'un identifiant commun. Toutefois, au-delà des limites de cet identifiant (changement d'identifiant lors d'un changement de régime, identifiant unique pour des jumeaux de même sexe, etc), le périmètre des données disponibles n'est pas similaire entre les soins en ville et à l'hôpital (Tableau 2). Par exemple, des médicaments dispensés lors d'un séjour hospitalier hors médicament avec un

financement spécifique (en sus du GHM) sont absents. La complétude des données d'un parcours de soins ville/hôpital ne peut donc être atteinte avec le SNDS comme seule source de données.

2 Objectif de thèse et axes de recherche

L'objectif principal de cette thèse était de montrer quel pouvait être l'apport des bases médico-administratives en épidémiologie et en pharmacovigilance.

Les travaux de cette thèse s'articulaient ainsi autour de deux axes de recherche :

- ✓ **Axe 1** : montrer l'apport de travaux méthodologiques pour prendre en compte les spécificités des données issues de bases médico-administratives
- ✓ **Axe 2** : illustrer, au travers de cas d'usage, l'apport de l'utilisation des données médico-administratives en épidémiologie et en pharmacovigilance

3 Axe 1 : montrer l'apport de travaux méthodologiques pour prendre en compte les spécificités des données issues de bases médico-administratives

Afin de prendre en compte les spécificités des données issues des BMA, nous proposons des travaux méthodologiques pour optimiser leur utilisation en recherche : l'identification de pathologies dans les données du SNDS, la validation d'un score pronostique, la gestion de données absentes des BMA. Deux approches ont été explorées : i) le développement d'un proxy ; ii) l'enrichissement de données de BMA par un appariement à des données issues d'un entrepôt de données hospitalières.

3.1 Adaptation et validation du score de comorbidités de Charlson aux données du SNDS : Etude 1

Contrairement aux données spécifiquement recueillies pour une étude épidémiologique classique, les données disponibles dans les BMA n'ont pas été collectées dans une finalité répondant directement aux définitions d'une exposition et/ou d'un critère de jugement. Il est donc nécessaire de traduire ces définitions en termes de données issues des BMA (code CIM10 de pathologie, acte CCAM spécifique, médicaments spécifiques...), via des algorithmes plus ou moins complexes. Ainsi l'utilisation de scores pronostiques développés en épidémiologie classique nécessite une adaptation et une validation de leurs performances avant d'être appliquée aux données de BMA.

3.1.1 Contexte

Aussi bien pour les études sur données médico-administratives que celles sur données ad hoc, la prise en compte des facteurs de confusion est indispensable pour établir une association entre une exposition et un évènement de santé en absence de randomisation. Parmi ces facteurs de confusion potentiels, les comorbidités du sujet en sont un des éléments majeurs (50). Quatre scores ont montré de bonnes performances pour évaluer cette comorbidité. Parmi eux, le score de Charlson (51), très largement utilisé dans la littérature médicale, est recommandé lorsque la mortalité est l'évènement à prédire. Il a été adapté sur plusieurs bases médico-administratives (52–55). Son utilisation sur les données du SNDS nécessite de définir au préalable chaque comorbidité du score au travers des

données disponibles dans la base puis de vérifier les performances de ce score pour prédire la mortalité. Les objectifs de ce travail étaient donc d'adapter le score de comorbidité de Charlson aux données du SNDS pour prédire la mortalité à 1 an de patients hospitalisés et de comparer la discrimination et la calibration de différentes versions de ce score.

3.1.2 Méthode

Nous avons constitué une cohorte de 6 602 641 patients sortis d'une hospitalisation en 2010 et ayant bénéficié d'au moins un remboursement de soins en 2009 par le régime général. Le décès n'étant exhaustif dans le SNDS que pour le régime général nous avons exclu les sujets ayant bénéficié d'un remboursement dans un autre régime. Cette cohorte a été divisée aléatoirement en 2 parties de taille égale pour créer un jeu de données d'apprentissage (cohorte A) et un jeu de données de validation (cohorte B). Un suivi de 12 mois après la sortie d'hospitalisation d'un établissement français de court séjour MCO a permis de recueillir le critère de jugement qui était le décès toute cause.

Quatre sources de données ont été utilisées pour définir les comorbidités dans les 12 mois précédant la sortie d'hospitalisation : les affections longue durée (code CIM10), les diagnostics d'hospitalisation (code CIM10), les actes médicaux (code CCAM), les délivrances médicamenteuses (code identifiant de présentation (CIP)). Les codes CIM 10 de l'algorithme de Quan (54) ont été adaptés aux données du PMSI français.

Un modèle de Cox a ensuite été construit afin de déterminer l'association de chaque pathologie avec le décès et de dériver des poids de la même manière que Charlson dans le score princeps.

Quatre versions du score ont été calculées pour chaque sujet : le score princeps (51), le score ajusté sur l'âge proposé par Charlson (56), le score avec des poids proposés par Quan (57) et le score obtenu avec les poids déterminés sur notre cohorte A d'apprentissage. La discrimination (c-statistique (58) pour les modèles de survie) et la calibration (courbe) de chaque version du score en variable quantitative et catégorielle ont été étudiées.

3.2 Résultats

Parmi les 6 602 641 patients inclus, 34,1 % présentaient au moins une comorbidité identifiée dans le score de Charlson. Les plus fréquentes étaient les cancers, les pathologies broncho-pulmonaires chroniques, et le diabète. La survie à 1 an était de 94,9 %, variant de 98,4 % pour les sujets avec un Charlson à 0 jusqu'à 71,6 % pour ceux avec un score supérieur ou égal à 5. Le modèle de Cox a déterminé de nouveaux poids sur la survie pour plusieurs pathologies. Certaines ont vu leur impact diminuer comme l'infarctus myocardique, les connectivites, la pathologie ulcéreuse, le diabète (poids ramené à 0 au lieu de 1 sauf pour le diabète compliqué de type 2), le SIDA/VIH (poids ramené à 1 au lieu de 6) alors que d'autres ont vu leur poids augmenter comme l'insuffisance cardiaque, les démences, les pathologies hépatiques modérées (poids de 2 au lieu de 1), les pathologies métastatiques (poids de 11 au lieu de 6). La version du score de Charlson qui présentait les meilleures performances était celle du score ajusté de l'âge en 4 classes (0, 1-2, 3-4 et ≥ 5) avec une c-statistique de 0,91 et une bonne calibration.

The Best Use of the Charlson Comorbidity Index With Electronic Health Care Database to Predict Mortality

Aurélie Bannay, MD,* Christophe Chaignot, MSc, † Pierre-Olivier Blotière, MSc, †
Mickaël Basson, MD, MSc, † Alain Weill, MD, † Philippe Ricordeau, MD, †
and François Alla, MD, PhD †

Background: The most used score to measure comorbidity is the Charlson index. Its application to a health care administrative database including International Classification of Diseases, 10th edition (ICD-10) codes, medical procedures, and medication required studying its properties on survival. Our objectives were to adapt the Charlson comorbidity index to the French National Health Insurance database to predict 1-year mortality of discharged patients and to compare discrimination and calibration of different versions of the Charlson index.

Methods: Our cohort included all adults discharged from a hospital stay in France in 2010 registered in the French National Health Insurance general scheme. The pathologies of the Charlson index were identified through ICD-10 codes of discharge diagnoses and long-term disease, specific medical procedures, and reimbursement of specific medications in the past 12 months before inclusion.

Results: We included 6,602,641 subjects at the date of their first discharge from medical, surgical, or obstetrical department in 2010. One-year survival was 94.88%, decreasing from 98.41% for Charlson index of 0–71.64% for Charlson index of ≥ 5 . With a discrimination of 0.91 and an appropriate calibration curve, we retained the crude Cox model including the age-adjusted Charlson index as a 4-level score.

Conclusions: Our study is the first to adapt the Charlson index to a large health care database including >6 million of inpatients. When mortality is the outcome, we recommended using the age-adjusted Charlson index as 4-level score to take into account comorbidities.

Key Words: comorbidity, administrative database, prognosis, mortality (*Med Care* 2016;54: 188–194)

From the *Faculté de médecine de Nancy, Université de Lorraine, Vandoeuvre les Nancy; and †CNAMTS, Paris, France.

Presented at 3 national French congresses: (i) 2èmes Journées scientifiques de l'assurance maladie, October 9–10, 2012, Paris, France; (ii) XXVIème Congrès national EMOIS, March 21–22, 2013, Nancy, France; (iii) 8ème Conférence Francophone d'Epidémiologie Clinique, May 14–16, 2014, Bordeaux, France.

The authors declare no conflict of interest.

Reprints: Aurélie Bannay, MD, Faculté de médecine de Nancy, Université de Lorraine, 9 avenue de la forêt de Haye, Vandoeuvre les Nancy 54500, France. E-mail: aurelie.bannay@univ-lorraine.fr.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medical-care.com.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 0025-7079/16/5402-0188

188 | www.lww-medicalcare.com

Health care administrative databases are increasingly used to conduct epidemiological studies, especially for postapproval drug efficacy and/or safety assessments. Regardless of the quality or quantity of data available, these studies are subject to the same limitations as classic observational studies. One of the main weaknesses of observational studies is the lack of control for confounders. Treatment arms cannot be directly compared in the absence of randomization, as various factors can influence the physician's choice of therapy, and the patient's medical history is one of the most important factors.¹ Comorbid conditions are frequently linked to clinically relevant outcomes, which can lead to an indication bias. Controlling for this major confounder requires measurement of each patient's global medical history. Several methods designed to assess comorbidity have been described in the literature. A review conducted by De Groot et al¹ screened 13 different methods and identified 4 indices that provided sufficient data to validate their clinimetric properties. Among these indices, the Charlson index² was recommended to measure comorbidity when mortality is the outcome of interest. The Charlson comorbidity index is the most extensively studied and most widely used comorbidity index in the medical literature (10,398 citations in articles or reviews on the Scopus database on April 17, 2014). The widespread use of this index could be explained by the fact that it is not designed for patients with a specific disease. It has also been adapted to administrative databases,^{3–9} frequently using a discharge database without medication to predict in-hospital mortality or 6-month mortality. The French health insurance administrative database, including national health insurance claims and a discharge database, is one of the largest health databases covering about 96% of the French population and also recording ambulatory care. To use the Charlson index on this French administrative database, its predictive properties for survival had to be studied.

The objectives of this study were to: (i) adapt the Charlson comorbidity index to the French National Health Insurance database to predict 1-year mortality of discharged patients, and (ii) compare discrimination and calibration of different versions of the Charlson index.

METHODS

Study Design and Data Collection Procedures

A population-based prognostic study was conducted on the French National Health Insurance Information System

Medical Care • Volume 54, Number 2, February 2016

Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved.

[Système national d'information Inter-Régimes de l'Assurance Maladie (SNIIRAM)],¹⁰ a claim database covering about 96% of the inhabitants of France. Several schemes provide mandatory benefits for all inhabitants of France. The general scheme covers approximately 75% of the population (including all French salaried workers and their dependants, as well as unemployed and retired salaried workers). Civil servants, nonworking students, rural workers, and self-employed benefit from other health insurance schemes. The SNIIRAM database contains individual, anonymous, and comprehensive data on all health expenditure reimbursements for patients¹¹ as well as demographic data (sex, age, and vital status). Patients with certain severe and costly chronic diseases [long-term disease (LTD)] can receive additional benefits. LTD is composed of a list of 30 chronic diseases categories (such as diabetes, multiple sclerosis, heart failure...) allowing a 100% reimbursement rate for related treatment. In this case, the LTD diagnosis is recorded according to the International Classification of Diseases, 10th edition (ICD-10). The SNIIRAM can be linked to the French Hospital Discharge database [Programme de Médicalisation des Systèmes d'Information (PMSI)]. In particular, the PMSI contains discharge diagnoses (with ICD-10 codes) and medical procedures performed during each hospital stay. Implementation of the SNIIRAM database was approved by the French data protection agency (Commission Nationale de l'Informatique et des Libertés).

Study Cohort

The study cohort included all adults discharged from surgical or obstetric wards in France between January 1, 2010 and December 31, 2010 registered in the French National Health Insurance general scheme. When patients had multiple hospital stays fulfilling inclusion criteria, only the first was considered. To ensure that patients had been covered by the general scheme for at least 1 year before inclusion in the study, patients were required to have received at least 1 reimbursement from the general scheme between January 1, 2009 and December 31, 2009. Participants were excluded if they had received a reimbursement from another scheme between January 1, 2010 and December 31, 2011. The cohort was randomly divided into 2 populations with a 50:50 probability; a test population called population A and a validation population called population B were distinguished.

Measures

Comorbidities

Comorbidity was assessed using the Charlson comorbidity index.² Comorbid diseases were defined from the following 4 data sources during the 12 months before inclusion in the study: (i) in-hospital diagnoses; (ii) LTD diagnoses; (iii) specific medical procedures performed during a hospital stay; and (iv) specific medications. The presence of comorbidity in at least 1 source was sufficient to attribute the comorbidity to the patient. The algorithms developed by Quan et al⁷ were adapted to the first 2 data sources (in-hospital diagnoses and LTD diagnoses). These algorithms have been shown to be the most highly discriminant of 3

different approaches to predict in-hospital mortality.¹² In addition to these ICD-10 codes, comorbidities were also defined by the presence of specific medical procedures: vascular stenting for peripheral vascular disease, dialysis for moderate or severe renal disease, laser surgery for diabetic retinopathy for diabetes with end-organ damage, or by the consumption of specific medication: antidiabetic drugs for diabetes, bronchodilator drugs for chronic pulmonary disease, and anti-Alzheimer drugs for dementia. Moreover, the definition of diabetes with end-organ damage was completed by the identification of diabetes plus at least one of the following conditions: myocardial infarction, moderate or severe renal disease, or cerebrovascular disease. This kind of algorithm has been previously used to define chronic conditions such as cancer from data of the SNIIRAM and PMSI databases.¹³ Table 1 presents details of the algorithm used for each comorbidity according to the 4 available sources.

Scores

We compared 4 versions of the Charlson index: (i) the initial Charlson index; (ii) the age-adjusted Charlson index; (iii) the Charlson index modified by Quan¹⁴; and (iv) the updated Charlson index with weights obtained on our database.

The Charlson index for a given participant corresponds to the sum of the weights (Table 4) of the subject's current comorbidities. All comorbid diseases were cumulative except for diabetes and diabetes with end-organ damage, and any tumor (including leukemia and lymphoma) and metastatic solid tumor. To take age into account, Charlson reevaluated the index and proposed an age-adjusted version¹⁵ by adding 1 point to the total score for each decade after the age of 50 years. In 2011, Quan et al¹⁴ provided new weights of the Charlson index obtained from administrative data in 6 countries. In the same way of the initial Charlson index, the sum of weights of disease affecting a given subject corresponds to the "Charlson index modified by Quan." The fourth score was obtained by the sum of the weights estimated from our database. They were obtained by a Cox model used to assess the crude or age-adjusted and sex-adjusted hazard ratio (HR) of each comorbid disease on 1-year mortality. It was hereafter referred to as the "updated Charlson index."

In our study, these 4 scores were used either as a total score or divided into 4 levels: 0 point, 1 or 2 points, 3 or 4 point, and ≥ 5 points.

Outcome and Follow-up Duration

Inclusion date in the study was the date of first discharge of the participant from a nonpsychiatric acute hospital in 2010. Outcome was all-cause mortality from the day of discharge. Participants were followed until the date of death or during 12 months; no loss to follow-up was observed due to the comprehensive nature of the database. Participants who died during their hospital stay had a follow-up duration equal to 0 day.

TABLE 1. Definition of Comorbidities According to Available Sources

	ICD-10 Codes (In-Hospital* or LTD Diagnoses)	Medical Procedures*	Medications*
Myocardial infarction	I21;I22;I252;I255		
Congestive heart failure	I110;I130;I132;I50		
Peripheral vascular disease	I70;I71;I731;I738;I739;I771;I790;I792;K551; K558;K559; Z958;Z959	Peripheral vascular stenting	
Cerebrovascular disease	G45;G46;H340;I60-I69		
Dementia	F00-F03;F051;G30;G311		At least 3 reimbursements of anti-Alzheimer drugs
Chronic pulmonary disease	I278;I279;J40-J47;J60-J67;J684;J701;J703		At least 2 reimbursements of bronchodilator drugs
Connective tissue disease	M05;M06;M315;M32;M33;M34;M351;M353;M360		
Ulcer disease	K25-K28		
Mild liver disease	B18;K700-K703;K709;K713K715;K717;K73; K74;K760; K762-K764;K768;K769;Z944		
Diabetes	E100;E101;E106;E108;E109;E110;E111;E116; E118;E119; E120;E121; E126;E128-E131;E136; E138-E141; E146;E148; E149		At least 3 reimbursements (or at least 2 in case of large pack sizes) of oral antidiabetic agents and/or insulin
Hemiplegia	G041;G114;G801;G802;G81;G82; G830; G831- G834;G839		
Moderate or severe renal disease	I120;I131;N032-N037;N052-N057;N18;N19; N250;Z490;Z491; Z492;Z940;Z992	Dialysis	
Diabetes with end-organ damage [†]	E102-E105;E107;E112-E115;E117; E122-E125; E127;E132-E135;E137; E142-E145;E147	Laser surgery for diabetic retinopathy	
Any tumor (including lymphoma and leukemia except for malignant neoplasm of skin)	C00-C26;C30-C34;C37-C41;C43; C45-C58; C60- C76;C81-C85;C88; C90-C97		
Moderate or severe liver disease	I850;I859;I864;I982;K704;K711; K721;K729; K765- K767		
Metastatic solid tumor	C77-C80		
HIV-AIDS	B20-B22;B24;Z21		

*Recorded during the 12 months before inclusion.

[†]The definition of diabetes with end-organ damage was completed by identification of diabetes plus at least one of the following conditions: myocardial infarction, moderate or severe renal disease, or cerebrovascular disease.

ICD-10 indicates International Classification of Diseases, 10th edition.

Statistical Analysis

For each comorbidity, the proportion of each source was determined from the number of subjects presenting a given comorbidity identified by one source over the total number of subjects presenting this comorbidity identified by at least 1 source. One-year survival was assessed by Kaplan-Meier curves. In-hospital mortality was obtained directly by the percentage of participants who had died at discharge. Eight Cox models were built, each one included one of the 4 scores as an independent variable alone or with adjustment for age (as a categorical variable) and sex to predict 1-year mortality. The performance of each model was assessed in terms of discrimination and calibration. Discrimination was measured using Harrell et al's¹⁶ *c*-statistic calculated by the SAS macro SurvCsTD¹⁷ according to the method of Kremers.¹⁸ Because of the size of the cohort, *c*-statistics were calculated over 1000 bootstrap samples of 10,000 subjects. The 95% confidence interval (CI) limits were the 2.5th and 97.5th percentiles obtained with 1000 bootstrap samples.

Calibration was examined by calibration curves plotting the predicted probability of survival versus the observed survival for each level of Charlson index. Because of the large size of our cohort, the assumption of proportional

hazards was assessed graphically by plotting $\log(-\log(\text{survival}))$ versus $\log(\text{time})$.

Analyses performed on our test cohort (population A) were repeated on our validation cohort (population B). All analyses involved the use of SAS statistical software, version 9.2 (SAS Institute, Cary, NC). A *P*-value of <0.05 was considered statistically significant.

RESULTS

A total of 6,602,641 subjects were included with a mean number of comorbid diseases of 0.52 per subject and a mean of 1.45 comorbid disease per subject among those subjects with at least 1 comorbid disease.

The distribution of the Charlson index depicted in Figure (Supplement Digital Content 1, <http://links.lww.com/MLR/B71>) showed that about 64% of subjects did not suffer from any comorbid disease. The main reasons for hospitalization of these patients were childbirth, elective abortion, gastrointestinal endoscopy, and cataract surgery. Their hospital stay was not related to chronic condition screened by the Charlson index. A total of 29.6% of patients with at least 1 comorbidity were identified by the PMSI, 63.4% by the

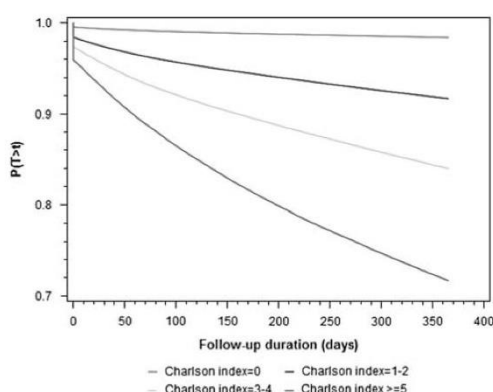


FIGURE 1. One-year survival according to the levels of the Charlson index.

LTD, and 40.6% by specific medications or procedures when appropriate. Details for each comorbid disease are available in Table (Supplement Digital Content 2, <http://links.lww.com/MLR/B72>). The most common comorbidities were cancer and chronic pulmonary disease presented by 9.9% and 9.7% of subjects, respectively. In contrast, ulcer disease, HIV/AIDS, and moderate and severe liver disease were reported in only 0.2%, 0.5%, and 0.5% of the cohort, respectively. Associations of at least 2 comorbidities were observed in 11% of subjects and the most common associations were diabetes with end-organ damage/chronic pulmonary disease, cancer/chronic pulmonary disease, and cancer/diabetes with end-organ damage. The test cohort (population A: n=3,299,213) and the validation cohort (population B: n=3,303,428 subjects) were balanced in

terms of age, sex, and comorbidities, as shown in Table (Supplement Digital Content 3, <http://links.lww.com/MLR/B73>).

One-year survival was 94.88% (338,271 deaths) (Figure, Supplement Digital Content 4, <http://links.lww.com/MLR/B74>), in-hospital survival was 99.03% (64,163 deaths), and postdischarge survival was 95.81% (274,108 deaths). Survival curves (Fig. 1) showed decreasing survival with increasing Charlson score from 98.41% for a Charlson index of 0–71.64% for a Charlson index of ≥ 5 .

A Cox model was used to estimate HRs for each comorbidity in the 2 subcohorts to compare the weights obtained in our study with those initially determined by Charlson et al in 1987.² After adjustment for age and sex, the HR of each comorbidity was closer to 1 except for 2 comorbidities with an initial weight of 6: (i) metastatic solid tumor with an HR of 11.36 (95% CI, 11.18–11.55); and (ii) HIV/AIDS with an HR of 1.43 (95% CI, 1.32–1.55) (Table 2). From the HR, we derived new weights to calculate the updated Charlson index according to the same HR ranges as used by Charlson.

The 4 version of Charlson index scores were significantly associated with 1-year mortality (Table 3). HR associated with the total score ranged between 1.477 and 1.525 in crude models and 1.364 and 1.409 in age-adjusted and sex-adjusted models.

The *c*-statistics of each model are presented in Table 4. A higher discrimination was observed for models including the Charlson index as a 4-level score. The model providing the best discrimination was for the crude Cox model including the age-adjusted Charlson index as a 4-level score. The calibration curves highlighted the relevance of the model including a 4-level Charlson index (Figure, Supplement Digital Content 5, <http://links.lww.com/MLR/B75>). According to these performances, the model selected was the crude model including the age-adjusted Charlson index as a 4-level score offering a discrimination of 0.91439 and the

TABLE 2. Comparison of Weights of Each Comorbidity on 1-Year Mortality With Charlson's Initial Weights

	Original Charlson Index		Population A				New Weight
	Initial Weight	Initial Range of HR	Model Adjusted for Sex and Age			P	
			HR	95% CI	P		
Myocardial infarct	1	1.2–1.5	1.042	1.025	1.059	<0.0001	0
Congestive heart failure	1	1.2–1.5	1.735	1.706	1.764	<0.0001	2
Peripheral vascular disease	1	1.2–1.5	1.313	1.290	1.336	<0.0001	1
Cerebrovascular disease	1	1.2–1.5	1.232	1.207	1.258	<0.0001	1
Dementia	1	1.2–1.5	1.874	1.846	1.903	<0.0001	2
Chronic pulmonary disease	1	1.2–1.5	1.306	1.290	1.323	<0.0001	1
Connective tissue disease	1	1.2–1.5	1.099	1.061	1.138	<0.0001	0
Ulcer disease	1	1.2–1.5	1.151	1.086	1.221	<0.0001	0
Mild liver disease	1	1.2–1.5	1.889	1.827	1.953	<0.0001	2
Diabetes	1	1.2–1.5	1.185	1.167	1.202	<0.0001	0
Diabetes with end-organ damage	2	1.5–2.5	1.155	1.132	1.179	<0.0001	0
Moderate or severe renal disease	2	1.5–2.5	1.401	1.371	1.432	<0.0001	1
Hemiplegia	2	1.5–2.5	1.869	1.812	1.928	<0.0001	2
Any tumor (including leukemia and lymphoma)	2	1.5–2.5	2.286	2.260	2.311	<0.0001	2
Moderate or severe liver disease	3	2.5–3.5	2.476	2.369	2.587	<0.0001	3
Metastatic solid tumor	6	6.0 to +∞	11.363	11.177	11.551	<0.0001	11
HIV/AIDS	6	6.0 to +∞	1.432	1.320	1.554	<0.0001	1

CI indicates confidence interval; HR, hazard ratio.

TABLE 3. One-Year Mortality Models

	Population A		Population B		Population A		Population B		
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	
Charlson index					Charlson index modified by Quan				
Crude models including									
Continuous score	1.477	1.475–1.480	1.477	1.475–1.480	1.524	1.522–1.527	1.525	1.523–1.528	
Categorical score									
0	1		1		1		1		
1–2	5.411	5.341–5.481	5.367	5.298–5.437	4.751	4.697–4.806	4.731	4.676–4.785	
3–4	10.813	10.649–10.980	10.803	10.639–10.970	9.597	9.453–9.744	9.706	9.651–9.854	
≥ 5	20.364	20.047–20.685	20.398	20.083–20.719	19.613	19.322–19.908	19.633	19.342–19.928	
Models adjusted for age and sex including									
Continuous score	1.365	1.363–1.368	1.366	1.363–1.368	1.409	1.407–1.412	1.409	1.406–1.411	
Categorical score									
0	1		1		1		1		
1–2	2.457	2.424–2.490	2.446	2.414–2.479	2.420	2.391–2.449	2.412	2.384–2.441	
3–4	3.885	3.823–3.947	3.890	3.828–3.952	3.739	3.681–3.797	3.776	3.717–3.835	
≥ 5	9.634	9.479–9.790	9.644	9.491–9.801	10.475	10.318–10.635	10.517	10.359–10.677	
Age-adjusted Charlson index*					Updated Charlson index				
Crude models including									
Continuous score	1.517	1.515–1.519	1.503	1.501–1.505	1.481	1.479–1.483	1.480	1.478–1.483	
Categorical score									
0	1		1		1		1		
1–2	4.710	4.537–4.890	4.758	4.583–4.940	5.355	5.287–5.424	5.306	5.239–5.375	
3–4	20.294	19.612–21.000	20.430	19.741–21.144	10.666	10.503–10.830	10.636	10.475–10.800	
≥ 5	77.265	74.735–79.880	78.095	75.528–80.750	20.249	19.933–20.570	20.290	19.975–20.610	
Models adjusted for age and sex including									
Continuous score	1.381	1.378–1.383	1.372	1.370–1.375	1.364	1.361–1.367	1.364	1.362–1.367	
Categorical score									
0	1		1		1		1		
1–2	2.981	2.851–3.117	3.077	2.942–3.218	2.435	2.403–2.467	2.422	2.390–2.454	
3–4	11.742	11.231–12.276	12.158	11.629–12.712	3.798	3.737–3.859	3.797	3.737–3.858	
≥ 5	34.648	33.158–36.206	36.087	34.532–37.712	9.481	9.328–9.635	9.484	9.332–9.638	

*One point was added to the total score of the Charlson index for each decade after the age of 50 years.
CI indicates confidence interval; HR, hazard ratio.

calibration curve presented in Figure (Supplement Digital Content 5, <http://links.lww.com/MLR/B75>).

DISCUSSION

This is the first study adapting the Charlson index to a large health care database including >6 million patients discharged from hospital to predict 1-year mortality. We used the same methodological approaches as those used by Charlson to select our adult population without any restriction in terms of age or reason for hospitalization. No patients were lost to follow-up, as these databases include all hospital stays and all reimbursements provided by the French National Health Insurance general scheme. We used a Cox model, as used by Charlson, and time-to-death was considered to be the most relevant event even if follow-up duration was limited to 1 year.

Although the Charlson index was initially developed from medical records, several studies have adapted this index to administrative databases by translating each comorbid disease into ICD codes. We selected the version proposed by Quan et al⁷ because of its discriminant properties. To limit underestimation of diseases not requiring hospitalization inherent to administrative database studies, we completed these algorithms with medical procedure and medication data available in our database. Comparison of the prevalence

of comorbidities according to 3 versions of ICD-10 coding algorithms¹² demonstrated similar results, especially with Canadian and Swiss populations. The algorithms of these studies were compared with data obtained from medical charts. If even difficulties to define individual comorbidities were highlighted,¹⁹ the mortality prediction was not affected by them.²⁰ Our algorithms could not be directly validated, as no matching of our database with medical records was possible. However, mortality prediction did not appear to be affected.

Our cohort showed proportions of comorbidities similar to the study conducted by Quan et al,⁷ especially regarding patients with no comorbidities. Indeed 63.8% of patients in our database had no comorbid disease; in their study Quan and colleagues found that between 63.8% and 65.5% of patient had a Charlson index equal to 0.

Two significant differences were observed concerning the weights determined in our study after adjustment for age and sex: the first concerned the effect of HIV/AIDS on mortality, which was decreased to 1.4, probably due to progress in therapy since 1987, and the second concerned metastatic solid tumor, which was increased to 11.3, also possibly due to progress in diagnosis and therapy. Improved cancer staging could result in more patients classified with metastatic cancer. This classification bias could decrease the effect of cancer only and increase the effect of metastatic

TABLE 4. Discrimination of Each 1-Year Mortality Cox Model

	Population A		Population B	
	c-Statistic	95% CI	c-Statistic	95% CI
Charlson index				
Crude model				
Continuous score	0.84147	0.84094–0.84201	0.84191	0.84137–0.84246
Categorical score	0.86011	0.85955–0.86067	0.86026	0.85968–0.86083
Model adjusted for age and sex				
Continuous score	0.86623	0.86578–0.86669	0.86535	0.86489–0.86582
Categorical score	0.86870	0.86825–0.86915	0.86788	0.86742–0.86834
Age-adjusted Charlson index				
Crude model				
Continuous score	0.88307	0.88264–0.88349	0.88266	0.88223–0.88310
Categorical score	0.91439	0.91393–0.91486	0.91405	0.91357–0.91453
Model adjusted for age and sex				
Continuous score	0.86819	0.86774–0.86864	0.86771	0.86724–0.86817
Categorical score	0.86942	0.86898–0.86986	0.86891	0.86845–0.86937
Charlson index modified by Quan				
Crude models				
Model including continuous score	0.85163	0.85107–0.85220	0.84636	0.84590–0.84682
Model including categorical score	0.86083	0.86034–0.86152	0.86122	0.86063–0.86161
Models adjusted for age and sex				
Model including continuous score	0.87074	0.87028–0.87120	0.87023	0.86976–0.87069
Model including categorical score	0.87255	0.87209–0.87300	0.87269	0.87224–0.87315
Charlson index updated version				
Crude models				
Model including continuous score	0.84046	0.83990–0.84102	0.84082	0.84027–0.84138
Model including categorical score	0.85913	0.85855–0.85972	0.85912	0.85854–0.85970
Models adjusted for age and sex				
Model including continuous score	0.86430	0.86384–0.86477	0.86410	0.86363–0.86458
Model including categorical score	0.86717	0.86672–0.86762	0.86684	0.86638–0.86730

CI indicates confidence interval.

cancer on mortality in our model. At the same time, progress in therapy, especially clinical research, has allowed more aggressive treatments and higher hospitalization rates.

The weights observed for other comorbidities were similar to those reported by several studies proposing a modified Charlson index²¹ by reducing the number of significant comorbid diseases. For example, myocardial infarction was no longer associated with 1-year mortality after adjustment for age and sex.^{14,21}

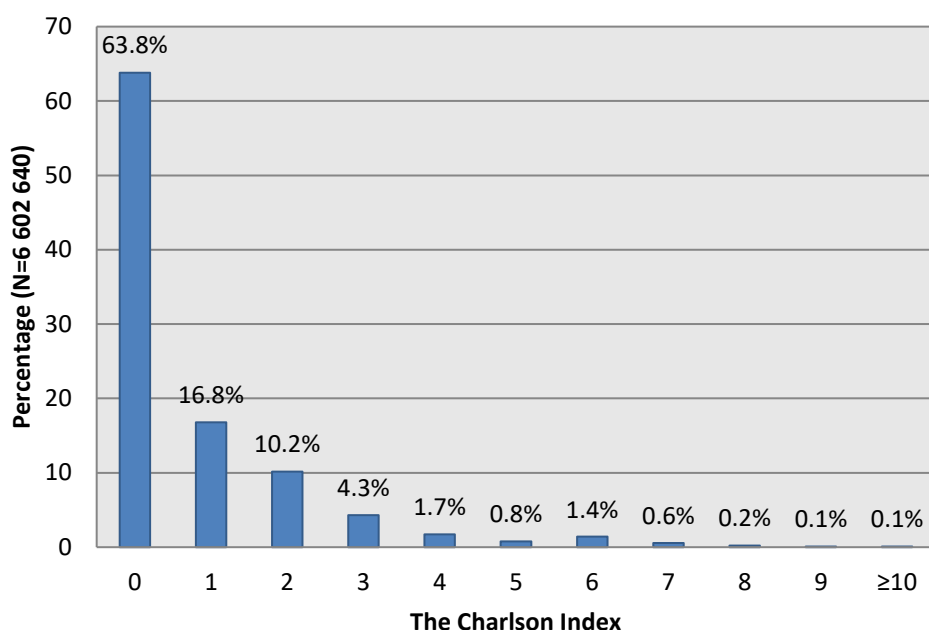
Discrimination of our models was slightly better than that obtained by previous adaptations of the Charlson index to administrative databases,^{5,12,20,22} as we obtained a concordance of >0.9 with our model, whereas other studies obtained lower concordance, even after adding age and clinical scores. Calibration curves showed discrepancies between predicted and observed survival especially for the highest Charlson scores with lowest predicted survival probabilities, possibly suggesting a nonlinear effect of comorbid disease beyond a cut-off of 5, as shown by D’Hoore et al.⁵ The use of the age-adjusted Charlson index allowed age to be taken into account, as age, along with sex, remains the best predictor of survival.¹ The choice of using the Charlson index or the age-adjusted Charlson index depends on the objective of the conducted study. It can be more relevant to use the Charlson index and to adjust for age in the main model. However, in the light of these findings, we recommend the use of the age-adjusted Charlson index as a 4-level scale.

This study showed that the Charlson index can be adapted to the French National Health Insurance database matched to the French hospital discharge database. The Charlson index was significantly associated with 1-year survival. When mortality is the outcome of interest, we recommended using the age-adjusted Charlson index as a 4-level score to take comorbidities into account. Although several studies have provided modified weights for the Charlson index, further studies are required to develop a score for ambulatory patients.

REFERENCES

1. De Groot V, Beckerman H, Lankhorst GJ, et al. How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol.* 2003;56:221–229.
2. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40:373–383.
3. Chae J-W, Song CS, Kim H, et al. Prediction of mortality in patients undergoing maintenance hemodialysis by Charlson Comorbidity Index using ICD-10 database. *Nephron Clin Pract.* 2011;117:c379–c384.
4. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol.* 1992;45:613–619.
5. D’Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *J Clin Epidemiol.* 1996;49:1429–1433.
6. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol.* 1993;46:1075–1079, discussion 1081–1090.

7. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–1139.
8. Halfon P, Egli Y, van Melle G, et al. Measuring potentially avoidable hospital readmissions. *J Clin Epidemiol*. 2002;55:573–587.
9. Sundararajan V, Henderson T, Perry C, et al. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol*. 2004;57:1288–1294.
10. Moulis G, Lapeyre-Mestre M, Palmaro A, et al. French health insurance databases: what interest for medical research? *Rev Médecine Interne*. 2015;36:411–417.
11. Tuppin P, de Roquefeuil L, Weill A, et al. French national health insurance information system and the permanent beneficiaries sample. *Rev Épidémiologie Santé Publique*. 2010;58:286–290.
12. Sundararajan V, Quan H, Halfon P, et al. Cross-national comparative performance of three versions of the ICD-10 Charlson index. *Med Care*. 2007;45:1210–1215.
13. Kudjawu Y, de Maria F, Decool E, et al. Crossing two health administrative databases: methodology and descriptive study for application to epidemiological surveillance of cancer in France. *Bull Epidemiol Hebd*. 2013; Spec No:49–58.
14. Quan H, Li B, Couris CM, et al. Updating and validating the Charlson Comorbidity Index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011;173:676–682.
15. Charlson M, Szatrowski TP, Peterson J, et al. Validation of a combined comorbidity index. *J Clin Epidemiol*. 1994;47:1245–1251.
16. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.
17. Kremers W. Calculates the C-statistic (concordance, discrimination index) for survival data with time dependent covariates and corresponding SE and 100(1-alpha)% CI. Available at: <http://www.mayo.edu/research/departments-divisions/departement-health-sciences-research/division-biomedical-statistics-informatics/software/locally-written-sas-macros>. Accessed February 10, 2014.
18. Kremers W. Concordance for survival time data: fixed and time-dependent covariates and possible ties in predictor and time. Mayo Foundation. 2007. Available at: <http://www.mayo.edu/research/documents/biostat-80pdf/DOC-10027891>. Accessed February 10, 2014.
19. Kieszak SM, Flanders WD, Kosinski AS, et al. A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *J Clin Epidemiol*. 1999;52:137–142.
20. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol*. 2000;29:891–898.
21. Ghali WA, Hall RE, Rosen AK, et al. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. *J Clin Epidemiol*. 1996;49:273–278.
22. Kim KH. Comparative study on three algorithms of the ICD-10 Charlson comorbidity index with myocardial infarction patients. *J Prev Med Public Health Yebang Uihakhoe Chi*. 2010;43:42–49.



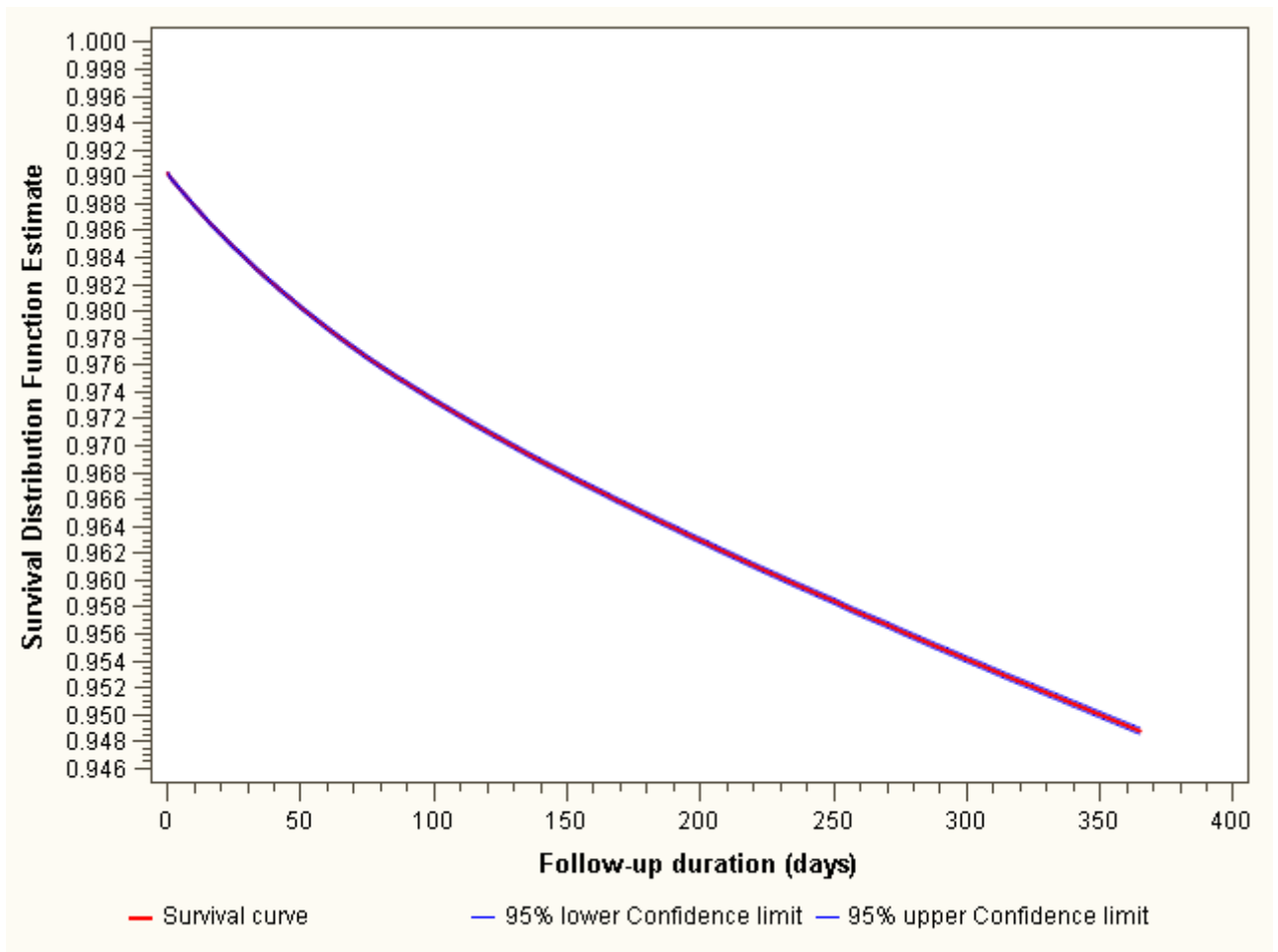
Supplement digital content 1. The Charlson Index distribution

Supplement digital content 2. Definition of comorbidities according to available sources and **proportion identified by each source**

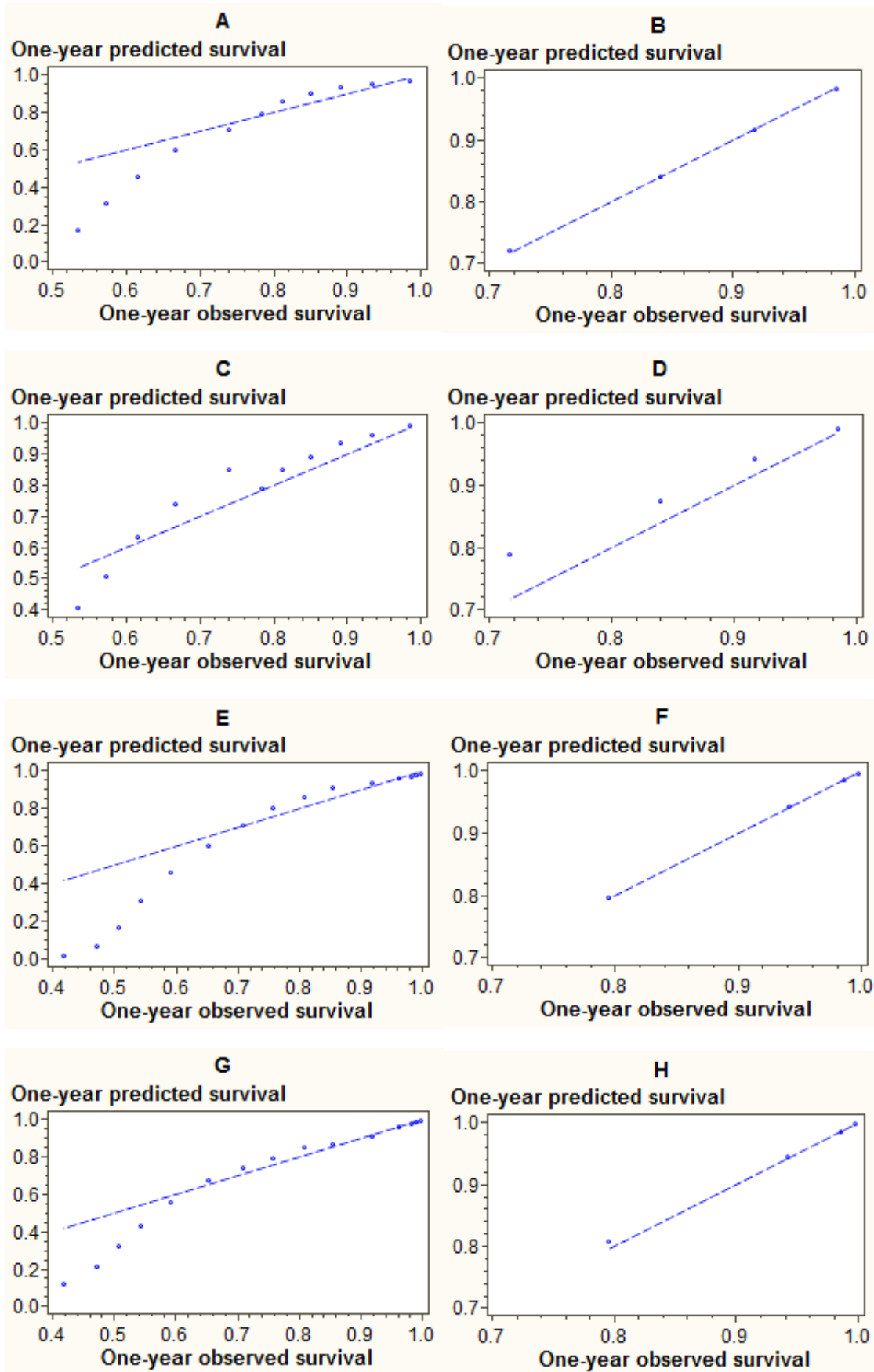
	Sensitivity of each source (%)		
	In-hospital diagnoses	LTD diagnoses	Medical procedures/ medication
Myocardial infarction	21.9	87.9	-
Congestive heart failure	55.1	52.2	-
Peripheral vascular disease	31.2	69.1	15.7
Cerebrovascular disease	42.1	66.7	-
Dementia	24.6	50.8	89.0
Chronic pulmonary disease	14.9	10.1	93.8
Connective tissue disease	24.3	92.8	-
Ulcer disease	100.0	-	-
Mild liver disease	45.4	72.8	-
Diabetes	19.9	76.9	82.3
Diabetes with end-organ damage	-	-	-
Moderate or severe renal disease	76.0	45.9	22.5
Hemiplegia	37.2	72.3	-
Any tumor (including leukemia and lymphoma)	22.1	95.6	-
Moderate or severe liver disease	45.0	67.8	-
Metastatic solid tumor	94.4	7.9	-
HIV/AIDS	38.0	94.1	-

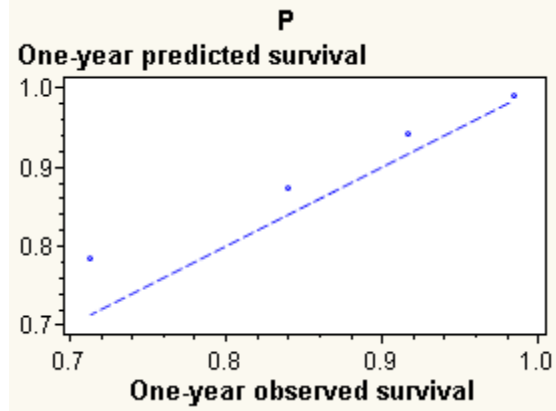
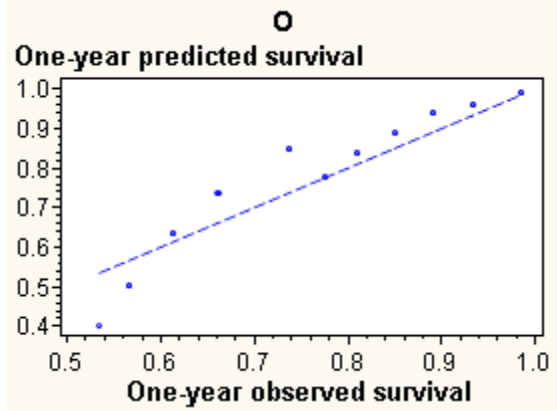
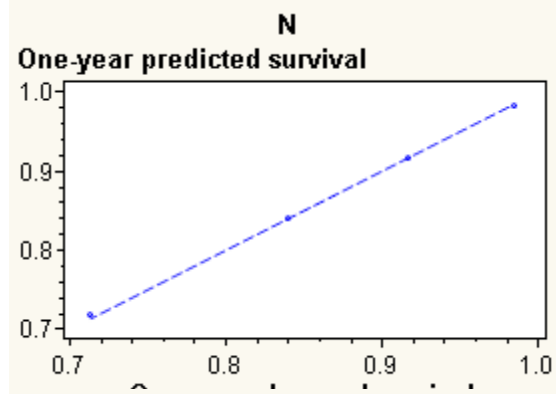
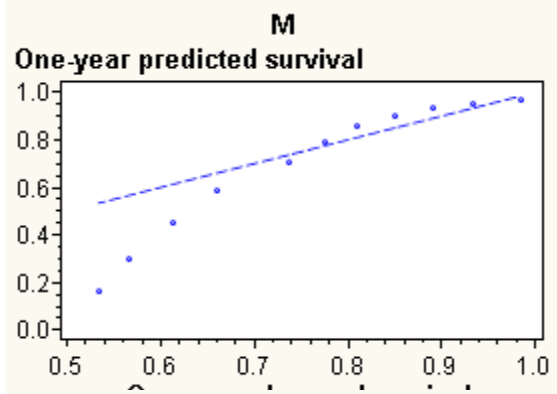
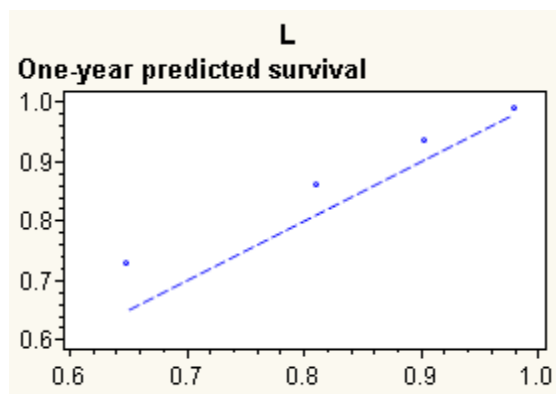
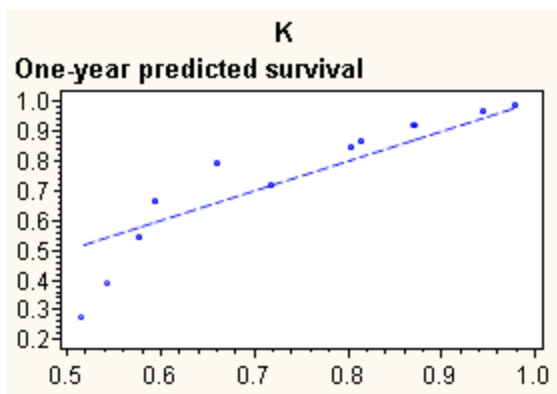
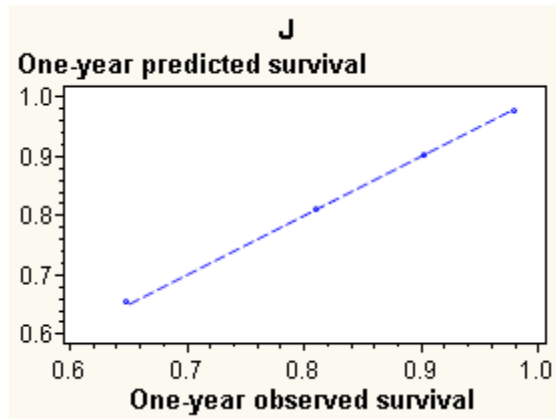
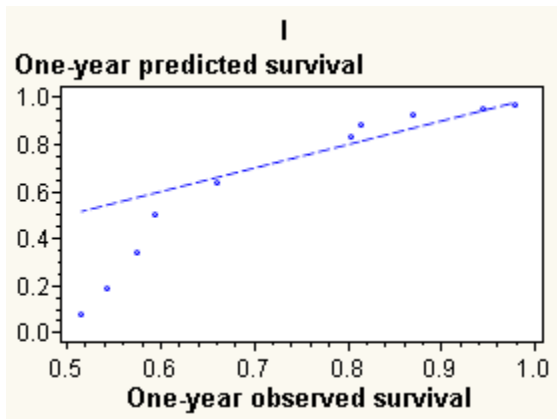
Supplement digital content 3. Participant characteristics in the 2 cohorts

	Population A		Population B	
	N= 3,299,213	%	N=3,303,428	%
Male	1,317,407	39.9	1,321,265	40.0
Age (years)				
18-24	193,957	5.9	194,710	5.9
25-34	515,935	15.6	515,148	15.6
35-44	468,308	14.2	468,091	14.2
45-54	490,511	14.9	491,166	14.9
55-64	560,133	17.0	561,231	17.0
65-74	453,397	13.7	455,836	13.8
75-84	432,265	13.1	431,905	13.1
≥85	184,707	5.6	185,341	5.6
In-hospital mortality	32,086	0.97	32,077	0.97
Charlson index comorbidities				
Myocardial infarction	146,419	4.4	146,378	4.4
Congestive heart failure	66,895	2.0	67,176	2.0
Peripheral vascular disease	103,008	3.1	103,677	3.1
Cerebrovascular disease	68,038	2.1	68,046	2.1
Dementia	85,817	2.6	85,203	2.6
Chronic pulmonary disease	319,750	9.7	321,135	9.7
Connective tissue disease	43,977	1.3	44,155	1.3
Ulcer disease	7,190	0.2	7,288	0.2
Mild liver disease	48,341	1.5	48,577	1.5
Diabetes	273,100	8.3	273,091	8.3
Hemiplegia	29,774	0.9	29,873	0.9
Moderate or severe renal disease	50,727	1.5	51,366	1.6
Diabetes with end-organ damage	90,082	2.7	90,578	2.7
Any tumor (including leukemia and lymphoma)	325,603	9.9	325,748	9.9
Moderate or severe liver disease	15,334	0.5	15,498	0.5
Metastatic solid tumor	38,473	1.2	38,488	1.2
HIV/AIDS	17,602	0.5	17,568	0.5



Supplement digital content 4 One-year survival: Kaplan-Meier Curve





Supplement digital content 5. Calibration curves:

- A) Crude model including the Charlson index as a continuous score,
- B) Crude model including the Charlson index as a 4-level score,
- C) Age- and gender-adjusted model including the Charlson index as continuous score,
- D) Age- and gender-adjusted model including the Charlson index as a 4-level score,
- E) Crude model including age-adjusted Charlson index as continuous score,
- F) Crude model including the age-adjusted Charlson index as a 4-level score,
- G) Age- and gender-adjusted model including the age-adjusted Charlson index as continuous score, H) Age- and gender-adjusted model including the age-adjusted Charlson index as a 4-level score,
- I) Crude model including the Charlson index modified by Quan as a continuous score,
- J) Crude model including the Charlson index modified by Quan as a 4-level score,
- K) Age- and gender-adjusted model including the Charlson index modified by Quan as continuous score,
- L) Age- and gender-adjusted model including the Charlson index modified by Quan as a 4-level score,
- M) Crude model including the updated Charlson index as a continuous score,
- N) Crude model including the updated Charlson index as a 4-level score,
- O) Age- and gender-adjusted model including the updated Charlson index as continuous score,
- P) Age- and gender-adjusted model including the updated Charlson index as a 4-level score.

3.2.1 Synthèse de l'étude 1

Cette étude est la première à avoir adapté un score de comorbidités aux données du SNDS. Chacune des 17 pathologies a pu être traduite en critères d'identification dans les données intra- et extra-hospitalières disponibles dans les BMA françaises (codes CIM10, actes CCAM, code CIP de médicaments).

Ce score de comorbidités de Charlson était associé de façon très significative à la mortalité toute cause à 1 an. De nouveaux poids ont pu être déterminés mettant ainsi en évidence le moindre impact de certaines pathologie (infarctus du myocarde, connectivites, pathologie ulcéreuse) pouvant probablement s'expliquer par des progrès dans leur prise en charge.

La comparaison de différentes versions de ce score a permis de montrer que l'utilisation de ce score prenant en compte l'âge comme une comorbidité présentait les meilleures performances.

Cette étude souffre cependant de certaines limites. La définition des groupes de pathologie était basée sur un algorithme en fonction des sources accessibles (ALD, diagnostics d'hospitalisation, actes médicaux, consommation médicamenteuse). Aucun retour aux dossiers médicaux pour validation n'était possible. Toutefois, une validation externe par comparaison avec la prévalence des principales pathologies en France a conforté nos résultats. Une autre limite concerne le critère de jugement retenu sur la mortalité toute cause. Son choix reposait sur la volonté de se placer dans les mêmes conditions que l'étude princeps de Charlson. Cependant, l'utilisation de ce score de comorbidité comme facteur d'ajustement dépasse largement le cadre des études utilisant le décès comme critère de jugement. L'utilisation de tout autre critère nécessiterait de revalider les performances de ce score sur cet autre critère.

Cette étude a néanmoins montré la possibilité de prendre en compte un facteur majeur de confusion dans les études épidémiologiques que sont les comorbidités. Son important référencement (plus de 146 citations depuis sa publication) est un marqueur de l'intérêt que les chercheurs portent aux données issues de bases médico-administratives et de la mise à disposition d'outils utilisables pour

mener des études sur données s'appuyant sur des approches méthodologiques améliorant leur niveau de preuve.

3.3 Construction et validation d'un algorithme d'identification des hospitalisations non programmées dans le PMSI chez les patients atteints de maladie inflammatoire chronique de l'intestin : Etude 2

3.3.1 Contexte

La réhospitalisation est un critère de jugement habituellement utilisé que ce soit en épidémiologie ou dans l'évaluation de la qualité des soins des établissements de santé (59,60). Distinguer les hospitalisations programmées des hospitalisations non programmées pourrait s'avérer indispensable afin de définir de manière plus pertinente cet indicateur. Toutefois la notion de programmation n'est pas une donnée recueillie dans la base médico-administrative du PMSI et donc de fait, non disponible dans le SNDS. L'objectif de cette étude était de construire et de valider un algorithme d'identification des hospitalisations non programmées dans le PMSI. Pour ce faire, nous nous sommes focalisés sur une pathologie chronique dont les hospitalisations sont fréquentes, qu'elles soient programmées ou non : les maladies inflammatoires chroniques de l'intestin (MICI).

3.3.2 Méthode

Un échantillon aléatoire de patients sortis d'un séjour hospitalier au CHRU de Nancy entre le 01/01/2016 et le 31/12/2018 mentionnant un code diagnostic CIM10 de maladie de Crohn (K50) ou de rectocolite hémorragique (K51), que ce soit en diagnostic principal, relié ou associé. Les séjours pour une prise en charge exclusivement en séances et ceux ayant un mode d'entrée par les urgences ont été exclus. Seul le premier séjour de chaque patient sur la période a été inclus dans l'étude. Les séjours hospitaliers ont été répartis aléatoirement en 2 jeux de données : un jeu d'apprentissage (75 %) et un jeu de validation (25 %). Un retour au dossier a permis de déterminer pour chaque séjour s'il avait été programmé (un délai de 48h a été retenu pour définir le caractère programmé) ou non programmé.

Un modèle de régression logistique a été utilisé pour rechercher une association significative entre le caractère non programmé de l'hospitalisation et les variables disponibles dans le PMSI. Cet algorithme

a ensuite été appliqué sur le jeu de validation et ses performances ont été évaluées par la sensibilité, spécificité, les valeurs prédictives et négatives ainsi que de l'aire sous la courbe ROC.

3.3.3 Résultats

Parmi les 2 656 séjours répondant aux critères de notre étude, 1 484 ont été inclus et répartis en 845 dans le jeu d'entraînement et 362 dans le jeu de validation. Un taux d'hospitalisation non programmée de 18,7 % a été observé. En majorité féminins (56,6 %), les patients avaient une moyenne d'âge de 41 ans. Leur durée médiane de séjour était de 2 jours.

Le modèle de régression logistique a identifié 6 variables associées à une hospitalisation non programmée : la provenance du patient (domicile, autre établissement de santé), la catégorie de diagnostic principal (chapitres de la CIM10), le type de prise en charge (médicale vs chirurgicale), la sévérité du GHM, la durée de séjour et une hospitalisation dans les 30 jours précédents.

Un seuil combinant une VPP de plus de 70 % et une sensibilité de plus de 66 % a été retenu pour l'algorithme. Appliqué sur le jeu de validation, il a montré des performances similaires avec une aire sous la courbe ROC de 0,912 $CI_{95\%}=[0,8731-0,944]$.

1 **Construction and evaluation of an algorithm for detecting**
2 **unplanned hospitalizations in patients with chronic inflammatory**
3 **bowel disease by using the French discharge database**

4 Luc Person^{1,2}, Yoann Dominique¹, Nicolas Jay^{1,2}, Aurélie Bannay^{1,2}

5 ¹Lorraine University, CNRS, INRIA, LORIA, F-54000 Nancy, France

6 ²CHRU-Nancy, Department of medical information, F-54000 Nancy, France

7 *Corresponding author: Aurélie Bannay, Department of Medical Information, Nancy

8 University Hospital

9 Corresponding author email: a.bannay@chru-nancy.fr

10

11 **Abstract (310)**

12 **Background**

13 Unplanned hospitalizations (UH) and readmissions are a recurring problem in medical care. In
14 order to identify them, predictive modeling algorithms can be created using data from
15 administrative databases concerning stays for inflammatory bowel diseases (IBD), a condition
16 that often leads to UH.

17 **Objective**

18 We aimed to build an algorithm, based on data available in the French national discharge
19 database, for identifying planned or unplanned hospitalizations of subjects with IBD.

20 **Method**

21 A random sample of 1484 hospital stays with a ICD-10 diagnosis code of IBD at the Nancy
22 University hospital between 2016 and 2018 were selected. The sample was randomly splitted
23 into a training dataset (70%) and a validation dataset (30%). The gold standard was
24 determined by experts using a structured decision tool to categorize each stay into planned or
25 unplanned admission according to their electronic health records. A stepwise logistic
26 regression model was performed to select covariates associated with UH on the training
27 dataset. The outcomes were the performances of the UH identification algorithm: sensitivity,
28 specificity, positive/negative predictive values (PPV/NPV) and area under the receiver
29 operating characteristics curve.

30 **Results**

31 A 18.7% rate of UH was observed in Nancy University hospital. Six covariates associated
32 with unplanned admissions were identified: origin of patient, main diagnosis, length of stay,
33 surgical treatment, severity of illness within stay and hospital stay within 30 days prior to
34 admission. The discriminatory power of the model was excellent with an AUC of 0.912
35 (95%CI 0.873 to 0.944). For a target PPV of 70.9%, the algorithm showed a sensitivity of

36 67.5% (95%CI 56.6 to 77.1%), a specificity of 93.7% (CI_{95%} 91.2 to 96.1%), and a NPV of
37 92.6% (CI_{95%} 90.5 to 94.7%) for UH.

38 **Conclusion**

39 According to our results, it possible to identify UH in the French national discharge databse
40 with promising performances. Future studies are required to validate and extend this
41 algorithm in other centers anf for other diseases.

42

43 **Keywords: unplanned admission, quality of care, discharge database, logistic regression,**
44 **inflammatory bowel disease.**

45

46 **1. Introduction**

47 The rates of unplanned hospitalizations (UH) and rehospitalizations are increasingly used
48 worldwide as indicators of the quality of care [1–5]. Their avoidance is currently a priority for
49 policy-makers in many countries [6]. Indeed, UH are frequent, expensive [7, 8] and
50 potentially avoidable [1, 9], indicating suboptimal healthcare [2, 10] and insufficiently
51 coordinated transitions between hospital and home care [11]. Numerous studies seek to
52 identify UH and rehospitalizations, especially by using predictive models [9, 12–16], in order
53 to reduce their rate and the associated healthcare costs [17].

54 Many retrospective studies on UH and rehospitalization indicators were conducted with
55 hospital administrative databases [8, 12–14]. Using these databases helps developing
56 population-level management guidelines [7] and could be of great value in creating screening
57 tools needed for improving the quality of care [1, 2, 18].

58 In France, each hospital stay is recorded in the *Programme de Médicalisation des Systèmes*
59 *d'Information* (PMSI) [19], a French adaptation of the diagnosis related groups (DRG)
60 system. This national discharge database contains administrative and medical data that have

61 previously been used to monitor quality indicators such as 30-day readmission (RH30) [8, 14,
62 20]. However, no direct information allowing to determine whether hospitalizations are
63 planned (PH) or unplanned is currently collected in this database. Hence, the lack of
64 specificity of RH30 could be improved by focusing on unplanned readmissions.
65 Planning of admissions is particularly important for patients with recurrent admissions such as
66 those with chronic diseases [21]. Inflammatory bowel diseases (IBD), are frequent chronic
67 diseases affecting over 1,5 million and 2 million people in North America and Europe [22,
68 23]. In France, Crohn's disease (CD) and ulcerative colitis (UC) concern one person per
69 thousand each [24]. These diseases, characterized by relapsing and remitting phases, regularly
70 lead to morbidity and high medical costs secondary to repeated hospitalizations [25]. Most
71 patients with IBD require long-term care and are managed in outpatient settings [26]. But as
72 disease severity progresses and complications arise, UH and avoidable hospitalizations often
73 occurs [27]. While literature shows conflicting figures for trends of rehospitalization rates ,
74 there is a need for improvement in identifying which type of hospitalizations occurs in an
75 effort to reduce unplanned admissions and to better monitor organization of hospital care [9,
76 28, 29].

77 **Objective**

78 The goal of our study was to propose an algorithm aimed at identifying unplanned admission
79 for patients with inflammatory bowel disease using the French national discharge database.

80

81 **2. Material & methods**

82 **Study design**

83 Our study included a random sample of all hospital stays discharged between January 1, 2016
84 and December 31, 2018, from the university hospital of Nancy, a tertiary center in the east
85 part of France. Only first hospitalization during the inclusion period for each patient was

86 taken into account. Inpatients with IBD were identified in the discharge database through the
87 International Classification of Diseases – 10th revision (ICD-10) codes for Crohn's disease
88 (K50) and ulcerative colitis (K51) as primary or secondary diagnoses. Ambulatory care for
89 chemotherapy, radiotherapy, dialysis, apheresis, blood transfusion and hyperbaric oxygen
90 therapy were excluded, as well as hospitalizations via the emergency department. Hospital
91 stays were also excluded if the planned or unplanned nature of the admission could not be
92 determined from the medical record.

93

94 **Data sources**

95 This study used two sources of de-identified data: the patient's electronic health record (EHR)
96 and the PMSI database. Collecting data process and analysis performed in this study were
97 conducted in accordance with the legal framework of the French data protection agency.
98 EHR was used as a gold standard to determine if stays were planned or unplanned.
99 Variables available in the PMSI were used to build an algorithm predicting unplanned
100 admission determined by the gold standard.

101

102 ○ EHR: Gold Standard

103 In accordance with literature [30], an admission was considered unplanned if at least one of
104 these criteria occurred in EHR: i) admission reported as unplanned , ii) admission reported as
105 planned less than 48 hours before admission.

106 EHR were reviewed by two investigators (LP, YD) using a structured decision tool to
107 categorize each stay into planned or unplanned admission. In case of uncertainty, the expertise
108 of a third investigator (AB) was required. This status concerning unplanned admission was
109 used as the gold standard.

110

111 ○ PMSI database: principles and variables of interest

112 The PMSI is the French national discharge database implementing the diagnosis related-
113 groups (DRG) system [31]. It contains medical information such as discharge diagnoses
114 (coded with ICD-10) and medical procedures performed during each hospital stay and
115 administrative data such as length of stay (LOS), origin and destination of patients. Several
116 variables available in the PMSI were considered as candidates for the unplanned admissions
117 predictive model: gender, age at admission, entry and discharge modes, LOS, main diagnosis,
118 DRG characteristics (major diagnostic category, medical or surgical type, severity level),
119 hospital stay within 30 days prior to admission.

120

121 **Outcomes**

122 The outcome was the performance of the UH identification algorithm measured by sensitivity,
123 specificity, Positive Predictive value (PPV), Negative Predictive Value (NPV) and AUC.

124

125 **Statistical analysis**

126 Based on a hypothesis of an expected positive predictive value of 75% with a confidence level
127 of 0.95, a width of the confidence interval of 10 points and an estimated proportion of UH at
128 nearly 20%, 1359 hospital stays were required.

129 Continuous variables were summarized using means or medians. Proportions were used for
130 categorical variables. Association between unplanned admission and the main variables
131 available in the PMSI database were assessed by Student's t-test for continuous variables and
132 the Chi-square or Fisher's test for categorical variables. Covariates significantly associated
133 with an unplanned admission in bivariate analysis at a threshold of 0.05 were included in a
134 multivariate model.

135 We randomly split study population into 70% as the training dataset and 30% as the validation
136 dataset. Multivariate analysis based on logistic regression was performed on the training
137 dataset to predict UH. A stepwise selection was used to select variables associated with
138 unplanned admission. Collinearity diagnostic was performed on covariates with statistically
139 significant association and covariates with variance inflation factor (VIF) < 2 were included
140 in the model. Discrimination refers to the ability to distinguish UH from PH and model
141 discrimination was quantified by the area under the ROC curves on the training dataset. The
142 model was considered robust if the AUCs between the training and validation datasets were
143 similar. The best prediction threshold was chosen by experts in order to reach the best
144 compromise between the predictive positive value and the sensitivity. We then externally
145 assessed the model using the validation dataset. Statistical analyses were performed using R
146 version 3.5.3 software.

147

148 **3. Results**

149 In the university hospital of Nancy, 10,547 hospital stays (1,728 patients) were recorded in the
150 2016-2018 PMSI database with inflammatory bowel disease diagnosis. A total of 7891 stays
151 met exclusion criteria: 7616 stays for previously described ambulatory criteria, 275
152 hospitalizations via the emergency department. Among the eligible 2,656 stays (1,457
153 patients), a random sample of 1520 stays was included. We additionally excluded 36
154 hospitalizations from which determination of the planned or unplanned procedure was
155 impossible. In the final sample of 1484 stays, a rate of 18.7% unplanned hospitalizations
156 (277) was observed. Mean age at inclusion was 41.1 ± 17.6 years. Men accounted for 43.4%
157 of hospitalizations. The mean length of stay was 4.2 days, with a median of 2.0 days. The
158 main discharge diagnosis was IBD for 21.0%: 13.5% for Crohn's disease and 7.5% for
159 ulcerative colitis.

160 For each of training and validation samples, characteristics of patients and hospital stays
161 according to their planned or unplanned admission are presented in [table 1](#). Referring to the
162 training dataset, patients with unplanned hospitalizations were comparable for age ($p= 0.232$)
163 and gender ($p= 0.583$). Patients with UH were more frequently transferred from another
164 hospital than patients with PH (13.9% vs 0.4%, $p<0.001$). UH were longer than PH
165 (median=6 days vs 0 day, $p<0.001$). Patients with UH suffered more frequently from digestive
166 conditions than patients with PH (57.2% vs 31.6% respectively, $p<0.001$), were more likely
167 hospitalized for mild, moderate or severe condition with nearly (74.2% vs 18.6%, $p<0.001$),
168 were more likely to undergo surgery (76.3% vs 44.7%, $p<0.001$) and had more frequently
169 previous hospitalizations within 30 days prior to admission (20.6% vs 6.7%, $p<0.001$).
170 [Table 2](#) presents PMSI covariates associated with unplanned hospitalization in standard
171 logistic regression model applied on the training dataset. The logistic regression model
172 retained five independent covariates: origin of patient, main diagnosis, surgery or medical
173 care, DRG severity, length of stay.
174 According to Figure 1, experts selected a cut-off of 0.5 corresponding to a PPV of 72.5% and
175 sensitivity of 66.5% on the training dataset. The evaluation of the model against the validation
176 dataset ([table 3](#)) confirmed the robustness of the model with AUCs of 0.91 (95% CI:0.87-
177 0.94) and a PPV of 70.9% (95% CI :62.4- 80.0%).

178

179 **4. Discussion**

180 Our study is, to our knowledge, the first to propose an algorithm identifying unplanned
181 hospitalizations of patients with inflammatory bowel disease in the French administrative
182 discharge database. The high rate of unplanned admission underlines its importance in the
183 context of a frequent chronic disease. Our findings also stress the relevance of distinguishing
184 unplanned from scheduled admissions by showing highly different patient profiles.

185 The observed UH rate of 18.7% appears to be consistent with results from previous reports.
186 Prior single center studies have reported 90-day readmission rates between 21% and 31% for
187 patients with inflammatory bowel diseases [32, 33]. In these studies, the definition of
188 unplanned hospitalizations could explain the small gap with our results. Indeed, most studies
189 focused on readmissions themselves and on patient's risk factors for readmission but were not
190 often aimed at identifying features of unplanned admissions. In the few studies performed on
191 all-cause unplanned readmissions, the rates were often lower (i.e. 16.7% [5], 11.3% to 28.1%
192 [34], 2.8% to 38% [35]). In another study on patients with chronic renal failure requiring
193 hemodialysis, nearly 25% of admissions were followed by a 30-day unplanned readmission
194 [36]. In France, the 30-day readmission indicator ranged from 12% to 15% between 2010 and
195 2015 [8, 37], without focusing on unplanned readmissions.

196 As admission via the emergency department is already well known predictor of unplanned
197 admission in the literature [12, 14, 35], we chose to focus on other types of admission. Our
198 study found characteristics that have already identified in previous studies as good predictors of
199 unplanned admission such length of stay [14] or digestive condition [38]. Nevertheless, some
200 of our findings differ from previous research. Although the association between age or gender
201 with unplanned hospitalization had already been confirmed in prior studies [5, 14, 16], we did
202 not find such association in our study. Moreover, when adjusting for confounding variables,
203 discharge mode (especially return to home) was no more associated with UH while a previous
204 study [14] claimed the opposite.

205 Our study presents some methodological strengths. Planned status of each hospital stays was
206 validated by reviewing clinical information of the EHR including medical notes and nursing
207 charts. This reviewing process by two physicians led the exclusion of very few stays from the
208 sample. The random selection of a sample from a large panel of IBD patients without age or
209 disease severity restrictions limited selection bias. However, the Nancy University Hospital is

210 a tertiary center, a referral bias could occur by including patients with more serious
211 pathologies [39]. In this way, Rinne and al. [40] reported that higher readmission rates were
212 observed in university hospitals.

213 Regarding our statistical approach, logistic regression modeling methods are widely used in
214 the literature to predict readmission and provide good predictive models [41].

215 In term of performances, our model presents a satisfying balance between a sensitivity of
216 67.5% and a specificity of 93.7%, while still achieving good predictive positive value (PPV=
217 70.9%). Our algorithm yielded a better AUC compared to the similar work of Pauly and al.
218 [14] studying 30-day readmission (0.91 vs 0.74, respectively).

219 Even if the performances are promising, it is necessary to validate these results with other
220 data. Our approach could be generalized to all French hospitals. Although it was monocentric
221 our study is based on covariates that are available in the national discharge database
222 facilitating further assessment of its external validity and multicentric studies. As the given
223 organization may vary from one hospital to another, we avoided including settings dependent
224 variables such as contact with one or more hospital units.

225 Our study is a large analysis based on the French national discharge database of hospital
226 readmissions among patients with IBD. The use of predictive models for targeting preventive
227 interventions has a great potential in public health: identifying patients at risk for readmission
228 and unplanned admission is the major step in any strategy aimed at improving chronic
229 patients.

230

231 **5. Conclusion**

232 Unplanned readmissions are a burden for health care and appropriate identification of patients
233 at risk of readmission can result in effective interventions to reduce unnecessary costs and
234 improve the quality of care. This study targets the key elements of a prediction model and

235 evaluates their contribution in terms of identifying patients at risk of readmission at the point
236 of discharge.

237 In summary, the algorithm predicts unplanned hospitalizations with satisfactory sensitivity
238 (67.5%) and positive predictive value (70.9%). It may be used to identify and classify at-risk
239 patients who should be targeted for interventions to decrease unplanned hospitalization. The
240 rate of UH in the Nancy University hospital can be used as a baseline to measure impact of
241 interventions. Future research must overcome the limitations of the model and may extend our
242 algorithm to other chronic diseases and other hospitals.

243

244 **Abbreviations**

245 PMSI: *Programme de Médicalisation des Systèmes d'Information*, IBD: Inflammatory Bowel
246 Disease, CD: Crohn's disease, UC: ulcerative colitis, UH: unplanned hospitalization, PH:
247 planned hospitalization, AUC: area under the curve, PPV: positive predictive value, NPV:
248 negative predictive value, CI: confidence interval,

249

250 **Declarations**

- 251 • Ethics approval and consent to participate

252 Not applicable.

- 253 • Consent for publication

254 Not applicable.

- 255 • Availability of data and materials

256 The underlying protected health information is not available. High-level data, such as
257 that described in the figures and tables herein, are available upon reasonable request
258 only if sharing that data would not endanger anyone's privacy.

- 259 • Competing interests
 260 The authors declare that they have no competing interests.
- 261 • Funding
 262 This research was supported by the Nancy University Hospital.
- 263 • Authors' contributions
 264 LP wrote the final manuscript. NJ and AB supervised the project. LP and YD
 265 collected the data and LP analyzed the data. All authors discussed the results,
 266 reviewed and approved the final manuscript.
- 267 • Acknowledgements
 268 Not applicable.
- 269

270 **References**

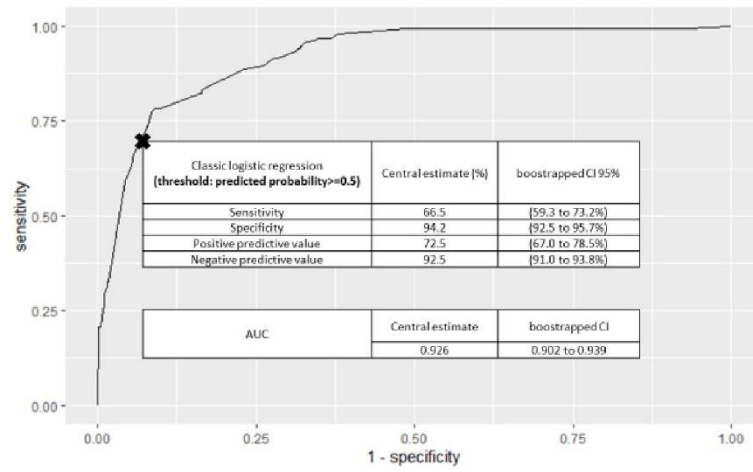
- 271 1. Halfon P, Eggli Y, Prêtre-rohrbach I, Meylan D, Marazzi A, Burnand B. Validation of the
 272 Potentially Avoidable Hospital Readmission Rate as a Routine Indicator of the Quality of
 273 Hospital Care. *Med Care.* 2006;44:972–81. doi:10.1097/01.mlr.0000228002.43688.c2.
- 274 2. Balla U, Malnick S, Schattner A. Early Readmissions to the Department of Medicine as a
 275 Screening Tool for Monitoring Quality of Care Problems. *Medicine (Baltimore).*
 276 2008;87:294. doi:10.1097/MD.0b013e3181886f93.
- 277 3. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care:
 278 advantages and limitations. *Arch Intern Med.* 2000;160:1074–81.
- 279 4. François P, Bertrand D, Beden C, Fauconnier J, Olive F. [Early readmission as an indicator
 280 of hospital quality of care]. *Rev Epidemiol Sante Publique.* 2001;49:183–92.
- 281 5. Wong EL, Cheung AW, Leung MC, Yam CH, Chan FW, Wong FY, et al. Unplanned
 282 readmission rates, length of hospital stay, mortality, and medical costs of ten common
 283 medical conditions: a retrospective analysis of Hong Kong hospital data. *BMC Health Serv*
 284 *Res.* 2011;11:149. doi:10.1186/1472-6963-11-149.
- 285 6. Kristensen SR, Bech M, Quentin W. A roadmap for comparing readmission policies with
 286 application to Denmark, England, Germany and the United States. *Health Policy Amst Neth.*
 287 2015;119:264–73.
- 288 7. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among Patients in the Medicare
 289 Fee-for-Service Program. *N Engl J Med.* 2009;360:1418–28. doi:10.1056/NEJMs0803563.

- 290 8. Gusmano M, Rodwin V, Weisz D, Cottenet J, Quantin C. Comparison of rehospitalization
291 rates in France and the United States. *J Health Serv Res Policy*. 2015;20:18–25.
292 doi:10.1177/1355819614551849.
- 293 9. Halfon P, Eggli Y, van Melle G, Chevalier J, Wasserfallen JB, Burnand B. Measuring
294 potentially avoidable hospital readmissions. *J Clin Epidemiol*. 2002;55:573–87.
- 295 10. Ashton CM, Del Junco DJ, Soucek J, Wray NP, Mansyur CL. The association between
296 the quality of inpatient care and early readmission: a meta-analysis of the evidence. *Med*
297 *Care*. 1997;35:1044–59.
- 298 11. Coleman EA, Parry C, Chalmers S, Min S-J. The care transitions intervention: results of a
299 randomized controlled trial. *Arch Intern Med*. 2006;166:1822–8.
- 300 12. Billings J, Blunt I, Steventon A, Georghiou T, Lewis G, Bardsley M. Development of a
301 predictive model to identify inpatients at risk of re-admission within 30 days of discharge
302 (PARR-30). *BMJ Open*. 2012;2:e001667. doi:10.1136/bmjopen-2012-001667.
- 303 13. Zhu K, Lou Z, Zhou J, Ballester N, Kong N, Parikh P. Predicting 30-day Hospital
304 Readmission with Publicly Available Administrative Database. A Conditional Logistic
305 Regression Modeling Approach. *Methods Inf Med*. 2015;54:560–7.
- 306 14. Pauly V, Mendizabal H, Gentile S, Auquier P, Boyer L. Predictive risk score for
307 unplanned 30-day rehospitalizations in the French universal health care system based on a
308 medico-administrative database. *PloS One*. 2019;14:e0210714.
- 309 15. Shadmi E, Flaks-Manov N, Hoshen M, Goldman O, Bitterman H, Balicer RD. Predicting
310 30-day readmissions with preadmission electronic health record data. *Med Care*.
311 2015;53:283–9.
- 312 16. Brünger B, Blozik E. Hospital readmission risk prediction based on claims data available
313 at admission: a pilot study in Switzerland. *BMJ Open*. 2019;9:e028409.
- 314 17. Tabak YP, Sun X, Nunez CM, Gupta V, Johannes RS. Predicting Readmission at Early
315 Hospitalization Using Electronic Clinical Data: An Early Readmission Risk Score. *Med Care*.
316 2017;55:267–275. doi:10.1097/MLR.0000000000000654.
- 317 18. Hekkert K, van der Brug F, Borghans I, Cihangir S, Zimmerman C, Westert G, et al. How
318 to identify potentially preventable readmissions by classifying them using a national
319 administrative database. *Int J Qual Health Care*. 2017;29:826–32. doi:10.1093/intqhc/mzx110.
- 320 19. Guide Méthodologique MCO 2019 | Publication ATIH. [https://www.atih.sante.fr/guide-](https://www.atih.sante.fr/guide-methodologique-mco-2019)
321 [methodologique-mco-2019](https://www.atih.sante.fr/guide-methodologique-mco-2019). Accessed 11 Jul 2019.
- 322 20. DGOS. Les indicateurs de réhospitalisation et de coordination. Ministère des Solidarités et
323 de la Santé. 2019. [https://solidarites-sante.gouv.fr/soins-et-maladies/qualite-des-soins-et-](https://solidarites-sante.gouv.fr/soins-et-maladies/qualite-des-soins-et-pratiques/qualite/les-indicateurs/article/re-hospitalisation-coordination)
324 [pratiques/qualite/les-indicateurs/article/re-hospitalisation-coordination](https://solidarites-sante.gouv.fr/soins-et-maladies/qualite-des-soins-et-pratiques/qualite/les-indicateurs/article/re-hospitalisation-coordination). Accessed 23 Jul 2019.
- 325 21. Malhotra A, Phatharacharukul P, Thongprayoon C. Risk factors for 90-day readmission in
326 veterans with inflammatory bowel disease-Does post-discharge follow-up matter? *Mil Med*
327 *Res*. 2018;5:5.

- 328 22. Burisch J, Jess T, Martinato M, Lakatos PL, ECCO -EpiCom. The burden of
329 inflammatory bowel disease in Europe. *J Crohns Colitis*. 2013;7:322–37.
- 330 23. Kaplan GG. The global burden of IBD: from 2015 to 2025. *Nat Rev Gastroenterol*
331 *Hepatol*. 2015;12:720–7.
- 332 24. Peyrin-Biroulet L, Bouhnik Y, Roblin X, Bonnaud G, Hagège H, Hébuterne X.
333 Algorithmes de prise en charge de la maladie de Crohn en 2016 : Consensus National
334 Français. *Hépto-Gastro Oncol Dig*. 2016;23:619–33. doi:10.1684/hpg.2016.1332.
- 335 25. Dulai PS, Singh S, Ohno-Machado L, Sandborn WJ. Population Health Management for
336 Inflammatory Bowel Disease. *Gastroenterology*. 2018;154:37–45.
337 doi:10.1053/j.gastro.2017.09.052.
- 338 26. Weizman AV, Nguyen GC. Quality of care delivered to hospitalized inflammatory bowel
339 disease patients. *World J Gastroenterol*. 2013;19:6360–6.
- 340 27. Bewtra M, Su C, Lewis JD. Trends in Hospitalization Rates for Inflammatory Bowel
341 Disease in the United States. *Clin Gastroenterol Hepatol*. 2007;5:597-601.e1.
342 doi:10.1016/j.cgh.2007.01.015.
- 343 28. Poojary P, Saha A, Chauhan K, Simoes P, Sands BE, Cho J, et al. Predictors of Hospital
344 Readmissions for Ulcerative Colitis in the United States: A National Database Study. *Inflamm*
345 *Bowel Dis*. 2017;23:347–56.
- 346 29. Hazratjee N, Agito M, Lopez R, Lashner B, Rizk MK. Hospital Readmissions in Patients
347 With Inflammatory Bowel Disease. *Am J Gastroenterol*. 2013;108:1024–32.
348 doi:10.1038/ajg.2012.343.
- 349 30. Malafaye N, Aubas P. Comparaison de deux méthodes de calcul du taux de réadmission
350 dans les 30jours après chirurgie de la cataracte. *Rev D'Épidémiologie Santé Publique*.
351 2018;66:S5–6. doi:10.1016/j.respe.2018.01.004.
- 352 31. Boudemaghe T, Belhadj I. Data Resource Profile: The French National Uniform Hospital
353 Discharge Data Set Database (PMSI). *Int J Epidemiol*. 2017;46:392–392d.
354 doi:10.1093/ije/dyw359.
- 355 32. Tinsley A, Naymagon S, Mathers B, Kingsley M, Sands BE, Ullman TA. Early
356 readmission in patients hospitalized for ulcerative colitis: incidence and risk factors. *Scand J*
357 *Gastroenterol*. 2015;50:1103–9.
- 358 33. Allegretti JR, Borges L, Lucci M, Chang M, Cao B, Collins E, et al. Risk Factors for
359 Rehospitalization Within 90 Days in Patients with Inflammatory Bowel Disease. *Inflamm*
360 *Bowel Dis*. 2015;21:2583–9.
- 361 34. Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, et al. An administrative
362 claims measure suitable for profiling hospital performance based on 30-day all-cause
363 readmission rates among patients with acute myocardial infarction. *Circ Cardiovasc Qual*
364 *Outcomes*. 2011;4:243–52.

- 365 35. van Walraven C, Bennett C, Jennings A, Austin PC, Forster AJ. Proportion of hospital
366 readmissions deemed avoidable: a systematic review. *CMAJ Can Med Assoc J J Assoc*
367 *Medicale Can.* 2011;183:E391-402.
- 368 36. Chan L, Chauhan K, Poojary P, Saha A, Hammer E, Vassalotti JA, et al. National
369 Estimates of 30-Day Unplanned Readmissions of Patients on Maintenance Hemodialysis. *Clin*
370 *J Am Soc Nephrol CJASN.* 2017;12:1652–62.
- 371 37. Indicateurs de coordination | Stats ATIH.
372 <https://www.scansante.fr/applications/indicateurs-de-coordination>. Accessed 20 Aug 2019.
- 373 38. Bailey MK, Weiss AJ, Barrett ML, Jiang HJ. Characteristics of 30-Day All-Cause
374 Hospital Readmissions, 2010–2016: Statistical Brief #248. In: *Healthcare Cost and Utilization*
375 *Project (HCUP) Statistical Briefs.* Rockville (MD): Agency for Healthcare Research and
376 Quality (US); 2006. <http://www.ncbi.nlm.nih.gov/books/NBK538941/>. Accessed 17 Aug
377 2019.
- 378 39. Salive ME. Referral Bias in Tertiary Care: The Utility of Clinical Epidemiology. *Mayo*
379 *Clin Proc.* 1994;69:808–9. doi:10.1016/S0025-6196(12)61105-7.
- 380 40. Rinne ST, Castaneda J, Lindenauer PK, Cleary PD, Paz HL, Gomez JL. Chronic
381 Obstructive Pulmonary Disease Readmissions and Other Measures of Hospital Quality. *Am J*
382 *Respir Crit Care Med.* 2017;196:47–55.
- 383 41. Artetxe A, Beristain A, Graña M. Predictive models for hospital readmission risk: A
384 systematic review of methods. *Comput Methods Programs Biomed.* 2018;164:49–64.
385 doi:10.1016/j.cmpb.2018.06.006.
- 386

387 **Figures**



388

389 Figure 1 Performance curve from the training dataset. The selected threshold is highlighted
390 by a black cross.

391

392 **Tables**

393 Table 1 Characteristics of patients and hospital stays according to their planned or unplanned
394 admission

	Train dataset				P valu e	Test dataset			
	Planned hospitalization		Unplanned hospitalization			Planned hospitalization		Unplanned hospitalization	
	N	%	N	%		N	%	N	%
	845	81.3	194	18.7		362	81.3	83	18.7
Patients									
Male	365	43.2	88	45.5	0.583	155	42.8	36	43.4
Age (years) mean ± SD	41.0 ± 17.5		43.5 ± 19.4		0.232	40.5 ± 17.2		38.4 ± 15.5	
Hospital stay characteristics									
Origin of patient					<0.001				
From home	842	99.6	167	86.1		358	98.9	68	81.9
Transfert from another hospital	3	0.4	27	13.9		4	1.1	15	18.1
Destination on discharge					0.013				
Return to home	831	98.3	184	94.8		358	98.9	79	95.2
Transfert to another hospital	13	1.5	9	4.6		3	0.8	3	3.6
Deceased	1	0.1	1	0.5		1	0.3	1	1.2
Length of hospital stay (days)									
Mean ± SD	2.9 ± 6.3		10.5 ± 16.7		<0.001	2.9 ± 7.2		8.6 ± 9.7	
Median	0		6			0		6	
Classes of length of stay of more than one day	420	50.3	192	99.0	<0.001	172	47.5	82	98.8
ICD-10 chapter used for coding main diagnose					<0.001				
Certain infectious and parasitic diseases	1	0.1	6	3.1		1	0.3	5	6.0
Diseases of the digestive system	267	31.6	111	57.2		113	31.2	42	50.6
Diseases of the musculoskeletal system and connective tissue	56	6.6	13	6.7		27	7.5	6	7.2
Pregnancy, childbirth and the puerperium	2	0.2	4	2.1		1	0.3	6	7.2
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	15	1.8	14	7.2		8	2.2	5	6.0
Injury, poisoning and certain other consequences of external causes	2	0.2	8	4.1		1	0.3	1	1.2
Factors influencing health status and contact with health services	403	47.7	6	3.1		183	50.6	3	3.6
Other chapters	99	11.7	32	16.5		28	7.7	15	18.1
Major Diagnostic Category					<0.001				
Diseases and Disorders of the Digestive System	573	67.8	121	62.4		249	68.8	43	51.8
Factors Influencing Health Status and Other Contacts with Health Services	81	9.6	5	2.6		37	10.2	3	3.6
Other categories	191	22.6	68	35.1		76	21.0	37	44.6
DRG type					<0.001				
Medical groups	378	44.7	148	76.3		164	45.3	68	81.9
Surgical groups									
Major procedure	154	18.2	42	21.6		61	16.9	15	18.1

Minor procedure	313	37.0	4	2.1		137	37.8	0	0.0
DRG severity					<0.001				
None	688	81.4	50	25.8		298	82.3	28	33.7
Mild, moderate, severe	157	18.6	144	74.2		64	17.7	55	66.3
Hospital stay within 30 days prior to admission	57	6.7	40	20.6	<0.001	30	8.3	18	21.7
Outpatient care within 30 days prior to admission	62	7.3	18	9.3	0.360	25	6.9	11	13.3

Note : N=Number, SD = Standard deviation; DRG = Diagnosis Related Group; ICD-10 = International Classification of Disease - 10th revision

395 Table 2 Association between unplanned hospitalisation and characteristics of hospital stays
 396 from the French discharge database

	Estimates	SE	p-value
Origin of patient			<0.001
From home	Reference		
Transfert from another hospital	3.2	0.7	<0.001
ICD-10 chapter used for coding main diagnose			
Factors influencing health status and contact with health services	Reference		<0.001
Certain infectious and parasitic diseases	3.6	1.2	0.003
Diseases of the digestive system	2.1	0.5	<0.001
Diseases of the musculoskeletal system and connective tissue	1.9	0.6	0.001
Pregnancy, childbirth and the puerperium	6.9	1.4	<0.001
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	3.7	0.6	<0.001
Injury, poisoning and certain other consequences of external causes	4.2	1.1	<0.001
Other chapters	1.5	0.5	0.002
DRG type			<0.001
Medical groups	Reference		
Surgical group with major procedure	-1.5	0.3	<0.001
Surgical group with minor procedure	0.7	0.8	0.375
DRG severity			<0.001
Low	Reference		
Mild, moderate and severe	1.8	0.3	<0.001
Length of hospital stay			
Outpatient settings	Reference		
One night and more	4.0	1.0	<0.001
Hospital stay within 30 days prior to admission	0.8	0.3	0.008

Note : SE = Standard Error

397

398 Table 3 Performances of the algorithm obtained by classic logistic regression on validation

399 data

	Classic logistic regression (threshold: predicted probability>=0.5)	
	Central estimate	bootstrapped CI 95%
Sensitivity (%)	67.5	(56.6 to 77.1%)
Specificity (%)	93.7	(91.2 to 96.1%)
Positive predictive value (%)	70.9	(62.4 to 80.0%)
Negative predictive value (%)	92.6	(90.5 to 94.7%)
AUC	0.912	(0.873 to 0.944)

Note : CI = Confidence Interval

400

3.3.4 Synthèse de l'étude 2

Cette étude s'est attachée à développer un algorithme de reconstruction d'une donnée absente des BMA françaises : le caractère programmé ou non des hospitalisations chez des patients atteints de MICI.

La validation de cet algorithme a reposé sur le statut réel du caractère programmé de l'hospitalisation déterminé par un retour manuel au dossier du patient par un expert médecin.

L'évaluation de cet algorithme sur un jeu de validation a montré de bonnes performances.

Ce travail présente toutefois quelques limites. L'échantillon de patients est extrait d'un centre de référence dont les taux d'hospitalisation non programmée pourraient y être plus élevés que dans des centres moins spécialisés. De plus, l'étude était centrée sur une pathologie chronique unique, celle des patients atteints de MICI, non représentative de l'ensemble des pathologies sujettes à des hospitalisations non programmées. Une extension sur un échantillon plus varié et dans d'autres pathologies est envisagée.

Cette étude a néanmoins démontré que l'absence d'une donnée dans les bases médico-administratives pouvait potentiellement être imputée par l'utilisation d'algorithmes permettant d'enrichir la description des séjours. La mise en place d'un proxy est ainsi une solution possible pour pallier l'absence d'une donnée recueillie directement. Leur emploi dépend cependant d'une méthodologie rigoureuse permettant de les valider à partir d'études dédiées. Ces proxy peuvent prendre la forme d'un algorithme probabiliste comme dans cette étude, mais ils peuvent également correspondre à des alternatives déterministes. Différents exemples de proxy déterministes dans les SNDS sont décrits dans la littérature. On peut évoquer l'utilisation de substituts nicotiques pour identifier un statut tabagique (61) ou encore un comportement de doc shopping c'est à dire un chevauchement de prescription de praticiens différents d'une même molécule pour déterminer un mésusage d'opioïdes (62).

3.4 Complémentarité des données issues des bases médico-administratives et des entrepôts hospitaliers : illustration par une étude de cohorte sur les patients consommateurs de statine : Etude 3

3.4.1 Contexte

Dans le cadre de la réutilisation des données en recherche, les bases médico-administratives permettent de travailler sur un nombre limité de données mais sur un grand nombre de sujets. A l'inverse, les données d'entrepôt hospitalier offrent un plus large éventail de données (clinique, biologiques...) sur une population spécifique de patients. Pouvoir chaîner ces données pour un même sujet permettrait d'obtenir une vision plus fine de la trajectoire de soins du sujet. En France, il demeure encore complexe d'apparier des données du SNDS avec des données d'entrepôts hospitaliers à partir du numéro d'inscription au répertoire des personnes physiques (NIR). L'objectif de ce travail était de montrer, au travers d'un cas d'usage en pharmacovigilance, la facilité de l'intégration, partage et la réutilisation des données issues du SNDS et de l'entrepôt de données hospitalières du CHU de Rennes, par l'utilisation de plateformes technologiques adaptées. Le cas d'usage s'intéressait à la description des trajectoires de soins de patients consommateurs de statines et de la survenue d'interactions médicamenteuses impliquant ces dernières.

3.4.2 Méthode

Les patients sortis d'une hospitalisation MCO hors séances entre 2015 et 2017 du CHU de Rennes étaient inclus dans l'étude.

Trois sources de données ont été utilisées : une extraction SNDS, une extraction de l'entrepôt de données du CHU de Rennes (eHop) et la base de connaissance Thériaque sur le médicament. Un algorithme semi-déterministe a été développé pour apparier les sujets à partir des données PMSI communes aux extractions SNDS et eHop. Le cas d'usage a restreint les sujets inclus aux sujets avec une délivrance de statines dans les 30 jours précédant leur première sortie d'hospitalisation au CHU de Rennes sur la période 2015-2017 ou une administration de statines lors de ce séjour hospitalier. Les caractéristiques des séjours et des patients ont pu être examinés au travers des données disponibles

dans eHop. Les distributions des résultats des principaux dosages réalisés à l'hôpital ont pu être décrits en fonction de la survenue d'interactions.

3.4.3 Résultats

L'appariement des données SNDS et eHop a retrouvé des taux de succès très élevés avec 98,7 % des 161 532 sujets issus de l'extraction SNDS et 97,0 % des 164 316 sujets issus de eHop. L'extraction des 16 806 sujets de la cohorte de patients consommateurs de statines et l'obtention de leurs données de consommation médicamenteuse enrichies par les informations de la base thériaque a pu être traitée dans un temps raisonnable inférieur à 90 minutes pour l'ensemble des analyses grâce à l'utilisation de technologies de calculs distribués.

Parmi les 16 806 sujets inclus, près de 38 % (6 382) ne consommaient des statines qu'en ville, 49 % (8 293) poursuivaient les statines une fois entrés à l'hôpital et près de 13 % (2 131) initiaient les statines au décours de leur séjour hospitalier. En fonction de niveaux de sévérité allant de précautions d'usage (niveau 3) à contre-indication (niveau 1), 5 579 sujets ont été identifiés avec une délivrance (en ville) et/ou une administration (à l'hôpital) de molécule pouvant être en interaction avec la statine consommée. Les séjours des patients présentant une interaction avaient des caractéristiques plus sévères que ceux n'en présentant pas (durée de séjour allongée, niveau de sévérité des GHM plus élevé...) Les molécules les plus fréquemment retrouvées dans le cadre d'une contre-indication avec les statines étaient la cyclosporine, l'acide fusidique, l'itraconazole, le fénofibrate et l'érythromycine. Des élévations de glycémie ont pu être observées en particulier chez les sujets ayant reçu de l'acide fusidique durant leur séjour hospitalier. De la même manière, une élévation des enzymes hépatiques et de la kaliémie ont été observées chez les sujets ayant reçu de l'itraconazole.

Original Paper

Aurelie Bannay^{1,2}, MD, MSc
Mathilde Bories^{2,3,4},
Pascal Le Corre^{3,4,5}, PharmD, PhD
Christine Riou², MD
Pierre Lemordant²,
Pascal Van Hille²,
Emmanuel Chazard⁶, MD, PhD
Xavier Dode^{7,8}, PharmD
Marc Cuggia², MD, PhD
Guillaume Bouzillé², MD, PhD

¹Université de Lorraine, CHRU-Nancy, CNRS, Inria, LORIA, F-54000 Nancy, France

²Univ Rennes, CHU Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France

³Pôle Pharmacie, Service Hospitalo-Universitaire de Pharmacie, CHU de Rennes,
35033 Rennes, France

⁴Laboratoire de Biopharmacie et Pharmacie Clinique, Faculté de Pharmacie,
Université de Rennes 1, 35043 Rennes, France

⁵CHU Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement
et travail)-UMR_S 1085, Univ Rennes, F-35000 Rennes, France

⁶Univ Lille, CHU Lille, CERIM EA2694, F-59000 Lille, France

⁷Centre National Hospitalier d'Information sur le Médicament, Paris, France.

⁸Pharmacy Department, Hospices Civils de Lyon University Hospital, Lyon, France

Corresponding author:

Guillaume Bouzillé

Données massives en santé, Université de Rennes 1, Villejean - Bâtiment 6 - Etage 4

2 avenue du professeur Léon Bernard

35043 Rennes France

guillaume.bouzille@chu-rennes.fr

+33 223234749

Leveraging national claim and hospital big data integration: a cohort study on a statin-drug interaction use case.

Abstract

Background: Linking different sources of medical data is a promising approach to analyse care trajectories. The INSHARE project aim was to provide the blueprint of a technological platform that facilitates integration, sharing and reuse of data from two sources: the eHOP clinical data warehouse (CDW) of Rennes academic hospital, and a dataset extracted from the French national claim data warehouse (SNDS).

Objective: Using a pharmacovigilance use case based on statin consumption and statin-drug interactions, the present work demonstrated how the INSHARE platform can support big data analytical tasks in the health field.

Methods: A Spark distributed cluster-computing framework was used for the record linkage procedure and all the analyses. A semi-deterministic record-linkage method based on the variables common between the chosen data sources was developed to identify all patients discharged after at least one hospital stay at Rennes academic hospital between 2015 and 2017. The use case study focused on a cohort of patients treated with statins prescribed by their general practitioner and/or during their hospital stay.

Results: The whole process (record-linkage procedure and use case analyses) required 88 minutes. Among the 161,532 and 164,316 patients from the SNDS dataset and eHOP CDW, respectively, 159,495 patients were successfully linked (98.7% and 97.0% of patients from SNDS and eHOP CDW, respectively). Among the 16,806 patients with at least one statin delivery, 8,293 patients started the consumption before and continued during the hospital stay, 6,382 patients stopped statin consumption at hospital admission, and 2,131 patients initiated taking statins in hospital. Statin-drug interactions occurred more frequently during hospitalization than in the community (36.4% and 22.2%, respectively $p < .001$). Only 121 patients had the most severe level of statin-drug interaction. Hospital stay burden (length of stay and in-hospital mortality) was more severe in patients with statin-drug interactions during hospitalization.

Conclusions: This study demonstrated the added value of combining and re-using clinical and claim data to provide large-scale measures of drug-drug interaction prevalence and care pathways outside hospitals. It builds the path to move the current healthcare system towards a Learning Health System using knowledge generated from research on real-world health data.

Keywords: Drug Interactions; Statins; Administrative Claims; Healthcare; Big Data; Data Linking; Data Warehousing

Introduction

The secondary use of healthcare data offers the opportunity to carry out observational studies in real life [1–3]. Indeed, hospital Clinical Data Warehouses (CDW) supply fine-grained information from electronic medical records (EHR), such as laboratory test results and drug administration, but are restricted to hospitalized patients. Conversely, National claim databases offer limited information (e.g. drug reimbursement and healthcare consumption data), but on a large part of the population. Therefore, matching data from these two different databases could be informative, but it is challenging. Patients existing in the two databases have to be correctly identified using appropriate record linkage methods. The first option is deterministic record linkage that relies on the presence of a unique common identifier or a combination of different variables used as the key to join tables [4]. More complex rules to link records can also be added, such as acceptable distance between string variables or between dates. The second option is probabilistic record linkage that is based on a model to assess the discriminative power of each variable used in the record linkage strategy. The result is the probability that an entity in the first database is the same entity in the second database [5,6]. Several studies demonstrated that in most cases, probabilistic approaches give better results than deterministic methods [7–10]. However, the choice of record linkage also heavily depends on the characteristics of the two databases to be linked. The quality of the data used in the record linkage is an especially important factor. Indeed, if high quality data (e.g. few missing values) are available, deterministic methods can achieve good results and are easier to develop [11].

In France, the national health database (SNDS: *Système National des Données de Santé*) [12] links the nationwide outpatient claim database (SNIIRAM), the national discharge database (PMSI), and the Epidemiology Centre of Medical Causes of Death (CepiDC; vital status data). Rennes academic hospital (CHU Rennes) uses eHOP [13], a CDW that includes EHR and discharge data on all stays in this hospital. Linking SNDS and eHOP is a promising strategy to analyse the patient care trajectories. However, legal, methodological, and technical barriers still remain. Health data are sensitive and in France, their use is regulated by the European General Data Protection Regulation (GDPR)[14]. Therefore, studies based on the use of health data entail various regulatory steps, such as the scientific evaluation of the project and of the patient information material and also the assessment of the impact on data protection. In France, the use of SNDS data for external research requires the development of a data repository that complies with strict security specifications to host the SNDS sample for the study.

In this context, the aim of the INSHARE project was to provide the blueprint of a technological platform (INSHARE platform) that facilitates data integration, sharing and reuse by following the FAIR Guiding Principles [15]. The present work demonstrates through a use case in pharmacovigilance how the INSHARE platform can support health big data analysis.

Our use case focused on statin consumption and statin-related drug-drug interactions (DDIs). Indeed, 36.9% [16] of 34-65-year-old French people have hypercholesterolemia, and statins are the most prescribed lipid-lowering treatment in France [17]. The current European treatment guidelines [18] recommend statins as the first-choice drug for hypercholesterolemia management. However, 10-25% of patients treated with statins experience muscle side effects [19], including rhabdomyolysis (incidence: 1-3 in 100,000

persons per year) [20]. Statin-induced rhabdomyolysis is related to DDIs in 60% of cases [20], which suggests that avoiding DDIs has an important role in reducing statin adverse events. Due to their wide use and DDI potential, statins are an interesting study topic to assess the value of our technological platform for clinical data reuse. Moreover, literature data indicate that DDIs are preventable, but this is hindered by the clinicians' lack of easy access to comprehensive information. Indeed, healthcare delivery is fragmented across the system and this creates an environment susceptible to medication-related issues [21]. Polypharmacy has been associated with higher risk of DDIs and of adverse drug events [22], and subsequently, with drug-related deaths in hospitals [23]. Therefore, it is important to precisely characterize the individual care pathways within the healthcare system using aggregated medical data.

Here, we present the technical aspects of the INSHARE platform and the methods and results of the care pathway analysis in patients with statin-drug interactions.

Methods

Data sources

Drug database: Thériaque

Thériaque [24] is a comprehensive dynamic knowledge database that provides exhaustive information on approved and marketed drugs. It contains highly structured information on each drug, such as indications, contraindications, and DDIs and their severity level. Each drug is referenced according to three mapped classifications: UCD (Unité Commune de Dispensation = medication-dispensing unit) used by the French hospital information system; CIP (Code Identifiant de la Présentation = drug package identifier) used by French community pharmacies; and ATC (Anatomic Therapeutic Codification) based on the drug active component(s).

French Claim database: SNDS

In France, the SNDS is a national claim data warehouse that covers 98.8% of the entire French population [25]. It contains data from outpatient care, such as medical consultations and drug deliveries by community pharmacies, and data from inpatient care, such as diagnosis and procedures performed during a stay in a private- or public-sector hospital. Each reimbursement of outpatient care is recorded at the individual level in a specific data mart called DCIR (Datamart de Consommation InterRégime) [12]. Data on inpatient care also are recorded at the individual level in an annual national discharge database called PMSI (Programme de Médicalisation des Systèmes d'Information) that is similar to the Diagnosis-related groups (DRGs). Individual data are deidentified and pseudonymised allowing the linkage, thanks to a unique identifier, between inpatient data (PMSI database) and outpatient data (DCIR). This claim data warehouse has been previously described [12].

We used a dataset extracted from the SNDS database that included all patients discharged after at least one hospital stay at Rennes academic hospital between 2015 and 2017. Due to the redundancy of information contained in the PMSI database, hospital stays following the primary diagnosis were excluded (e.g. stays for chemotherapy, radiotherapy, dialysis, apheresis, blood transfusion and hyperbaric oxygen therapy). All inpatient and outpatient data in the 12 months before each hospital stay were extracted.

Clinical data warehouse: eHOP

eHOP is the CDW developed and deployed at Rennes academic hospital [13]. It collects administrative and clinical data from EHRs: unstructured (e.g. clinical notes) and structured (e.g. drugs, laboratory results). Data are de-identified and a unique anonymous identifier allows the linkage among hospital stays of a given patient. The eHOP CDW currently allows searching 80 million unstructured data and 430 million structured elements. All these data are collected from EHRs and cover more than 1.4 million patients.

The dataset from the eHOP database included patients according to the same criteria used for the SNDS data: all data on hospital stays at Rennes academic hospital between 2015 and 2017. For this study, we used the following structured data:

- Demographic data
- Drug administered (UCD and date of administration)
- PMSI data: ICD-10 codes, procedure codes, mortality, length-of stay, etc.
- Laboratory results described with a local terminology.

Record linkage procedure

As no unique patient identifier is available to link SNDS and eHOP data due to regulatory issues, we developed a semi-deterministic record-linkage method based on PMSI variables that are common between the SNDS data source and the eHOP CDW data source (Figure 1). PMSI data are available from all French hospitals and are produced in a standardized way by each hospital. Once de-identified, PMSI data feed the nationwide PMSI database. This database is then integrated in the SNDS database to link PMSI data with claim data. In theory, PMSI data from the SNDS and hospitals should be exactly the same. However, during preliminary works, we identified some discrepancies concerning ICD-10 and procedure codes between these data sources. Therefore, we incorporated in the record linkage algorithm some fuzzy logic to solve inconsistencies. Specifically, ICD-10 codes comprise between three and six characters, but we kept only the first four characters. Procedure codes comprise seven characters and we kept all seven. We merged ICD-10 and procedure codes in alphabetical order in a unique string for each stay. We then tested different Levenshtein distance thresholds to consider a match between sets of codes (the distribution of the Levenshtein distances for ICD-10 codes and procedure codes is in supplementary table 1). We identified the threshold of 5 as the best choice for both ICD-10 and procedure codes. For the final matching, first we assessed whether a patient had at least one exact match. This was considered as the exact match if the other patients were fuzzy matches. If we did not find any exact match, we kept the fuzzy match first looking at procedure codes. If a patient had several exact matches or several fuzzy matches, we kept the one with the most fuzzy matches on ICD-10 codes. Remaining patients with several matches were considered them as duplicates that were excluded from the linkage results.

We also had to solve specific cases concerning twins who do not have an individual identifier (NUM_ENQ) in the PMSI. Indeed the same identifier (NUM_ENQ) is shared by twins of the same sex [12]. Thus, it was impossible to link their SNDS records with their records in eHOP. We chose to exclude twin patients from the record linkage results.

We assessed the linkage effectiveness by calculating the rate of SNDS and eHOP patients who could be matched in the other dataset. We also described some characteristics of the following groups: patients who were matched between data sources, and patients from the SNDS and eHOP datasets who could not be matched.

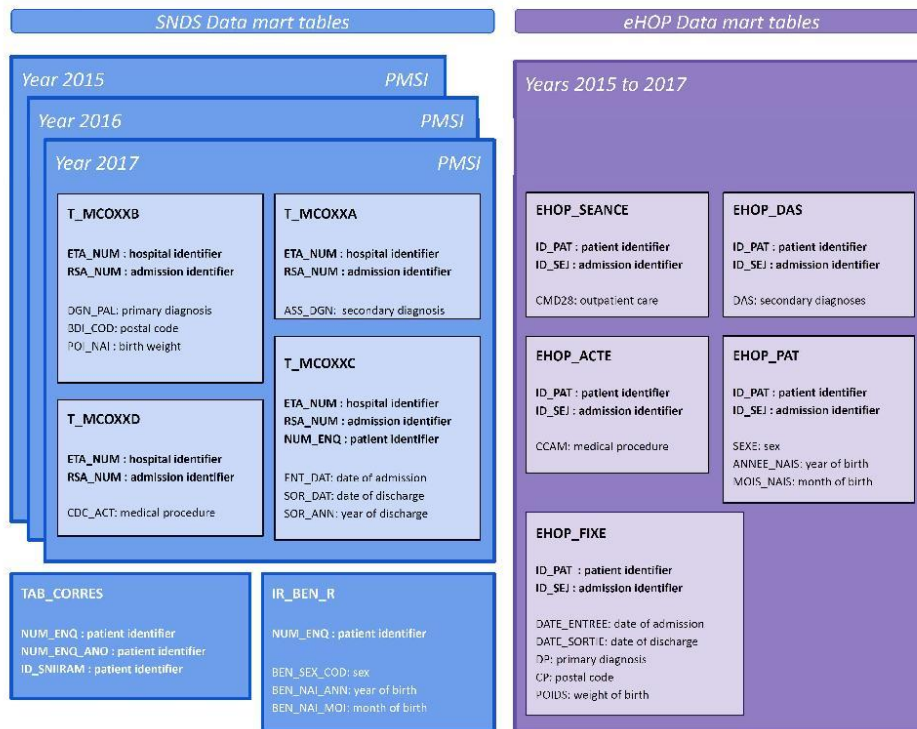


Figure 1. SNDS data mart tables (in blue) including PMSI tables, and eHOP data mart tables (in purple) with the different variables from the two data sources used for the linkage procedure.

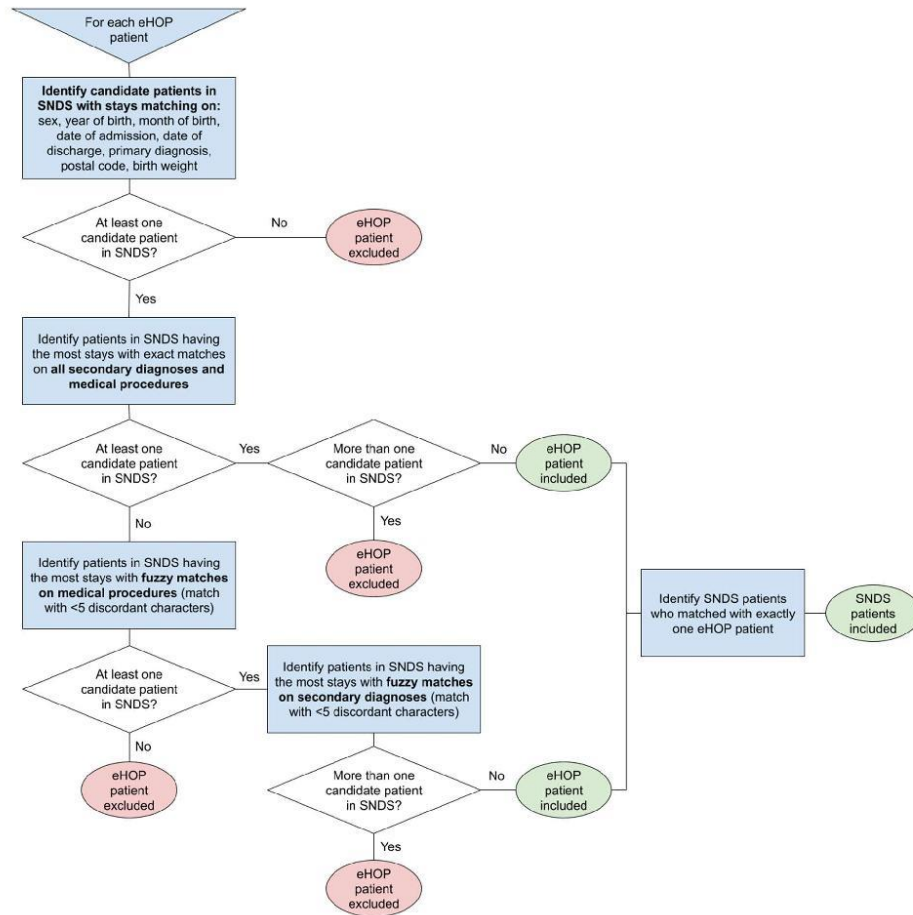


Figure 2. Decision tree for the record-linkage procedure.

INSHARE platform

The INSHARE platform comprises two parts: a data repository to gather any kind of data sources, and a computing infrastructure to perform data preparation, record linkage and analyses. The platform is available through Apache Mesos, a resource manager, to allow concurrent access to the computing server.

The data repository was the Apache Hadoop HDFS repository, and data were stored in parquet format files, with an appropriate stratification key. SNDS datasets were made available to us in csv files that were stored in a specific folder in the server. We extracted the data needed from the eHOP CDW and the Thériaque database with Spark SQL. This

extraction step avoided repeating long queries in the CDW and overloading the production CDW used for other purposes.

We used the Spark distributed computing framework, version 2.3.4, for the data preparation, the record linkage procedure and all the use case analyses.

We then accessed these data with Spark SQL that allowed us to merge data from the different sources in an efficient way and to perform all analyses. We used the R language as script language and particularly the sparklyr package. The overall data processing is depicted in Figure 3.

We used a single node cluster: a CentOS 7 Unix server with 2 Intel Xeon 5122 @ 3.6 GHz and 192 Gb of RAM. Thus, we did not replicate the HDFS repository, and executed the Spark master and slave nodes on the same machine.

Use case study design

We performed a cohort study on patients treated with statins prescribed by their general practitioner and/or during the hospital stay. We collected information on statins (ATC classes C10AA, C10BA, and C10BX) and the statin-drug interactions from the Thériaque database. We classified statin intake as i) community consumption if we found at least one statin delivery by a community pharmacy less than 1 month before hospitalization, and ii) hospital consumption if we found at least one statin administered during the hospital stay. Only the first hospital stay for each patient was retained for the use case.

For each patient, we extracted the following features: sex, age at admission, the international non-proprietary name of the used statin, consumption of drugs potentially interacting with the used statin, DDI severity, admission via the emergency department, length of hospital stay, in-hospital death, laboratory results: creatine phosphokinase (CPK), creatininaemia, glycaemia, haemoglobin, kalaemia, natremia, aspartate aminotransferase (AST), alanine aminotransferase (ALT), hospital care burden (i.e. DRG severity).

We classified patients in three subgroups according to their statin consumption status: i) patients treated with statins before and during their hospital stay, ii) patients treated with statins before admission, but not during the hospital stay, iii) patient who started taking statins in hospital without any statin treatment in the 12 previous months. We defined a statin-related DDI on the basis of the intake (hospital administration or community delivery within 8 days before or after the statin) of a drug that reacts with the statin taken by that patient. According to the Thériaque database, we classified all statin-drug interactions in three levels of severity (level 1: contraindication, level 2: relative contraindication, level 3: precaution of use).

Statistical analyses

We described categorical variables as numbers and percentages, and quantitative variables as mean and standard deviation (SD) in the case of symmetric distribution, and median with first and third quartiles (Q1-Q3), otherwise. We explored association between patients' characteristics or hospital stays and the occurrence of a statin-drug interaction with the Chi-square test (categorical variables) and one-way analysis of variance (quantitative variables). We built a logistic regression model to identify factors independently related to the occurrence of a statin-interaction.

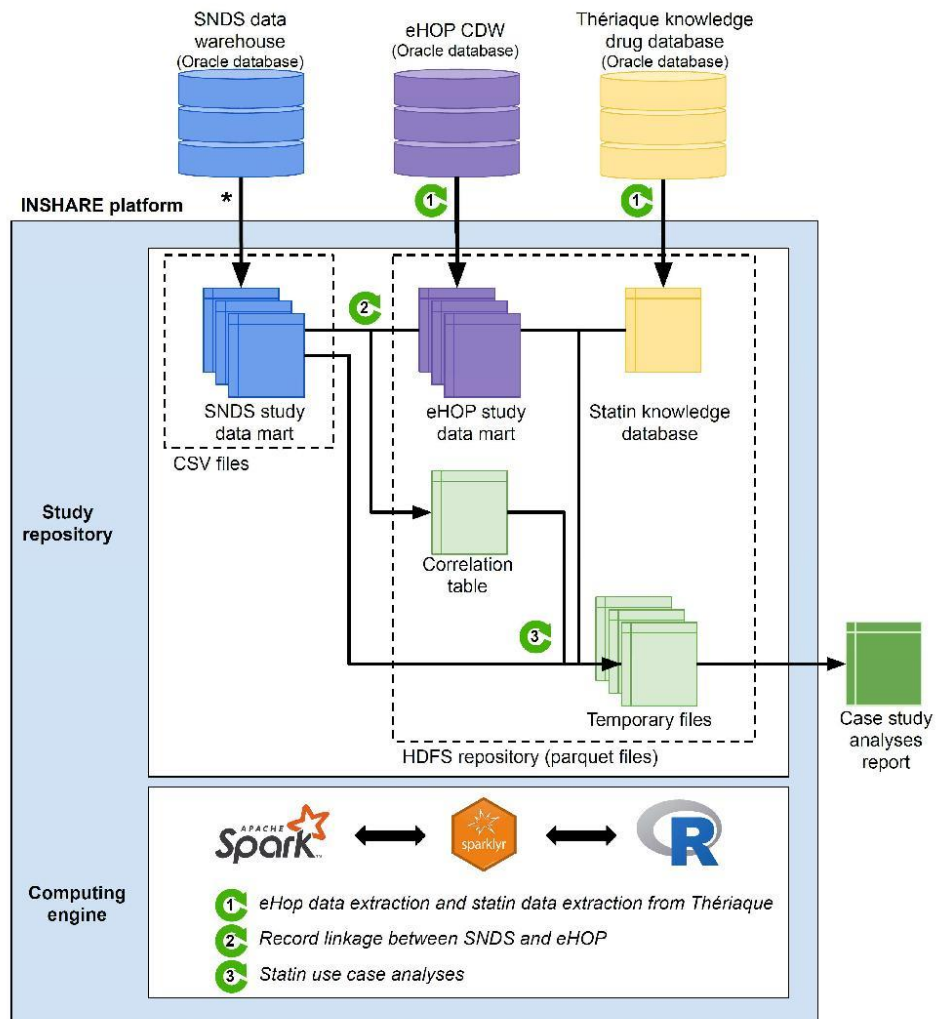


Figure 3. INSHARE platform and data processing workflow.
 * Data were extracted from the national SNDS database by a French national health insurance manager outside of this study workflow.

Ethical consideration

The record linkage and the use case study were approved by the Commission nationale de l'informatique et des libertés (French Data Protection Agency or CNIL, N 2206739). According to the French regulations, patients were informed on the use of their data, and no signed consent was required.

Results

Technological results

INSHARE overall computing performance

The time needed for the record-linkage procedure and the statin use case analysis was 88 minutes. The most time-consuming step was the detection of DDIs in the data of patients taking a statin. The time needed for each step is indicated in Figure 4.

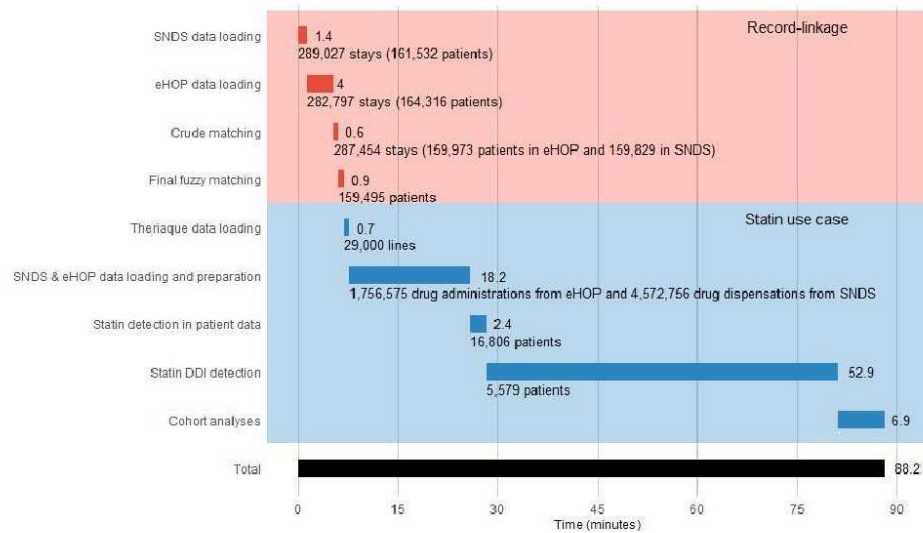


Figure 4. Time duration from data loading to the end of the use case study analysis.

Assessment of the record linkage procedure

The SNDS and eHOP datasets included 161,532 subjects (278,341 stays) and 164,316 subjects (265,089 stays), respectively, who had at least one hospital stay at Rennes academic hospital between 2015 and 2017.

We successfully linked 159,495 patients (98.7% and 97.0% patients from the SNDS and eHOP datasets, respectively). We excluded from the linkage results 199 patients from the SNDS dataset and 162 patients from the eHOP dataset because their records were linked with more than one patient in the other dataset. Patients that could not be linked were

younger (median age of the unmatched patients from the eHOP and SNDS datasets: 22.3 and 27.6 years, respectively, compared with 48.4 years for the matched patients). Moreover, women represented 51% of all matched patients and 57% and 18.5% of unmatched patients in the eHOP and SNDS datasets, respectively.

Use case results

Statin-taking population

Among the 159,495 matched patients, we retained 16,806 patients with at least one statin delivery. Specifically, 8,293 patients started statin treatment before admission and continued during the hospital stay (community and hospital consumption), 6,382 patients started statin treatment before admission and stopped at hospital admission (only community consumption), and 2,131 patients started statins in hospital (hospital initiation). The characteristics of the three subgroups are summarized in Table 1. Age (72.9% and 75.4% of ≥ 65 -year-old patients) and unplanned hospitalization rate (37.9% and 29.4%) were similar in patients with only community consumption and patients with community and hospital consumption. Type of hospital care was similar in patients with community and hospital consumption and in patients with hospital initiation (57.0% and 50.3% of surgery). The percentage of ≥ 65 -year-old patients and rate of unplanned hospitalizations were lower in patients with hospital initiation than in the other two subgroups.

The most dispensed statin was atorvastatin in all three subgroups. Simvastatin, rosuvastatin and pravastatin represented approximately 20% of prescriptions/each in patients with only community consumption. In hospital, only two statins were available (atorvastatin and pravastatin).

Statin-drug interaction detection

We identified 5579 patients with potential statin-related DDIs. Overall, statin-drug interactions occurred more frequently during hospitalization than in the community (36.4% and 22.2%, respectively). The most severe DDIs (level 1) concerned 0.87% of hospitalized patients.

Table 2 presents the hospital outcomes in patients with and without statin-drug interactions. Patients with statin-drug interactions were divided in three subgroups according to the place of DDI occurrence: i) during community consumption (regardless of their hospital consumption), ii) during hospital consumption (regardless of their community consumption), or iii) during both community and hospital consumption. Statin-drug interactions occurring in hospital were associated with longer hospital stay, more severe pathology, and higher in-hospital mortality. The logistic regression model identified characteristics that were significantly related to the occurrence of statin-drug interactions: men older than 64 years of age, admitted for medical care for severe pathology, and longer length of hospital stay.

Tables 4 and 5 present the frequency of patients according to their DDI severity and to the place of DDI occurrence and the details of the five most frequent drugs that interacted with statins according to the place of DDI occurrence.

Link between statin-drug interaction and laboratory results

Figure 5 illustrates the link between the five most frequent drug interactions of each statin and the laboratory results. Overall, we observed little variations in laboratory values

between patients with level 3 statin-drug interactions and patients without statins or taking statins but without DDI. However, glycaemia was higher in patients in whom a potential statin interaction (level 1) with sodium fusidate, itraconazole, or erythromycin was detected. Similarly, kalaemia and liver enzymes (ALT, AST) were altered in patients with a potential statin interaction (level 1) with itraconazole, or sodium fusidate, and with itraconazole, respectively. However, the sample sizes were too small (fewer than 20 patients for most laboratory data, particularly for CPK) to detect any significant variation.

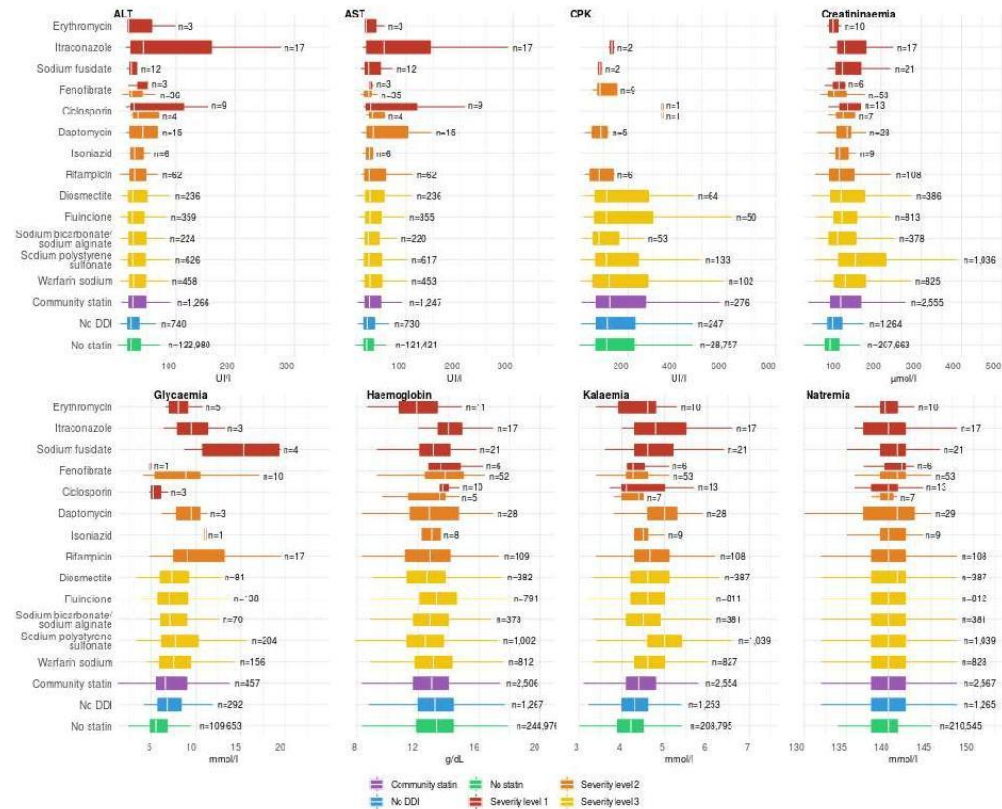


Figure 5. Boxplots of laboratory results for the top 5 DDIs of each statin. The three control groups are depicted in purple, blue and green. Boxplots in yellow, orange and red indicate the laboratory results of patients exposed to statin-related DDI with a level of severity of 3, 2 and 1 (the most severe). Patients can have more than one DDI and of different severity. Fenofibrate and ciclosporin have two boxplots because some of their DDIs are classified as level 2 and others as level 1

Table 1. Patients' characteristics according to their statin consumption

	Only community consumption n=6,382	Community and hospital consumption n=8,293	Hospital initiation n=2,131	P-value
	n (%)	n (%)	n (%)	
Sex, men	3,790 (59.4)	5,431 (65.5)	1,437 (67.4)	<.001
≥65-year-old patients	4,651 (72.9)	6,255 (75.4)	1,192 (55.9)	<.001
Unscheduled admission	2,416 (37.9)	2,434 (29.4)	1,155 (54.2)	<.001
Type of care				<.001
medical care	4,576 (71.7)	3,564 (43.0)	1,059 (49.7)	
surgery	1,806 (28.3)	4,729 (57.0)	1,072 (50.3)	
Chronic statin consumption (>3 months)	6,075 (95.2)	7,660 (92.4)	-	<.001
Statin type				<.001
atorvastatin	2,380 (37.3)	4,632 (55.9)	1,909 (89.6)	
fluvastatin	194 (3.0)	190 (2.3)	3 (0.1)	
pravastatin	1,374 (21.5)	2,004 (24.2)	183 (8.6)	
rosuvastatin	1,145 (17.9)	1,473 (17.8)	24 (1.1)	
simvastatin	1,301 (20.4)	1,540 (18.6)	24 (1.1)	
Patients with statin-drug interaction during				
community consumption	1,438 (22.5)	1,815 (21.9)		<.001
DDI severity				.069
1	30 (2.1)	20 (1.1)		
2	20 (1.4)	29 (1.6)		
3	1,404 (96.6)	1,784 (97.3)		
hospital consumption		3,215 (38.8)	585 (27.5)	<.001
DDI severity				<.001
1		72 (2.1)	10 (1.6)	
2		143 (4.2)	58 (9.4)	
3		3,154 (93.6)	552 (89.0)	

Table 2. Characteristics of patients and hospital stays according to the place of the statin-drug interaction occurrence

	Interaction only during community consumption n=1,779	Interaction only during hospital consumption n=2,326	Interaction during community and also hospital consumption n=1,474	No interaction n=11,227	P-value
Sex, men, n (%)	1,132 (63.6)	1,521 (65.4)	1,008 (68.4)	6,997 (62.3)	<.001
≥65-year-old patients, n (%)	1,394 (78.4)	1,750 (75.2)	1,215 (82.4)	7,739 (68.9)	<.001
Unscheduled admission, n (%)	698 (39.2)	926 (39.8)	544 (36.9)	3,837 (34.2)	<.001
Type of care					<.001
medical care, n (%)	1,233 (69.3)	1,149 (49.4)	787 (53.4)	6,030 (53.7)	
surgery, n (%)	546 (30.7)	1,177 (50.6)	687 (46.6)	5,197 (46.3)	
Length of stay (days), mean (SD)	8.3 (10.4)	11.9 (15.8)	8.4 (10.1)	7.6 (8.2)	<.001
Intensive care unit admission, n (%)	94 (5.3)	742 (31.9)	260 (17.6)	2,156 (19.2)	<.001
DRG severity					<.001
1 (least severe), n (%)	1,314 (73.9)	692 (29.8)	565 (38.3)	7,070 (63.0)	
2, n (%)	177 (9.9)	673 (28.9)	388 (26.3)	2,015 (17.9)	
3, n (%)	197 (11.1)	729 (31.3)	396 (26.9)	1,692 (15.1)	
4 (most severe), n (%)	91 (5.1)	232 (10.0)	125 (8.5)	450 (4.0)	
In-hospital mortality, n (%)	12 (0.67)	24 (1.03)	31 (2.1)	87 (0.78)	<.001

Table 3. Factors related to the occurrence of a statin-interaction

	OR	95% CI
Sex, men	1.14	1.04-1.25
≥65-year-old patients	1.48	1.34-1.62
Unscheduled admission	1.08	0.97-1.19
Medical care	1.56	1.43-1.71
Length of stay (days)	1.03	1.03-1.04
DRG severity		
1 (least severe)	1	
2	1.18	1.06-1.31
3	1.27	1.13-1.43
4 (most severe)	1.51	1.22-1.86

Table 3. Top 5 drugs interacting with statins during community consumption.

	Drug/statin	Rosuvastatin	Simvastatin	Atorvastatin	Pravastatin	Fluvastatin
Severity level: 1 (most severe)		n=25	n=19	n=4	n=2	n=0
T O P 5	Ciclosporin (n=29)	17	12			
	Sodium fusidate (n=8)	1	3	2	2	
	Fenofibrate (n=6)	6				
	Telithromycin (n=3)		1	2		
	Clarithromycin (n=3)		3			
Severity level: 2		n=7	n=23	n=11	n=3	n=6
T O P 5	Fenofibrate (n=23)	6	3	6	3	5
	Carbamazepine (n=15)		15			
	Ciclosporin (n=6)		4	2		
	Rifampicin (n=2)		1	1		
	Bezafibrate (n=2)	1				1
Severity level: 3 (least severe)		n=588	n=1,256	n=1,643	n=755	n=83
T O P 5	Fluindione (n=898)	156	164	365	194	19
	Warfarin sodium (n=674)	114	101	287	152	20
	Amlodipine besylate (n=322)		322			
	Sodium bicarbonate/sodium alginate ^a (n=285)	51	58	110	57	9
	Sodium polystyrene sulfonate (n=225)	38		102	52	

^asodium bicarbonate-containing antacid

Table 4. Top 5 drugs interacting with statins during hospital consumption.

	Drug/statin	Rosuvastatin	Simvastatin	Atorvastatin	Pravastatin	Fluvastatin
	Severity level: 1 (most severe)	n=16	n=29	n=34	n=5	n=0
T O P 5	Sodium fusidate (n=21)	3		14	4	
	Itraconazole (n=19)		2	17		
	Ciclosporin (n=15)	6	9			
	Erythromycin (n=12)		12			
	Fenofibrate (n=7)	7				
	Severity level: 2	n=12	n=27	n=188	n=13	n=2
T O P 5	Rifampicin (n=118)		11	107		
	Fenofibrate (n=56)	7	3	36	8	2
	Daptomycin (n=30)	5	3	19	3	
	Isoniazid (n=9)			9		
	Ciclosporin (n=8)		5	3		
	Severity level: 3 (least severe)	n=508	n=1,140	n=3,531	n=1,204	n=79
T O P 5	Sodium polystyrene sulfonate (n=1,142)	100	98	660	243	13
	Warfarin sodium (n=894)	83	65	529	206	11
	Fluindione (n=894)	92	88	504	196	14
	Diosmectite (n=431)	42	31	259	93	6
	Sodium bicarbonate/ sodium alginate ^a (n=421)	39	39	242	93	8

^asodium bicarbonate-containing antacid

Discussion

Strengths

Technical work

To the best of our knowledge, this is the first study that successfully linked EHR data, through a CDW, and claim data. However, there are some initiatives that integrate the two data types at the source into a common database [26,27].

The linkage process was efficient and is generic enough to be applied to any data source that contains PMSI data. Our goal was to demonstrate that for data reuse purposes, it is possible to link fine-grained EHR data and claim data, without a common patient identifier. Today, most hospitals have a CDW dedicated to research and fed with EHR data. Specifically, we used the eHOP CDW architecture that is currently the most widespread CDW type in France [13].

These two data sources can be bulky. For instance, the statin use case required to read and to filter all drug administrations (n=13,125,574) and all drug dispensations (n=6,019,432) to identify patients to be included in the study. To ensure fast computation, we developed a computing framework based on Spark and HDFS that showed good performances even on our small single-node cluster. These tools are widespread in the big data field, but they are still rarely used for data reuse in hospitals. According to Dolezel et al., their underuse, despite the massive amount of hospital data available, is explained by the lack of personnel with specific technological skills [28].

DDI use case

Our use case study found a statin-drug interaction prevalence of 22.2% and 36.4%, during community consumption and during hospital consumption, respectively. Few studies have provided statin-drug interaction rates during primary care and hospital care for the same population. A Bulgarian study [29] reported statin-drug interaction prevalence rates of 26.6% at hospital admission (used as a proxy of primary care prescription) and 24.4% at discharge. Regarding primary care, this rate ranges from 6.9% [30] to 33% in a systematic review [31] on elderly patients. However, the definition of interaction varies among studies. This could be explained not only by the choice of drug database, as reported in the literature [32,33], but also by the focus on the most severe interactions. Our study took into account different severity levels, from precaution of use to contraindication, using the Thériaque database.

By comparing the places where interactions occurred (community and/or hospital), our study showed that the most severe interactions in hospital led to more specialized and longer care, as previously reported [34]. This should be put in perspective with the larger number and types of drugs administered during hospital stays. Finally, we attempted to link DDIs and laboratory results and showed their potential impact on some laboratory parameters. Previous works reported the biological effects of some statin-drug interactions, such as i) liver toxicity (elevated ALT/AST) by the interaction with cyclosporin, ii) hyperkalaemia [34] with itraconazole or erythromycin [35], and iii) hyperglycaemia with fusidic acid [36]. These findings

should be interpreted with caution because some of them could be due to the adverse effects of statins [37] or of the other drug, such as itraconazole.

Limitations

Technical work

The pairing procedure showed that the data lifecycle introduced quality defects that explain the incomplete record linkage. We are still investigating the reasons of the match failures and how we can explain quality data defects. The record linkage procedure could be improved using more sophisticated linkage strategies, such as probabilistic methods. However, our study concerned a specific case where data variables used for the record linkage procedure originated from the same source (i.e. PMSI data produced by hospitals). Most of the unmatched patients were twins who cannot be distinguished in the SNDS data, even using more complex methods. We think that the deterministic approach is simpler to maintain and is more understandable for people who would like to use or to adapt our algorithm to their own purpose.

DDI use case

DDI prevalence remains dependent on the chosen definition. In our study, these interactions were based only on the simple presence of a drug that can interact with statins and did not capture dose-dependency or patient-specific factors that may influence DDI definitions. Moreover, only information on dispensation was available for primary care (community consumption), while administrations were considered for hospital stay.

Despite the large cohort of patients over a 3-year period, our use case study found only 121 patients with a severity level 1 DDI among whom only five had CPK data. This highlights the important sample size needed in pharmacoepidemiology and pharmacovigilance studies to detect rare adverse effects.

Conclusions

This study demonstrates the added value of combining and reusing clinical and claim data to provide large-scale measures of DDI prevalence and care pathways outside hospitals. In a complex healthcare system that involves multiple care providers, transitions of care are often the source of medication discrepancies and DDIs [38]. Linking CDW and community data is a promising tool to identify where gaps are occurring within the system.

Our approach also allows performing big data-driven analyses to generate new hypotheses. For instance, by linking laboratory data with DDIs, we demonstrated that our strategy allows exploring potential biological variations associated with DDI exposure. However, due to the small patient samples with laboratory results and the exploratory design of the study, we did not want to infer any causal effect or clinical impact at this step. In this context, data reuse should be complementary to hypothesis-driven pharmaco-epidemiological research, which is the appropriate way to confirm the plausibility of a given hypothesis generated using health data.

This builds the path to progress towards a Learning Health System, in which patient care is continuously improved using knowledge generated from research on real world health data and clinical research [39].

Since the INSHARE project, we have extended this approach in the HUGOSHARE project in which we plan to analyse, using the Health Data Hub platform [40], the DDIs for a larger number of drug classes in a much bigger dataset from SNDS and from the CDWs of six academic hospitals of the French western area. This may overcome the limitations of the present study concerning the limited sample sizes for rare events with the aim to generate high quality hypotheses and to consider building predictive models.

Future medical technological developments may also consider enriching community pharmacy reimbursement data with other community data, such as community laboratory results or ambulatory visits. This might enable researchers to identify system vulnerabilities that result in medication errors slipping through the holes of the “Swiss Cheese Model of System Errors” [41,42].

Acknowledgements

We would like to thank Daenna Wung who contributed to this work

Authors' contributions

MC, GB and AB conceived of and designed the study. AB, GB, PL and PVH collected and analysed the data. MC, GB, AB, MB, PLC, EC, XD, and CR interpreted the data. AB and GB drafted the manuscript, and MB, PLC, CR, PL, PVH, EC, XD, and MC critically revised the manuscript. MC obtained funding and supervised the study.

Funding

This work was supported by the French National Research Agency (ANR) for the INSHARE (INtegrating and Sharing Health dAta for Research) project (grant n°. ANR-15-CE19-0024), and by the GIS Epiphare for the record linkage procedure between the SNDS and Hospital clinical data.

Conflicts of Interest

None declared.

References

1. Chazard E, Ficheur G, Caron A, Lamer A, Labreuche J, Cuggia M, Genin M, Bouzille G, Duhamel A. Secondary Use of Healthcare Structured Data: The Challenge of Domain-Knowledge Based Extraction of Features. *Stud Health Technol Inform* 2018;255:15–19. PMID:30306898
2. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. *Yearb Med Inform* 2017 Aug;26(1):38–52. PMID:28480475
3. Safran C. Reuse of clinical data. *Yearb Med Inform* 2014 Aug 15;9:52–54. PMID:25123722

4. Pacheco AG, Saraceni V, Tuboi SH, Moulton LH, Chaisson RE, Cavalcante SC, Durovni B, Faulhaber JC, Golub JE, King B, Schechter M, Harrison LH. Validation of a hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency virus-infected persons and mortality databases in Brazil. *Am J Epidemiol* 2008 Dec 1;168(11):1326–1332. PMID:18849301
5. Asher J, Resnick D, Brite J, Brackbill R, Cone J. An Introduction to Probabilistic Record Linkage with a Focus on Linkage Processing for WTC Registries. *Int J Environ Res Public Health* 2020 Sep 22;17(18):E6937. PMID:32972036
6. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. *Int J Epidemiol* 2016 Jun;45(3):954–964. PMID:26686842
7. Oliveira GP de, Bierrenbach AL de S, Camargo KR de, Coeli CM, Pinheiro RS. Accuracy of probabilistic and deterministic record linkage: the case of tuberculosis. *Rev Saude Publica* 2016 Aug 22;50:49. PMID:27556963
8. Oostema JA, Nickles A, Reeves MJ. A Comparison of Probabilistic and Deterministic Match Strategies for Linking Prehospital and in-Hospital Stroke Registry Data. *J Stroke Cerebrovasc Dis Off J Natl Stroke Assoc* 2020 Oct;29(10):105151. PMID:32912531
9. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol* 2011 May;64(5):565–572. PMID:20952162
10. Hagger-Johnson G, Harron K, Goldstein H, Aldridge R, Gilbert R. Probabilistic linkage to enhance deterministic algorithms and reduce data linkage errors in hospital administrative data. *J Innov Health Inform* 2017 Jun 30;24(2):891. PMID:28749318
11. Bosh KA, Coyle JR, Muriithi NW, Ramaswamy C, Zhou W, Brantley AD, Stockman LJ, VanderBusch L, Westheimer EF, Tang T, Green TA, Hall HI. Linking HIV and Viral Hepatitis Surveillance Data: Evaluating a Standard, Deterministic Matching Algorithm Using Data From 6 US Health Jurisdictions. *Am J Epidemiol* 2018 Nov 1;187(11):2415–2422. PMID:30099475
12. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, Roquefeuil L de, Maura G, Caillol H, Tajahmady A, Coste J, Gissot C, Weill A, Fagot-Campagna A. 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. *Wwwem-Premiumcomdatarevues03987620v65sS4S0398762017304315* [Internet] 2017 Sep 15 [cited 2020 Aug 6]; Available from: <https://www.em-premium.com/article/1140905/>

13. Madec J, Bouzillé G, Riou C, Van Hille P, Merour C, Artigny M-L, Delamarre D, Raimbert V, Lemordant P, Cuggia M. eHOP Clinical Data Warehouse: From a Prototype to the Creation of an Inter-Regional Clinical Data Centers Network. *Stud Health Technol Inform* 2019 Aug 21;264:1536–1537. PMID:31438219
14. LOI n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles. 2018-493 Jun 20, 2018.
15. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016 Mar 15;3:160018. PMID:26978244
16. Ferrières J, Bongard V, Dallongeville J, Arveiler D, Cotel D, Haas B, Wagner A, Amouyel P, Ruidavets J-B. Trends in plasma lipids, lipoproteins and dyslipidaemias in French adults, 1996-2007. *Arch Cardiovasc Dis* 2009 Apr;102(4):293–301. PMID:19427606
17. Haute Autorité de Santé. Efficacité et efficacité des hypolipémiants : une analyse centrée sur les statines [Internet]. Saint-Denis La Plaine: HAS; 2010. Available from: https://www.has-sante.fr/jcms/r_1499450/fr/efficacite-et-efficience-des-hypolipemiants-une-analyse-centree-sur-les-statines
18. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, Chapman MJ, De Backer GG, Delgado V, Ference BA, Graham IM, Halliday A, Landmesser U, Mihaylova B, Pedersen TR, Riccardi G, Richter DJ, Sabatine MS, Taskinen M-R, Tokgozoglul L, Wiklund O, ESC Scientific Document Group. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J* 2020 Jan 1;41(1):111–188. PMID:31504418
19. Stock J. Targeting LDL cholesterol: Early treatment is key to population health. *Atherosclerosis* 2020 May;300:37–38. PMID:32253009
20. Law M, Rudnicka AR. Statin safety: a systematic review. *Am J Cardiol* 2006 Apr 17;97(8A):52C-60C. PMID:16581329
21. Säfström E, Jaarsma T, Strömberg A. Continuity and utilization of health and community care in elderly patients with heart failure before and after hospitalization. *BMC Geriatr* 2018 Aug 13;18(1):177. PMID:30103688

22. Wastesson JW, Morin L, Tan ECK, Johnell K. An update on the clinical consequences of polypharmacy in older adults: a narrative review. *Expert Opin Drug Saf* 2018 Dec;17(12):1185–1196. PMID:30540223
23. Leelakanok N, Holcombe AL, Lund BC, Gu X, Schweizer ML. Association between polypharmacy and death: A systematic review and meta-analysis. *J Am Pharm Assoc JPhA* 2017 Dec;57(6):729-738.e10. PMID:28784299
24. Husson M-C. Theriaque: independent-drug database for good use of drugs by health practitioners. *Ann Pharm Fr* 2008 Dec;66(5–6):268–277. PMID:19061726
25. Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, Moore N. The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2017;26(8):954–962. [doi: 10.1002/pds.4233]
26. Olatosi B, Zhang J, Weissman S, Hu J, Haider MR, Li X. Using big data analytics to improve HIV medical care utilisation in South Carolina: A study protocol. *BMJ Open* [Internet] 2019 Jul 19 [cited 2021 Feb 11];9(7). PMID:31326931
27. McCoy RG, Dykhoff HJ, Sangaralingham L, Ross JS, Karaca-Mandic P, Montori VM, Shah ND. Adoption of New Glucose-Lowering Medications in the U.S.—The Case of SGLT2 Inhibitors: Nationwide Cohort Study. *Diabetes Technol Ther* 2019 Dec 1;21(12):702–712. PMID:31418588
28. Dolezel D, McLeod A. Big Data Analytics in Healthcare: Investigating the Diffusion of Innovation. *Perspect Health Inf Manag* [Internet] 2019 Jul 1 [cited 2021 Feb 11];16(Summer). PMID:31423120
29. Zhelyazkova-Savova M, Gancheva S, Sirakova V. Potential statin-drug interactions: prevalence and clinical significance. *SpringerPlus* 2014;3:168. PMID:24790817
30. Rätz Bravo AE, Tchambaz L, Krähenbühl-Melcher A, Hess L, Schlienger RG, Krähenbühl S. Prevalence of potentially severe drug-drug interactions in ambulatory patients with dyslipidaemia receiving HMG-CoA reductase inhibitor therapy. *Drug Saf* 2005;28(3):263–275. PMID:15733030
31. Thai M, Reeve E, Hilmer SN, Qi K, Pearson S-A, Gnjdic D. Prevalence of statin-drug interactions in older people: a systematic review. *Eur J Clin Pharmacol* 2016 May;72(5):513–521. PMID:26790666
32. Fung KW, Kapusnik-Uner J, Cunningham J, Higby-Baker S, Bodenreider O. Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support. *J Am Med Inform Assoc JAMIA* 2017 Jul;24(4):806–812. PMID:28339701

33. Morival C, Westerlynck R, Bouzillé G, Cuggia M, Le Corre P. Prevalence and nature of statin drug-drug interactions in a university hospital by electronic health record mining. *Eur J Clin Pharmacol* 2018;74(4):525–534. [doi: 10.1007/s00228-017-2400-6]
34. Wang J, Chi C-L, St Peter WL, Carlson A, Loth M, Pradhan PM, Liang Y, Chen W-Y, Lenskaia T, Robinson JG, Adam TJ. A Population-Based Study of Simvastatin Drug-Drug Interactions in Cardiovascular Disease Patients. *AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci* 2020;2020:664–673. PMID:32477689
35. Li DQ, Kim R, McArthur E, Fleet JL, Bailey DG, Juurlink D, Shariff SZ, Gomes T, Mamdani M, Gandhi S, Dixon S, Garg AX. Risk of adverse events among older adults following co-prescription of clarithromycin and statins not metabolized by cytochrome P450 3A4. *CMAJ Can Med Assoc J J Assoc Medicale Can* 2015 Feb 17;187(3):174–180. PMID:25534598
36. Eng H, Scialis RJ, Rotter CJ, Lin J, Lazzaro S, Varma MV, Di L, Feng B, West M, Kalgutkar AS. The Antimicrobial Agent Fusidic Acid Inhibits Organic Anion Transporting Polypeptide-Mediated Hepatic Clearance and May Potentiate Statin-Induced Myopathy. *Drug Metab Dispos Biol Fate Chem* 2016 May;44(5):692–699. PMID:26888941
37. Kim J, Lee HS, Lee K-Y. Effect of statins on fasting glucose in non-diabetic individuals: nationwide population-based health examination in Korea. *Cardiovasc Diabetol* 2018 Dec 5;17(1):155. PMID:30518364
38. Colombo F, Nunnari P, Ceccarelli G, Romano AV, Barbieri P, Scaglione F. Measures of Drug Prescribing at Care Transitions in an Internal Medicine Unit. *J Clin Pharmacol* 2018 Sep;58(9):1171–1183. PMID:29723431
39. Platt JE, Raj M, Wienroth M. An Analysis of the Learning Health System in Its First Decade in Practice: Scoping Review. *J Med Internet Res* 2020 Mar 19;22(3):e17026. PMID:32191214
40. Cuggia M, Combes S. The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare. *Yearb Med Inform* 2019 Aug;28(1):195–202. PMID:31419832
41. Yao B, Kang H, Wang J, Zhou S, Gong Y. Toward Reporting Support and Quality Assessment for Learning from Reporting: A Necessary Data Elements Model for Narrative Medication Error Reports. *AMIA Annu Symp Proc AMIA Symp* 2018;2018:1581–1590. PMID:30815204
42. Wiegmann DA, J Wood L, N Cohen T, Shappell SA. Understanding the “Swiss Cheese Model” and Its Application to Patient Safety. *J Patient Saf* 2021 Apr 14; PMID:33852542

Appendix

Step 1: to create SNDS_PMSI table

```

For each year
T_MCOXXD
  group by (RSA_NUM,ETA_NUM)
  substr(ASS_DGN,4)
  sorted array
  concat all _ASS_DGN
T_MCOXXA
  group by (RSA_NUM,ETA_NUM)
  substr(ASS_DNG,4)
  sorted array
  concat all _CDC_ACT
create table SNDS_PMSI_XR
  select * from T_MCOXXC
  left join T_MCOXXB, T_MCOXXA, T_MCOXXD
  on (ETA_NUM, RSA_NUM)
union all years
filter ETA_NUM= " Id Rennes teaching hospital "
left join IR_BEN_R
  on NUM_ENQ
  WHEN (all _ASS_DGN or all _CDC_ACT IS NULL) THEN ". "
  save SNDS_PMSI
  
```

Step 2 : to create EHOP_PMSI table

```

EHOP_DAS
  group by (ID_SEJ)
  substr(DAS,4)
  sorted array
  concat all_DAS
EHOP_ACTE
  group by (ID_SEJ)
  substr(DAS,4)
  sorted array
  concat all_CCAM
create table EHOP_PMSI
  select * from EHOP_FIXE
  anti join EHOP_SEANCE
  on ID_SEJ
  left join EHOP_PAT
  on ID_PAT
  left join EHOP_ACTE, EHOP_DAS
  on ID_SEJ
  WHEN (all _DAS or all _CCAM IS NULL) THEN ". "
  save EHOP_PMSI
  
```

Step 3 : to delete twins data from SNDS

```

create TAB_TWIN
TAB_CORRES
  group by NUM_ENQ
  count NUM_ENQ_ANO+nb_ID
  filter(nb_ID=1)
create TAB_without_twins
TAB_CORRES
anti join TAB_TWINS
  
```

Step 4 : to perform record linkage

```

Create table crude_matching
select* from SNDS_PMSI
full join EHOP_PMSI
  on BEN_COD_SEX=SEXE,
  BEN_NAI_ANN=ANNEE_NAIS
  BEN_NAI_MOI=MOI_NAIS
  ENT_DAT=DATE_ENTREE
  SOR_DAT=DATE_SORTIE
  DGRN_PAL=DP
  substr(BDI_CCD, 1, 2) = substr(CP, 1, 2)
  POI_NAI=POIDS
group by ID_SEJ
calculate 2 Levenshtein's distances
  one between all_CCAM and all_CDC_ACT = dist_CCAM
  one between all_ASS_DGN and all_DAS = dist_DAS
inner join TAB_without_twins
alter table
  if dist_CCAM+dist_DAS=0 THEN EXACT_match
  if dist_CCAM < 5 THEN FUZZY_CCAM
  if dist_DAS < 5 THEN FUZZY_DAS
filter(EXACT_match OR FUZZY_CCAM OR FUZZY_DAS)
save crude_matching
ID_SNIIRAM with max(nb_sej_FUZZY_CCAM)
filter
else
ID_SNIIRAM with max(nb_sej_FUZZY_DAS)
save final_matching
  
```

Supplementary Fig 1. Pseudocode describing the different steps of the record-linkage algorithm.

Supplementary Table 1. Distributions of Levenshtein distances for procedure and diagnostic codes.

Levenshtein distance	Procedure codes	Diagnostic codes
0	287283	276952
1	2	1
2	44	6
3		23
4	10	516
5	3	9243
6	2	12
7	21	11
8	46	5
9	2	52
10	2	541
11	2	2
12		13
13	2	5
14		7
15	15	30
16	2	5
17		3
18	2	
19	3	3
20		4
21	2	2
22	2	2
23	3	
24		7
25	2	2
29		2
30		3
39	4	
45		2

3.4.4 Synthèse de l'étude 3

Cette étude a fait la preuve de la faisabilité d'un appariement entre des données du SNDS et des données d'entrepôt hospitalier sur un grand nombre de patients (plus de 160 000) en absence du NIR avec des taux de succès supérieurs à 97%.

Cet appariement semi-probabiliste a utilisé les données du PMSI présentes dans ces 2 sources. Elle a mis en évidence l'intérêt de l'enrichissement des données issues du SNDS par des données cliniques et biologiques intra-hospitalières. En effet, les données médicamenteuses étant quasiment absentes des données issues du PMSI, aucune interaction médicamenteuse ne pouvait y être retrouvée. L'enrichissement a également bénéficié aux données issues de l'entrepôt hospitalier. En effet, bien qu'offrant des données plus détaillées sur la partie hospitalière du parcours de soins (résultats cliniques, biologiques, d'imagerie...), le parcours de soins extra-hospitalier y est absent. L'antériorité des traitements, par exemple, ne peut être exploré.

Le cas d'usage sur des patients consommateurs de statines a permis de mettre en lumière des profils de séjours hospitaliers plus lourds lors de la survenue d'une interaction entre les statines et une autre molécule qu'en absence d'interaction. Le chaînage de données de consommation en ville, de données d'administration médicamenteuse à l'hôpital, de données cliniques et biologiques offre un cadre d'étude prometteur en pharmacovigilance et pharmaco-épidémiologie. L'extension de ce projet INSHARE à un réseau d'entrepôt hospitalier dans le cadre du projet HUGOSHARE mené par le groupement HUGO permettra de changer d'échelle en augmentant la taille des cohortes de patients concernés et de prendre en compte la diversité des pratiques hospitalières.

3.5 Synthèse de l'axe 1

Dans cette première partie, au travers de trois approches méthodologiques différentes, nous avons pu montrer qu'il était possible de proposer des outils pouvant répondre à certaines problématiques soulevées par la nature des données issues de base médico-administratives : la prise en compte de facteurs de confusion telles que les comorbidités, l'absence de données concernant le recours aux soins telle que la notion de programmation d'une hospitalisation, le manque de données cliniques et biologiques. Nous avons pu proposer des outils méthodologiques pour pallier les insuffisances liées à la nature de ce type de données : l'adaptation et la validation du score de comorbidités de Charlson aux données du SNDS, la construction d'un proxy pour identifier les hospitalisations non programmées des patients atteints de MICI dans le PMSI, le développement et l'évaluation d'un algorithme d'appariement semi-déterministe permettant de chaîner des données de consommation de ville et hospitalière.

Au travers d'un cas d'usage en pharmacovigilance l'étude INSHARE illustre comment les données issues de bases médico-administratives chaînées à des données hospitalières pouvaient contribuer à identifier des profils de patients présentant une interaction avec leur traitement par statines. Dans une seconde partie, nous nous sommes attachés à illustrer, au travers de 3 autres cas d'usage, l'intérêt des données issues des bases médico-administratives pour reconstituer des parcours de soins pour mener des études en pharmacovigilance et en épidémiologie.

4 Axe 2 : illustrer, au travers de cas d'usage, l'apport de l'utilisation des données médico-administratives en épidémiologie et en pharmacovigilance

Les travaux présentés visent à montrer l'intérêt de l'utilisation des données médico-administratives en épidémiologie et en pharmacovigilance : une fenêtre d'exposition étendue, un recrutement quasi-exhaustif en particulier dans le cadre de pathologies rares, l'accélération du recrutement de sujets et du recueil de données, notamment en situation de crise sanitaire, la complétude du parcours de soins.

Pour des raisons pratiques, le cas d'usage en pharmacovigilance, réalisé dans l'étude INSHARE, a été traité dans la partie précédente. Toutefois, ce cas d'usage sur les patients consommateurs de statines appartient également à cet axe de recherche illustrant l'apport des données médico-administratives en pharmacovigilance.

4.1 Impact des comorbidités maternelles sur l'issue de grossesse et le devenir de l'enfant : Etude 4

4.1.1 Contexte

Le nombre de femmes enceintes présentant des comorbidités pré-existantes à la grossesse augmente en raison d'une élévation de l'âge maternel et des progrès thérapeutiques. Ces avancées concernent aussi bien la prise en charge des maladies chroniques que les progrès en biologie de la reproduction. L'impact de ces comorbidités sur les issues de de grossesse peut se mesurer soit par pathologie, soit de manière globale au travers d'un score.

Afin de privilégier l'approche globale, notre choix s'est porté sur le score de Charlson, qui permet de se focaliser sur des pathologies pré-existantes à la grossesse et que nous avons préalablement validé sur des données médico-administratives françaises.

L'objectif de cette étude était d'évaluer l'impact des comorbidités maternelles pré-existantes à la grossesse sur l'issue de grossesse et de le devenir de l'enfant à partir des données du PMSI.

4.1.2 Méthodes

Notre étude de cohorte a identifié toutes les femmes de 10 à 55 ans enceintes de 15 SA et plus sorties d'une hospitalisation pour issue de grossesse entre 2016 et 2018 dans la base nationale du PMSI MCO en France métropolitaine. Les séjours pour la prise en charge d'une grossesse multiple ont été exclus et seule la première issue de grossesse pour chaque femme a été incluse.

La présence d'une comorbidité définie dans le score de Charlson a été recherchée dans les 6 ans précédant la date de conception dans les codes de diagnostic CIM10 des hospitalisations en s'appuyant la méthodologie présentée dans l'étude d'adaptation et de validation du score de Charlson dans le SNDS (63). Le critère de jugement principal était une issue de grossesse défavorable définie par la survenue du premier des événements suivants : fausse couche tardive, IMG, prématurité, mort-nés, décès de l'enfant dans la première semaine de vie. Les antécédents obstétricaux examinés dans les 6 ans précédents la conception ont été regroupés selon 3 modalités : absence d'antécédent de grossesse de de 15 SA et plus, au moins un antécédent de grossesse de 15 SA avec une issue défavorable, uniquement des grossesses de 15 SA et plus avec une issue favorable.

Un modèle de poisson a évalué l'association entre la présence d'au moins une comorbidité pré-existante à la grossesse et la survenue d'une issue défavorable de grossesse.

4.1.3 Résultats

Parmi les 1 871 647 femmes incluses, 47 720 (2,5 %) présentaient au moins une comorbidité du score de Charlson dont les plus fréquentes étaient les pathologies broncho-pulmonaires (dont l'asthme), le diabète, les cancers et les pathologies hépatiques peu sévères.

Une issue défavorable de grossesse est survenue chez 6,6 % des femmes avec, au premier rang, un accouchement prématuré (73,2 %). L'analyse bivariée a retrouvé une association significative entre la présence d'une comorbidité et une issue défavorable (RR = 1,83, 95 %IC = [1,79 - 1,88]). Dans l'analyse multivariée, une interaction entre les antécédents obstétricaux et la présence d'une comorbidité a été retrouvée de façon significative. Un modèle pour chaque modalité d'antécédents obstétricaux a été

réalisé. L'impact d'une comorbidité était estimé via un RR de 1,88 (95 %IC =[1,82-1,95]) pour les femmes sans antécédent de grossesse de 15 SA et plus dans les 6 ans, jusqu'à un RR de 1,60 (95%IC = [1,51 - 1,70]) pour les femmes avec uniquement des antécédents favorables. Une analyse exploratoire de l'effet du score de Charlson en lui-même a mis en évidence une association significative avec une issue de grossesse défavorable.

Impact of maternal Charlson index's comorbidities on pregnancy outcome: a population-based cohort study

Aurélie Bannay

Affiliation: Université de Lorraine, CHRU-Nancy, Medical information department, F-54000 Nancy,
France

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

Address: CHRU de Nancy, Hôpital Saint Julien DIM, 1 rue Foller CO60034 54035 NANCY Cedex
FRANCE

Vianney Guardiolle

Affiliation: Université de Lorraine, CHRU-Nancy, Medical information department, F-54000 Nancy,
France

Address: CHRU de Nancy, Hôpital Saint Julien DIM, 1 rue Foller CO60034 54035 NANCY Cedex
FRANCE

Sylvain Baillot

Affiliation: CHRU-Nancy, Medical information department, F-54000 Nancy, France

Address: CHRU de Nancy, Hôpital Saint Julien DIM, 1 rue Foller CO60034 54035 NANCY Cedex
FRANCE

Nicolas Jay

Affiliation: Université de Lorraine, CHRU-Nancy, Medical information department, F-54000 Nancy,
France

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

Address: CHRU de Nancy, Hôpital Saint Julien DIM, 1 rue Foller CO60034 54035 NANCY Cedex
FRANCE

Emilie Gauchotte

Affiliation: CHRU-Nancy, Obstetrics and Fetal Medicine, F-54000 Nancy, France

Address: CHRU de Nancy, Maternité régionale, 10 Avenue Dr Heydenreich, CO60034 54035 NANCY
Cedex FRANCE

Catherine Deneux-Tharaux MD, PhD

Affiliation: INSERM U1153, Obstetrical, Perinatal and Paediatric Epidemiology Research Team
(EPOPé), Research Center for Epidemiology and Biostatistics (CRESS), Université de Paris, Paris,
France

Address: Inserm U1153, équipe EPOPé, épidémiologie obstétricale périnatale et pédiatrique,
maternité Port-Royal, 53, avenue de l'Observatoire, 75014 Paris, France

Jeanne Fresson MD, MSc

Affiliation: CHRU-Nancy, Medical information department, F-54000 Nancy, France

INSERM U1153, Obstetrical, Perinatal and Paediatric Epidemiology Research Team (EPOPé), Research
Center for Epidemiology and Biostatistics (CRESS), Université de Paris, Paris, France

Address: CHRU de Nancy, Maternité régionale, 10 Avenue Dr Heydenreich, CO60034 54035 NANCY
Cedex FRANCE

Corresponding author: Aurelie Bannay

Service d'Évaluation et d'Information Médicales-Département d'Information Médicale

CHRU de Nancy

Hôpital Saint Julien

1, Rue Foller CO60034

54035 NANCY Cedex

Tel +33 3 83 85 18 39

Email aurelie.bannay@univ-lorraine.fr

Running title: Effect of maternal comorbidity on pregnancy outcome

Abstract:

Objective: To assess the impact of pre-existing maternal comorbid conditions, identified using the Charlson comorbidity index, on pregnancy outcome using data extracted from a nationwide hospital discharge database.

Design: Population-based cohort study

Setting: France

Population: all 10 to 55-year-old women with at least one hospital stay due to singleton pregnancy outcome at ≥ 15 weeks of gestation.

Methods: All claim data for hospital stays during the 6 years before the conception date were extracted. Women were considered as having a comorbidity if at least one ICD-10 code for a condition included in the Charlson comorbidity index was found in one of their previous hospital stays. Poisson regression models were used to estimate the risk ratios (RR) with 95% confidence intervals (CI) of an adverse pregnancy outcome in exposed relative to non-exposed women.

Main Outcome Measures: An adverse pregnancy composite outcome: abortion/miscarriage, termination of pregnancy, stillbirth, preterm birth, or death of a full-term new-born during the first 7 days of life.

Results: Among the 1 871 467 included women, 6.6% had an adverse pregnancy outcome, and 2.5% had at least one comorbidity included in the Charlson comorbidity index. Maternal pre-existing chronic comorbidities increased the risk of adverse pregnancy outcome (>15 gestational weeks) by 88% (adjusted RR=1.88 95% CI=[1.82-1.95]) for women without previous pregnancies and by 60% (RR=1.60 95% CI=[1.51-1.7]) for women with childbearing history.

Conclusions: These findings stress the importance of taking into account maternal comorbid conditions to evaluate the foetal outcome in perinatal healthcare studies.

Funding: None

Keywords: Comorbidity, pregnancy outcome, perinatal epidemiology

Tweetable abstract

Maternal comorbidity increases the risk of an adverse pregnancy outcome up to 88%. It must be taken into account in perinatal studies.

Introduction

In industrialized countries, the mean maternal age and the number of comorbid conditions (e.g. diabetes mellitus, high blood pressure) are increasing¹. Moreover, advances in assisted reproductive technologies and therapeutic breakthroughs have extended childbearing to women with chronic diseases. The impact of such maternal comorbidities on the pregnancy outcome can be investigated: i) by focusing on each specific pathology, and therefore in a limited number of patients; and ii) by using an overall comorbidity index to obtain a complete picture.

Three comorbidity indices²⁻⁴ have been assessed to predict the maternal outcome⁵. The Charlson comorbidity and the Elixhauser comorbidity scores take into account chronic non-obstetrical pathologies, whereas the Bateman obstetric comorbidity index predicts the maternal outcome using data on chronic diseases and obstetrical complications. In maternal health, several studies^{4,6} reported better performances of the Bateman score. Conversely, the associations between maternal comorbidity index and foetal and neonatal outcome have not been studied yet. Comorbidity indices could be useful in obstetrical practice to identify pregnant women at high risk by appraising their overall comorbidity burden¹. The Charlson index seems more suitable to predict the comorbidity effect before or at the start of pregnancy because it is not focused on obstetrical conditions. Comorbidity indices are also commonly used in observational studies, especially those based on administrative databases, to adjust for confounders.

In France, each hospital stay is recorded in the national hospital discharge database *Programme de médicalisation des systèmes d'information* (PMSI) ⁷. Thanks to the quality of the collected data and the complete coverage of the inpatient and outpatient hospital stays, this database is a key tool for monitoring healthcare indicators ⁸. Deterministic record linkage of all hospital claim data over the years is an asset for building individual healthcare pathways. The linkage of data on new-born and mother hospital stay for delivery offers opportunities for perinatal epidemiology. For instance, in France, perinatal and maternal healthcare monitoring and clinical practice evaluations are based on the analysis of the PMSI database. Moreover, the exhaustiveness of claim data allows assessing rare outcomes in population-based studies.

The aim of this study was to determine the impact of maternal comorbid conditions identified using the Charlson comorbidity index (CCI) on pregnancy outcome.

Material and methods

Data source

All women admitted for a pregnancy outcome between 01/01/2016 and 31/12/2018 were identified in the French national discharge database (PMSI database).

The PMSI database⁷ was established to monitor the activity of public and private hospitals. It provides detailed administrative and medical information on all inpatient and outpatient stays in French public and private hospitals, including the discharge diagnosis based on ICD-10 codes and medical procedures performed during the hospital stay. Each hospital stay is classified in one of the diagnosis-related groups used for the hospital billing system. The PMSI database allows two kinds of linkage: i) linking all hospitalizations of a given patient thanks to the patient's unique identifier, and ii) linking the mother and child's hospital stays for delivery thanks to a common identifier, since 2012.

Study design and participants

Our nationwide population-based study included all 10- to 55-year-old women with one hospital stay due to a pregnancy outcome after at least 15 weeks of pregnancy in mainland France between 2016

and 2018. Multiple pregnancies were excluded because of their specific features and linkage errors. For each woman, only the first hospital stay due to a pregnancy outcome during the study period was retained.

Collected data

For each woman, all hospital claim data for the 6 years before the conception date until the pregnancy outcome-related stay (index stay) were extracted. The following items were collected: maternal age at index stay admission, gestational weeks (GW), last normal menstrual period, delivery type, length of the index stay (for the mother), residence postal code, primary and secondary diagnoses, degree of specialization of the maternity ward (from primary level to referral level).

Other variables were derived from the collected data. Specifically, the conception date was estimated from the gestational age or the start of the last normal menstrual period. Women were divided in three categories according to their pregnancy history: i) no history of previous known pregnancies (at least 15 weeks of gestation), ii) history of at least one pregnancy with a duration of at least 15 weeks and adverse outcome, and iii) history of pregnancy longer than 15 weeks with only favourable outcome. Women were considered exposed if at least one ICD-10 code for a condition included in the CCI was found in one of their hospital stays in the 6 years before the conception date. The primary outcome was an adverse pregnancy composite outcome: miscarriage, pregnancy termination, stillbirth, preterm birth (<37 GW), or death of a full-term new-born during the first 7 days of life.

This study was carried out following the MR-005 methodological guidelines of the French Data Protection Agency. According to French regulations, all data were de-identified, and the signature of an informed consent was not required.

Statistical analysis

Categorical variables were described using numbers and percentages, and quantitative variables using means and their standard deviation (SD). The relationship between women's characteristics and pregnancy adverse outcome was assessed using Poisson regression models adjusted for covariates that were significantly associated in the univariate analysis. When appropriate, 95% confidence intervals (95% CI) were computed. Two sensitivity analyses were performed: i) one with an extended

exposure time until the pregnancy outcome-related stay, and ii) one restricted to women admitted to level I maternity wards. Statistical analyses were performed using SAS Enterprise Guide, version 7.1.

Results

Among the 1 871 467 women with a singleton pregnancy at ≥ 15 GW between 2016 and 2018, 47 720 (2.5%) women had at least one comorbidity listed in the CCI among the diagnoses (ICD-10 codes) of their hospital stay(s) in the 6 years before the conception date. Among the 17 CCI condition categories, the most common were: chronic pulmonary disease including asthma (33.1%), diabetes without complications (16.9%), cancer including leukaemia (9.2%), and mild liver disease (9.0%).

In the whole population, 124 232 women (6.6%) had an adverse pregnancy outcome during the study period (Table 1). Exposed women were older than non-exposed women (31.5% vs 21.8% of ≥ 35 -year-old women), and more likely to have had a previous pregnancy (54.8% vs 39.7%), especially with a previous adverse outcome (7.5% vs 3%). Exposed women were more likely to be admitted to a referral maternity ward (35.8% vs 24.9%) and to have a caesarean delivery (29.7%) than non-exposed women (19.2%). The main adverse pregnancy outcomes were preterm birth (73.2%), abortion or miscarriage, and termination of pregnancy before 22 GW (15.5%) (Table 2). According to their obstetric history, the adverse pregnancy outcome rate was 17.8% in women with at least one previous adverse outcome, 4.7% in women with only previous favourable outcome(s), and 7.3% in women without pregnancy history (Table 2).

Univariate analysis found an association between the occurrence of at least one condition included in the CCI in the 6 years before the conception date and an adverse pregnancy outcome (RR=1.83, 95% CI=[1.79-1.88]). In the sensitivity analysis, the length of the exposure window before the pregnancy outcome-related hospital stay increased the RR value to 1.90 (95% CI=[1.85-1.94]).

The multivariable model adjusted for maternal age and pregnancy history showed a significant interaction between pregnancy history and occurrence of at least one comorbidity. A subgroup analysis according to pregnancy history showed that the association between comorbidity and adverse pregnancy outcome was stronger in the subgroup of women without pregnancy history (adjusted RR=1.88 95% CI=[1.82-1.95]) than in the subgroups with pregnancy history [with history of adverse

outcome (adjusted RR=1.62 95% CI=[1.55-1.70]); with history of favourable outcome (adjusted RR=1.60 95%CI=[1.51-1.70])). In sensitivity analyses, the association of comorbidity and adverse pregnancy outcome remained significant at a lower level for women hospitalized in level I maternity wards (Table 3).

Exploratory analyses with the CCI score used as continuous or categorical variable (Table 4) also found an increased risk of adverse outcome with higher CCI scores, but this relationship seemed non-linear. Indeed, the group of women with the highest CCI score did not show the highest risk.

Discussion

Main findings

To the best of our knowledge, this is the first study assessing the association between global pre-existing maternal comorbidity and foetal and new-born health. Our study reported a significant risk increase by 88% and 60% in women without and with pregnancy history, respectively.

Strength and limitations

This study is based on data extracted from the national French hospital discharge database (PMSI) and therefore its strengths and limitations are mostly linked to the features of this database. Indeed, this nationwide database is almost exhaustive (99% of all deliveries)⁹. Consequently, it offers a big volume of data for studying uncommon conditions that cannot be captured by studies based on the collection of data on specific diseases. The linkage with hospital stays in the 6 years before the conception date gave a large exposure window for comorbidities and obstetric history. However, this 6-year window could have led to misclassify some women. For instance, some women could have been categorized as non-exposed or without history in case of linkage failure or in case of lack of hospitalizations during this period. The prevalence of comorbidities in our study (2.5% of pregnant women) was probably underestimated because the PMSI data allow identifying only women with severe conditions that led to hospitalization. Therefore, women with comorbid conditions that did not require hospitalization were considered as non-exposed in our study. The sensitivity analysis using an extended exposure window until the pregnancy outcome-related stay showed a prevalence of 3.03% (data not shown) of women with a comorbid condition and a similar association with adverse pregnancy outcomes (RR=1.90, 95% CI=[1.85-1.94]) (Table 3).

This hospital discharge database contains only data on hospital stays. In France, women with pregnancy outcome before the 15th GW (i.e. early miscarriage) are not always admitted to hospital. Therefore, to avoid a selection bias, only women with a pregnancy outcome after at least 15 GW were included. The threshold of 22 GW is typically used in the literature; however, we chose a lower cut-off to take into account the comorbidity effect as early as possible after pregnancy initiation. Moreover, this cut-off is above the legal limit of voluntary abortion in France. The database quality was previously validated for perinatal indicators by comparison with French national statistics and a national perinatal survey¹⁰.

Interpretations

Several studies^{4,6} have used comorbidity indices to assess maternal morbidity and mortality. The maternal comorbidity index developed by Bateman includes comorbidity and obstetric complications and showed better performances than other scores⁴. Obstetric complications have a major impact on the child health, but they could be intermediate factors between a pre-existing comorbidity and the pregnancy outcome. As our study was more in the framework of a predictive model to plan pregnancy care than within an explanatory model, it was more relevant to use a general score applicable to all women before pregnancy. The CCI² has been widely used since its description in 1987, and its robustness has been demonstrated. It has been adapted and updated for the ICD-10 classification^{11,12}, and it is used in French administrative databases¹³.

Our findings showed that 2.5% of pregnant women had at least one condition included in the CCI. The most frequent categories (chronic pulmonary disease including asthma, diabetes without complications, cancer, mild liver disease) were expected in a population of older childbearing women. In an exploratory analysis, we found an association between the CCI score and adverse pregnancy outcomes. The relationship seemed to be non-linear and suggested that CCI initial weights could be adapted to predict adverse pregnancy outcomes. In a recent study¹⁴, the CCI was used to assess maternal and paternal health before conception and their changes between 2009 and 2016. The authors took into account the CCI and other conditions, such as obesity, hypertension, and depression, using the database of an American claim company that included hospital stay data and outpatient care data. The prevalence of pre-existing comorbid conditions was higher than in our study (11.3% in 2009 and 18.9% in 2016) because of the specific population and the different pathologies taken into account.

Different individual maternal pathologies have been associated with adverse pregnancy outcome. Recent studies reported relative risks of premature birth of 1.26, 3.51 and 3.81 in women with hepatitis B ⁵, pre-existing diabetes ¹⁵, and chronic hypertension ¹⁵. However, their impact is difficult to assess and they are often not taken into account in epidemiologic studies, especially in international comparisons or in temporal trends of perinatal indicators. Our study shows the need to take into account maternal comorbid conditions as confounding factors to assess foetal and neonatal outcome. General comorbidity scores could be used in the absence of a specific maternal comorbidity score. The use of data from the French complete administrative database SNDS¹⁶ (Système national des données de santé) that includes also primary care consumption could be helpful to validate our findings by taking into account also medication use, disability data, and other pathologies that are known predictors of preterm birth, such as chronic hypertension ¹⁵. A first assessment of data on women who gave birth in 2015 showed its feasibility for identifying pre-existing comorbid conditions ¹⁷. This database could be useful for developing *a specific comorbidity score to predict adverse pregnancy outcome*.

Besides reducing confounders in epidemiologic and health service research, our findings could also have an impact on healthcare organization policies for perinatal care and obstetric practice (17). Since the late 1980s, the regionalization model has been implemented in France for perinatal healthcare, with the aim of adapting the level of care to the maternal and foetal risks (18–20). Our results could be explained by the effect of this strategy that led to the more frequent admission to a referral maternity ward of women with at least one comorbidity compared with women without comorbidity (35.8% vs 24.9%). In our sensitivity analysis restricted to level I maternity wards, the association between comorbidity and adverse pregnancy outcome remained, but with lower strength (from RR=1.32 for women with favourable obstetric history to 1.48 for women without obstetric history). A recent study on maternal health ¹ demonstrated the value of using a comorbidity score to identify mothers at risk of morbidity and mortality. A better early clinical identification of women at high risk and their referral to specialized maternity wards could improve their and their child outcome (21). The early detection of women at high risk could allow a more targeted patient education and might promote adherence to specialized pregnancy monitoring.

Conclusion

Our findings show the interest and the need to take into account maternal pre-existing comorbidities to analyse pregnancy outcome and new-born health. We showed that the CCI could be useful to assess their overall condition. However, a specific tool could be developed and implemented in administrative databases to accurately measure comorbidity before conception and to predict the individual risk for new-borns. Such tool could be very useful also for perinatal epidemiology, obstetric practice, and perinatal healthcare policy evaluations.

Disclosure

All authors report no conflict of interests.

Contribution to authorship

AB, VG, SB, NJ, EG, CTD, and JF contributed to the concept and design of the study. Data and statistical analysis were by SB, VG, AB and JF. AB, JF and VG drafted the manuscript and all authors performed a critical revision of the manuscript. AB, JF and SB had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Details of ethics approval

This study was conducted according to the French Reference Methodology MR-005 regarding access to the programme for the medicalisation of information systems approved by the French National Commission for Data Protection and Liberties (CNIL MR-005 number 2204437 on the 7th August 2018).

Funding

None

References

1. Easter SR, Bateman BT, Sweeney VH, Manganaro K, Lassey SC, Gagne JJ, et al. A comorbidity-based screening tool to predict severe maternal morbidity at the time of delivery. *Am J Obstet Gynecol.* 2019;221:271.e1-271.e10.

2. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40:373–83.
3. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998;36:8–27.
4. Bateman BT, Mhyre JM, Hernandez-Diaz S, Huybrechts KF, Fischer MA, Creanga AA, et al. Development of a comorbidity index for use in obstetric patients. *Obstet Gynecol.* 2013;122:957–65.
5. Liu J, Zhang S, Liu M, Wang Q, Shen H, Zhang Y. Maternal pre-pregnancy infection with hepatitis B virus and the risk of preterm birth: a population-based cohort study. *Lancet Glob Health.* 2017;5:e624–32.
6. Aoyama K, D’Souza R, Inada E, Lapinsky SE, Fowler RA. Measurement properties of comorbidity indices in maternal health research: a systematic review. *BMC Pregnancy and Childbirth.* 2017;17:372.
7. Boudemaghe T, Belhadj I. Data Resource Profile: The French National Uniform Hospital Discharge Data Set Database (PMSI). *Int J Epidemiol.* 2017;46:392–392d.
8. Goueslard K, Cottenet J, Benzenine E, Tubert-Bitter P, Quantin C. Validation study: evaluation of the metrological quality of French hospital data for perinatal algorithms. *BMJ Open.* 2020;10:e035218.
9. Bellamy V. Les 784 000 naissances de 2016 ont eu lieu dans 2 800 communes - Insee Focus - 92 [Internet]. [cited 2020 Dec 14]. Available from: <https://www.insee.fr/fr/statistiques/3047024>
10. Quantin C, Cottenet J, Vuagnat A, Prunet C, Mouquet M-C, Fresson J, et al. Quality of perinatal statistics from hospital discharge data: comparison with civil registration and the 2010 National Perinatal Survey. *J Gynecol Obstet Biol Reprod (Paris).* 2014;43:680–90.
11. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol.* 2011;173:676–82.

12. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–9.
13. Bannay A, Chaignot C, Blotière P-O, Basson M, Weill A, Ricordeau P, et al. The Best Use of the Charlson Comorbidity Index With Electronic Health Care Database to Predict Mortality. *Med Care*. 2016;54:188–94.
14. Sun AJ, Li S, Zhang CA, Jensen TK, Lindahl-Jacobsen R, Eisenberg ML. Parental comorbidity and medication use in the USA: a panel study of 785 000 live births. *Hum Reprod*. 2020;35:669–75.
15. Berger H, Melamed N, Davis BM, Hasan H, Mawjee K, Barrett J, et al. Impact of diabetes, obesity and hypertension on preterm birth: Population-based study. *PLOS ONE*. 2020;15:e0228743.
16. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, Roquefeuil L de, 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. *Revue d'Épidémiologie et de Santé Publique*. 2017; 65 S4:149-167.
17. Piffaretti C, Looten V, Rey S, Fresson J, Fagot-Campagna A, Tuppin P. Management of pregnancy based on healthcare consumption of women who delivered in France in 2015: Contribution of the national health data system (SNDS). *J Gynecol Obstet Hum Reprod*. 2018;47:299–307.
18. Grossesses à risque : orientation des femmes enceintes entre les maternités en vue de l'accouchement [Internet]. Haute Autorité de Santé. [cited 2020 Dec 14]. Available from: https://www.has-sante.fr/jcms/c_935540/fr/grossesses-a-risque-orientation-des-femmes-enceintes-entre-les-maternites-en-vue-de-l-accouchement
19. Suivi et orientation des femmes enceintes en fonction des situations à risque identifiées [Internet]. Haute Autorité de Santé. [cited 2020 Dec 14]. Available from: https://www.has-sante.fr/jcms/c_547976/fr/suivi-et-orientation-des-femmes-enceintes-en-fonction-des-situations-a-risque-identifiees
20. Décret no 98-900 du 9 octobre 1998 relatif aux conditions techniques de fonctionnement auxquelles doivent satisfaire les établissements de santé pour être autorisés à pratiquer les

activités d'obstétrique, de néonatalogie ou de réanimation néonatale et modifiant le code de la santé publique (troisième partie : Décrets) - Légifrance [Internet]. [cited 2020 Dec 14]. Available from: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000756322/>

21. Watson H, McLaren J, Carlisle N, Ratnavel N, Watts T, Zaima A, et al. All the right moves: why in utero transfer is both important for the baby and difficult to achieve and new strategies for change. *F1000Res.* 2020;9:979.

Table 1 Characteristics of the whole sample and according to the presence of at least one condition included in Charlson comorbidity index in the 6 years before the conception date

	Whole population (n= 1 871 467) %		Exposed women (n= 47 720) %		Non- exposed women (n= 1 823 747) %	
Maternal age at conception date						
≤19 years	33 780	1.8	232	0.5	33 548	1.8
20-24 years	218 327	11.7	3 105	6.5	215 222	11.8
25-29 years	574 756	30.7	12 531	26.3	562 225	30.8
30-34 years	632 738	33.8	16 826	35.3	615 912	33.8
35-39 years	330 523	17.7	11 414	23.9	319 109	17.5
≥40 years	81 343	4.3	3 612	7.6	77 731	4.3
Pregnancy history during the last 6 years						
at least one adverse outcome	58 219	3.2	3 570	7.5	54 649	3.0
only favourable outcome(s)	691 319	36.9	22 593	47.3	668 726	36.7
no previous outcome	1 121 929	59.9	21 557	45.2	1 100 372	60.3
Maternity hospital level						
I (primary level)	417 751	22.3	7 947	16.7	409 804	22.5
IIA	544 869	29.1	11 495	24.1	533 374	29.2
IIB	437 258	23.4	11 161	23.4	426 097	23.4
III (referral level)	471 397	25.2	17 104	35.8	454 293	24.9
no maternity hospital	192	0.01	13	0.03	179	0.01

Mother's length of stay						
0-2 days	82 857	4.4	2 157	4.5	80 700	4.4
3-4 days	1 100 239	58.8	23 992	50.3	1 076 247	59.0
5 days and more	688 371	36.8	21 571	45.2	666 800	36.6
Delivery type (n=1 852 207)						
vaginal or extraction	1 492 339	80.6	33 034	70.3	1 459 305	80.8
C-section	359 848	19.4	13 929	29.7	345 919	19.2
Adverse outcome						
abortion/miscarriage	11 097	0.6	463	1,0	10 634	0.6
termination of pregnancy						
15-21 GW	8 183	0.4	294	0.6	7 889	0.4
22 GW and more	5 702	0.3	186	0.4	5 516	0.3
still birth	7 710	0.4	313	0.6	7 397	0.4
preterm birth (<37 GW)						
22-27 GW	3829	0.2	192	0.4	3 637	0.2
including death during the first 7 days of life	960	0.05	49	0.1	911	0.05
28-31 GW	7720	0.41	384	0.8	7 336	0.4
including death during the first 7 days of life	218	0.01	8	0.02	210	0.01
32-36 GW	79356	4.24	3828	8.02	75 528	4.14
including death during the first 7 days of life	279	0.01	10	0.02	269	0.01
death of full-term new-born during the first 7 days of life	635	0.034	24	0.05	611	0.034

Table 2 Adverse pregnancy outcomes according to the obstetric history in the past 6 years

	Obstetric history during the past 6 years					
	at least one previous adverse outcome (n= 58 219)	%	only previous favourable outcome(s) (n= 691 319)	%	no previous outcome (n= 1 121 929)	%
Adverse outcome	10 375	17.8	32 178	4.7	81 679	7.3
abortion/miscarriage	591	1.0	3 628	0.6	6 878	0.6
medical termination of pregnancy						
15-21 GW	361	0.6	2 860	0.4	4 962	0.5
22 GW and more	237	0.4	1 912	0.3	3 553	0.3
still birth	430	0.8	2 158	0.3	5 122	0.5
preterm birth (<37 GW)	8 739	15.0	21 409	3.1	60 757	5.4
22-27 GW	329	0.57	677	0.1	2 823	0.25
including death during the first 7 days of life	84	0.14	176	0.03	700	0.06
28-31 GW	771	1.31	1477	0.21	5472	0.49
including death during the first 7 days of life	30	0.05	49	0.01	139	0.01
32-36 GW	7639	13.12	19255	2.79	52462	4.68
including death during the first 7 days of life	22	0.04	86	0.01	171	0.02
death of full-term new-born during the first 7 days of life	17	0.029	211	0.03 1	407	0.036

Table 3 Multivariable analysis of the association between comorbid conditions and adverse pregnancy outcome

	Crude model		Adjusted model †		Adjusted model † with interaction between pregnancy history and comorbidity occurrence	
	RR	95% CI	RR	95% CI	RR	95% CI
Whole population						
at least one comorbid condition						
in the past 6 years before the conception date	1.83	1.79-1.88	1.75	1.71-1.80	1.62	1.55-1.70
in the past 6 years until the pregnancy outcome-related hospital stay	1.90	1.85-1.94	1.83	1.78-1.87	1.69	1.62-1.77
Subgroups according the pregnancy history						
at least one previous adverse outcome (n= 58 219)	-	-	1.6	1.51-1.70	-	-
only previous favourable outcome(s) (n= 691 319)	-	-	1.62	1.55-1.70	-	-
no previous outcome (n= 1 121 929)	-	-	1.88	1.82-1.95	-	-
Pregnancy outcomes in primary-level maternity hospitals (n= 417 751)						
Whole population						
occurrence of least one comorbid condition						
in the past 6 years before the conception date	1.44	1.32-1.58	1.42	1.30-1.56	1.38	1.18-1.60
in the past 6 years until the pregnancy outcome-related hospital stay	1.49	1.37-1.62	1.47	1.35-1.60	1.40	1.21-1.61
Subgroups according to the pregnancy history						
at least one previous adverse outcome (n= 9 661)	-	-	1.32	1.02-1.70	-	-
only previous favourable outcome(s) (n= 159 082)	-	-	1.38	1.18-1.60	-	-
no previous outcome (n= 249 008)	-	-	1.48	1.31-1.68	-	-

Notes: †for maternal age and pregnancy history in the whole population analyses, and for maternal age in the subgroup analyses

Table 4. Exploratory analysis of the comorbidity assessment and its association with adverse pregnancy outcomes

	Crude model		Adjusted model †		Adjusted model † with interaction between obstetrical history and occurrence of comorbidity	
	RR	95% CI	RR	95% CI	RR	95% CI
Whole population						
continuous score	1.25	1.24-1.27	1.23	1.21-1.24	1.22	1.20-1.25
categorical score						
0 (n=1 823 747)	1		1		1	
1-2 (n=43 804)	1.77	1.72-1.81	1.70	1.65-1.75	1.58	1.51-1.66
3-4 (n=1 605)	3.18	2.87-3.53	2.79	2.52-3.10	2.10	1.62-2.72
≥5 (n=2 311)	2.16	1.94-2.40	1.92	1.73-2.14	2.22	1.84-2.67
Subgroups according obstetrical history						
at least one previous adverse outcome (n= 58 219)						
continuous score	-	-	1.18	1.15-1.21	-	-
categorical score						
0 (n=54 649)	-	-	1		-	-
1-2 (n=3251)	-	-	1.59	1.49-1.69	-	-
3-4 (n=151)	-	-	2.30	1.83-2.90	-	-
≥5 (n=168)	-	-	1.30	0.98-1.74	-	-
only previous favourable outcome (n= 691 319)						
continuous score	-	-	1.22	1.20-1.25	-	-
categorical score						
0 (n=668 726)	-	-	1		-	-
1-2 (n=21100)	-	-	1.58	1.50-1.66	-	-
3-4 (n=546)	-	-	2.11	1.62-2.73	-	-
≥5 (n=947)	-	-	2.21	1.83-2.67	-	-
no previous outcome (n= 1 121 000)						
continuous score	-	-	1.24	1.23-1.26	-	-
categorical score						
0 (n= 1 100 372)	-	-	1		-	-
1-2 (n=19453)	-	-	1.82	1.75-1.89	-	-
3-4 (n=908)	-	-	3.25	2.86-3.69	-	-
≥5 (n=1196)	-	-	1.98	1.72-2.27	-	-

Notes: †for maternal age and obstetrical history in the whole population analyses and for maternal age in subgroup analyses

4.1.4 Synthèse de l'étude 4

Cette étude a permis de mettre en évidence un impact conséquent des comorbidités maternelles pré-existantes à la grossesse sur l'issue de grossesse et le début de vie de l'enfant. Si l'effet de différentes pathologies (diabète...) prises de manière individuelle sur la grossesse étaient déjà connu dans la littérature, l'impact des comorbidités maternelles n'avait pas été étudié de manière globale. La modélisation de notre étude a mesuré cet impact au travers d'un risque relatif s'élevant jusqu'à 1,88 (95 %IC = [1,82-1,95]) pour les femmes sans antécédent de grossesse.

L'utilisation de données issues de BMA a permis de recruter une cohorte de plus 1,8 million de femmes enceintes, ce qui n'aurait pas été envisageable dans le cadre d'une étude épidémiologique avec un recueil spécifique. D'autre part, la disponibilité de ces données et le chaînage des différentes hospitalisations à l'aide d'un identifiant unique ont rendu possible la reconstitution du parcours de soins dans les 6 ans préalables à la grossesse. Cette fenêtre d'exposition étendue a notamment permis de retracer une grande partie des antécédents obstétricaux.

Pour des raisons d'accès, cette étude n'a utilisé que des données issues du PMSI, il en résulte un potentiel biais de mesure ne retenant que les comorbidités les plus sévères ayant fait l'objet d'une hospitalisation. Par ailleurs, un certain nombre de comorbidités bénéficiant d'une prise en charge entièrement en externe n'ont pas pu être identifiées. Toutefois il n'y a pas d'arguments pouvant laisser penser que ce biais soit différentiel selon que les femmes aient ou non présenté une issue de grossesse défavorable. Concernant le score de Charlson, il existe des pathologies pouvant avoir un impact sur l'issue de grossesse telle que la sclérose en plaque ou l'obésité qui n'y sont pas intégrées.

L'extension de ce travail sur les données SNDS est prévue et permettra de résoudre un certain nombre de ces écueils.

4.2 Pronostic des patients atteints de maladie inflammatoire chronique après un syndrome coronarien aigu : Etude 5

4.2.1 Contexte

Les patients avec une inflammation chronique présentent davantage de risque de développer de l'athérosclérose et des événements thrombotiques (64,65). Parmi les pathologies inflammatoires chroniques, les MICI sont associées à des remaniements menant à une rigidité artérielle jusqu'à l'athérosclérose (66,67). Si plusieurs études se sont intéressées aux facteurs de risque cardiovasculaires chez les patients atteints de MICI (68–71), peu de données sont disponibles concernant le pronostic après un syndrome coronarien. L'objectif de cette étude était de déterminer le pronostic après la sortie d'hospitalisation pour syndrome coronarien aigu de patients atteints de MICI.

4.2.2 Méthodes

Cette étude de cohorte a inclus tous les patients sortis d'hospitalisation MCO entre le 01/01/2012 et le 31/12/2017 admis pour prise en charge initiale d'un syndrome coronarien en France. Seule la première sortie d'hospitalisation pour ce motif sur cette période a été retenue (hospitalisation index). Les données extraites à partir des hospitalisations dans les 12 mois précédant l'hospitalisation index concernaient les données démographiques, les facteurs de risque cardiovasculaires, les type de prise en charge en particulier sur la revascularisation. Les sujets ont été répartis en 2 groupes : le groupe des patients MICI défini par la présence d'au moins un code CIM10 de MICI (Maladie de Crohn K50 ou rectocolite hémorragique K51) dans au moins un des séjours hospitaliers et le groupe des patients non MICI défini par l'absence de code de MICI. Un appariement entre les sujets du groupe MICI et les sujets sans MICI a été effectué sur l'âge et le sexe. L'association entre la présence d'une MICI et le décès intra-hospitalier toute cause, critère de jugement principal, a été recherchée à l'aide d'un modèle de régression logistique. Plusieurs critères de jugement à un an (décès, réhospitalisation pour récurrence de syndrome coronarien aigu, AVC, insuffisance cardiaque) ont été examinés au travers d'un modèle de Cox.

4.2.3 Résultats

Parmi les 59 831 patients pris en charge pour un syndrome coronarien aigu, 246 (0,4 %) présentaient une MICI. Les patients avec une MICI différaient sur de nombreuses caractéristiques des patients sans MICI : ils étaient plus jeunes, avec un IMC plus élevé, plus nombreux à avoir des antécédents de coronaropathie et de pathologie rénale. Les 245 patients atteints de MICI ont été appariés à 1 225 patients sans MICI. Des différences en termes de IMC et d'antécédents de coronaropathie subsistaient après l'appariement.

Les modèles multivariés n'ont mis en évidence aucune différence en matière de mortalité intrahospitalière, de mortalité à un an et de critère composite combinant la mortalité, les rehospitalisation pour syndrome coronarien, AVC, ou insuffisance cardiaque. Le seul critère de jugement pour lequel une différence a été observée en analyse bivariée était le recours à une transfusion sanguine dans l'année suivant l'hospitalisation index, plus fréquent pour les sujets avec une MICI.

Outcomes after acute coronary syndrome in patients with inflammatory bowel disease

B Popovic, Y Juillère, N Jay, A Bannay

Abstract

Background: Patients with chronic inflammatory conditions are at an increased risk of developing atherothrombotic events. We aimed to assess the 1-year prognosis after myocardial infarction (MI) in patients with inflammatory bowel disease (IBD).

Methods: From the PMSI (Program de Medicalisation des Systèmes d'informatique) database, 246 out of 39,835 consecutive MI patients, hospitalized between 2010 and 2015, were diagnosed with IBD and followed up for 1 year after discharge.

Results: Compared with MI patients without IBD, MI patients with IBD were younger (aged 69 vs. 70.8 years, $p=0.04$) with a higher rate of elevated body mass index (BMI) (21.5% vs 15%, $p=0.004$), previously diagnosed ischemic cardiopathy (18.3% vs 12.6%, $p<0.0008$) and chronic renal disease (8.9% vs 5.6%, $p=0.02$). From the PMSI database, we subsequently analyzed an age- and sex-matched cohort ($n=1,670$, 1 patient with IBD /5 patients without IBD) and found that mortality (9% vs 8.3, $p=0.729$), recurrent MI (2.9% vs 1.9%, $p=0.328$), stroke (0.8% vs 0.6%, $p=0.656$) and hospitalization resulting from heart failure (3.3% vs 3.5%, $p=0.846$) did not significantly differ between the IBD and non-IBD groups within the first year after initial admission. However, a significant increase in the blood transfusion rate at the 1-year follow up was observed in MI patients with IBD compared with MI patients without IBD (15.1% vs 9.4%, $p=0.0083$).

Conclusion: our findings suggest that both residual risk and bleeding events should be carefully monitored in MI patients diagnosed with chronic inflammation such as that observed in IBD.

Introduction

Patients with chronic inflammatory conditions are at an increased risk of developing atherosclerosis and ultimately atherothrombotic events.^{1,2} Inflammation plays a crucial role in the process of atherosclerosis, where the immune cells dominate early atherosclerotic lesions, and their effector molecules accelerate the progression of these lesions to adverse cardiovascular events.³⁻⁵

Among chronic inflammatory diseases, inflammatory bowel diseases (IBDs), including Crohn's disease and ulcerative colitis, are associated with significant subclinical changes in the forms of arterial stiffness and subclinical atherosclerosis.^{6,7}

The incidence of cardiovascular disease in patients with IBD has been described in large population-based studies.⁸⁻¹¹ These studies are focused primarily on identifying risks of cardiovascular disease in patients with IBD. Currently, the impact of IBD on the prognosis after myocardial infarction remains elusive. Understanding these impacts will be crucial in managing and preventing primary and secondary CV risk factors in IBD patients. Accordingly, we analyzed risk profiles and monitored one-year prognosis following MI in a large cohort of patients with IBD.

Methods

Settings and Data Sources

All MI patients discharged of a French hospital stay between January 2012 and December 2017 were identified from the national discharge database, the PMSI (Programme de Medicalisation des Systèmes d'information), derived from the Diagnosis-related group system. The reliability and validity of the PMSI data have been described elsewhere.^{12,13} This national database contains both medical and administrative data, including, age, sex, diagnosis and medical procedures. Patients' diagnoses encoded according to the *International Classification of Diseases, Tenth Revision (ICD-10)* were classified as primary diagnosis (one per stay), or secondary diagnoses (as many as necessary). All patients' data were de-identified to ensure confidentiality.

To avoid iatrogenic myocardial infarction (MI), we selected patients with a diagnosis of MI present at admission. This approach was taken to specifically focus on patients who presented with acute myocardial ischemia and not those with MI secondary to surgery, hypotension, anemia, or other events after admission.

In cases of multiple admissions of a patient for an MI, only the first stay was included in our final analysis to avoid counting the same patient twice.

Collected data

Data were extracted from the national discharge database until 12 months prior the index hospitalisation, including i) demographic data : age and sex. ii) cardiovascular risk factors such as hypertension, smoking, diabetes mellitus, obesity, dyslipidemia, and comorbidities: peripheral artery disease and renal failure, atrial fibrillation, known ischemic cardiomyopathy, iii) type of MI, revascularization procedures: coronary angiography, coronary angioplasty and coronary artery bypass grafting (CABG); and iv) occurrence and type of IBD : Crohn disease and ulcerative colitis.

The primary outcome was in-hospital mortality defined as death during the index hospitalization. Secondary outcomes were 30-day mortality, need for red blood cell transfusion within

30 days after admission, 1-year readmission with death, 1-year readmission with recurrent MI, 1-year readmission with stroke, 1-year readmission with heart failure, 1-year readmission with need for red blood cell transfusion, 1-year readmission with CABG. Two composite secondary outcomes were defined i) 1-year readmission with one of the following event: death, recurrent MI, stroke, ii) 1-year readmission with one of the following event: death, recurrent MI, stroke, heart failures

Statistical Analysis

Categorical variables were described using numbers and percentages, between-group differences (with vs without IBD) were compared using χ^2 or Fisher exact tests, and quantitative variables were described using means and their standard deviation (SD), between-group differences were compared using t-test. A matched cohort was built matching each exposed patient with IBD to 1 non-exposed patient without IBD using age.

The relationship between in-hospital outcomes and the occurrence of BD was assessed using logistic regression models adjusted for covariates that were significantly associated in the univariate analysis. The relationship between 1-year outcomes and the occurrence of BD was assessed using Cox regression models adjusted for covariates that were significantly associated in the univariate analysis.

A sensitivity analysis was performed on patients with STEMI including transmural MI (ICD-10 codes I21.0, I21.1, I21.2, I21.3, and I21.9) and patients without STEMI including subendocardial MI (code I21.4). Statistical analyses were performed using SAS Enterprise Guide, version 7.1.

Results

Overall cohort (n= 59,831 patients)

From 2012–2017, the national discharge database recorded 59,831 patients with acute MI. Among them, 246 were also diagnosed with IBD. The baseline characteristics of patients with and without IBD are shown in Table 1. Compared with patients without IBD, patients with IBD were younger (mean age: 69 vs. 70.8 years, $p=0.04$) with a higher rate of elevated BMI (21.5% vs 15%, $p=0.004$), previously diagnosed ischemic cardiopathy (18.3% vs 12.6%, $p<0.0008$), and chronic renal disease (8.9% vs 5.6%, $p=0.02$).

STEMI represented the predominant initial presentation in patients with or without IBD (58.5% vs 58.8%, $p=0.926$). The rate of coronary revascularization, including those of percutaneous coronary intervention (45.5% vs 49.3%, $p=0.235$) and surgical revascularization (8.5% vs 8.7%, $p= 0.948$), were not significantly different between MI patients with or without IBD. Deaths within the first year after the initial admission occurred in 22 (8.9%) MI patients with IBD and 5291 (8.9%) MI patients without IBD ($p=0.972$). Adverse cardiovascular events, including recurrent MI (2.8% vs 1.9%, $p=0.296$), stroke (0.8% vs 0.6%, $p=0.612$), and hospitalization resulting from heart failure (3.3% vs 3.5%, $p=0.83$) within the first year after initial admission as well as the composite endpoint: death within 1 year, recurrent MI, stroke, or hospitalization resulting from heart failure (15.5% vs 13.8%, $p=0.445$) were also not significantly different between the IBD and non-IBD groups. However, the rate of red blood cell transfusion at the 1-year follow up was significantly higher in MI patients with IBD than that in MI patients without IBD (15% vs 8.3%, $p=0.0001$).

Matched cohort (n=1,670 patients)

In our age- and sex-matched cohort, the IBD group had a significantly higher incidence of BMI>25 (21.6% vs 14.8%, $p=0.007$) and previously diagnosed ischemic cardiomyopathy (18.4% vs 13.1%, $p=0.029$) than the non-IBD group. The proportion of coronary revascularization during the index hospitalization remained similar in both groups. In this matched cohort, the 1-year mortality (9% vs 8.3%, $p=0.729$), recurrent MI (2.9% vs 1.9%, $p=0.328$), stroke (0.8% vs 0.6%, $p=0.656$), and hospitalization resulting from heart failure (3.3% vs 3.5%, $p=0.846$) were not significantly different between the IBD and non-IBD groups within the first year after initial admission as well as for the composite endpoint: 1-year death, recurrent MI, or stroke and 1-year death, recurrent MI, stroke, or hospitalization for heart failure (15.5% vs 13.8%, $p=0.381$). However, a significant increase in the blood transfusion rate at the 1-year follow up was observed in MI patients with IBD compared with that in MI patients without IBD (15.1% vs 9.4%, $p=0.0083$).

Discussion

Our study using a large nationwide database presents a contemporary and comprehensive epidemiological analysis of cardiovascular risk profiles in patients with IBD following MI.

The principal findings of our study are: 1. Patients with IBD who experienced MI present a radically different CV profile as expected with similar CV factors compared with patients without IBD, except that a higher proportion of patients with IBD are overweight; 2. Early mortality and ischemic event rates were similar between the IBD and non-IBD group, while a more frequent need for blood transfusion was observed in the IBD group. These data suggest that bleeding/transfusion events, as well as ischemic risks, are critical factors that should be monitored judiciously in the 1-year follow up of MI patients who are also diagnosed with IBD.

Traditionally, young age and female subgroups were overrepresented in the group of patients with IBD who experienced acute coronary syndrome.¹⁴⁻¹⁰ IBD diagnosis at a young age has been known to be associated with more aggressive outcomes, implying that an increased level of inflammation^{15,16} is associated with a higher MI incidence rate and worse cardiovascular prognosis in young patients.

In our study, the clinical profiles of patients who experienced MI are noticeably similar between the IBD and non-IBD groups with discrepancies between age intervals and CV factors between both groups less pronounced, as expected. Our study included patients with first-time MI but also patients with known ischemic cardiopathy including previous MI previously diagnosed MI or ischemic cardiopathy patients, which could have influenced the clinical profile of MI patients with and without IBD. Patients with IBD usually accumulate traditional risk factors as time passes,¹⁴ which may reduce the relative contribution of inflammation in atherosclerosis.

Overweight is a matter of concern in patients with chronic inflammatory diseases such as IBD. Although the impact of overweight on bowel disease severity or course in patients with IBD remains uncertain, our findings suggest that overweight/obese patients with IBD represent a high-risk subgroup of acute ischemic events which requires specific attention. Indeed, overweight reflects current wellness, no control of CV factors and sedentary lifestyle of patients.¹⁷

Several studies previously reported an increased long-term morbidity/mortality in patients with post-MI and IBD, and different observations stratified the risk of CV events and the outcomes after MI according to persistent bowel disease activity, which may suggest a role of shared pathophysiological inflammatory mechanisms.^{8, 18, 19, 20, 21}

Based on the results of our analysis, the 1-year mortality rate, and adverse CV events were not significantly different between MI patients with and without IBD. These findings could be attributed to the improved

management of MI with prompt revascularization, use of high-intensity statin therapy, improved antiplatelet agents and a better control of risk factors.²² Anti-inflammatory approaches, such as statins and anti-interleukin-1 β antibodies, have recently been found to decrease the rate of cardiovascular events in those with elevated high-sensitivity C-reactive protein. These medications may be also highly effective in the prevention of MI in patients with autoimmune diseases including IBD.^{23,24}

Ischaemic recurrences after MI and stenting have also dropped considerably last decades thanks to the progressive refinement of antithrombotic strategies, including prolonged dual antiplatelet therapy and use of more potent P2Y₁₂ inhibitors. This residual ischemic risk remains greater in inflammatory disorders, inflammation and coagulation playing crucial roles in the pathogenesis of multiple chronic inflammatory disorders as IBD.²⁵

However, due to more potent and prolonged platelet inhibition, the incidence of major bleeding is increased.²⁶

Both early and late bleeding events have been associated with significant mortality risk, higher health costs and concerned either important or moderate bleedings.²⁷

Our study, for the first time, reports a higher blood transfusion rate in one-year follow up of MI patients with IBD compared with MI patients without IBD.

This transfusion rate in the IBD group became significantly higher after the first month of MI diagnosis, most likely due to antiplatelet therapy.

Our findings suggest that the bleeding risk in patients with IBD should be judiciously monitored with predictive bleeding scores,²⁸ as recommended by current guidelines.²⁹

While our investigation was retrospective, we chose the more quantifiable endpoint of blood transfusion events as a surrogate marker for bleeding. The bleeding events, particularly their origin and importance, have not been captured in our study as in other cohorts because blood transfusion may be the consequence of bleeding or the need to correct iron-deficiency anemia, which is relatively common in patients with IBD and active bowel disease.

While the high prevalence and adverse prognostic impacts of anemia in patients with ischemic heart disease are widely recognized, the treatment modality and hemoglobin cut off level for the onset of treatment remains debatable.^{30,31} This is particularly true when anemia is part of the systemic inflammatory response syndrome, which is seldom associated with severely decreased hemoglobin level.³²

Mechanistic studies evaluating the benefit of transfusion in patients with MI remain controversial.^{30, 31, 33} Some data suggest that blood transfusion could even be deleterious in the context of MI, as it may increase platelet activation and aggregation,³⁴ thereby inducing an inflammatory response.³⁵

The retrospective design of the present study and the incomplete data as concerning the extent of coronary artery disease and the treatment regimen at discharge drive its limitations.

Given the reliance on *ICD-10* codes for the selection of patients and their outcomes, there was a potential for misclassification- or underdetection-related biases. Even though we cannot exclude the bias

of misclassification or underdiagnosis, they may have only a minor impact on the findings.

Moreover, we did not classify disease activity according to the extent of inflammation or disease activity indices.

Because the course of IBD can vary from disease in remission to severe disease over time, no adjustment for disease activity was performed. However, the inflammation level, clinical symptoms, and mucosal disease activity in intestinal bowel disease seem to be disconnected.^{36, 37}

Moreover, we have no information concerning left ventricular and valvular functions, which may characterize post-MI scar extent and predict long-term prognosis and heart failure occurrence in our study cohort.

Conclusion

In this large population-based cohort study, the blood transfusion rate was significantly higher in MI patients with IBD compared with that in MI patients without IBD and this association was independent of the periprocedural course. There was no significant difference in 1-year mortality rates and ischemic events between MI patients with or without IBD. These data collectively suggest that MI patients with IBD are at a high risk of bleeding disorders; hence, they should be judiciously monitored.

References

1. Frostegard J. Atherosclerosis in patients with autoimmune disorders. *Arterioscler Thromb Vasc Biol.* 2005;25:1776-85.
2. Roifman I, Beck PL, Anderson TJ, Eisenberg MJ and Genest J. Chronic inflammatory diseases and cardiovascular risk: a systematic review. *Can J Cardiol.* 2011;27:174-82.
3. Ross R. Atherosclerosis--an inflammatory disease. *N Engl J Med.* 1999;340:115-26.
4. Libby P. Inflammation in atherosclerosis. *Nature.* 2002;420:868-74.
5. Wolf D and Ley K. Immunity and Inflammation in Atherosclerosis. *Circ Res.* 2019;124:315-327.
6. Wu GC, Leng RX, Lu Q, Fan YG, Wang DG and Ye DQ. Subclinical Atherosclerosis in Patients With Inflammatory Bowel Diseases: A Systematic Review and Meta-Analysis. *Angiology.* 2017;68:447-461.
7. Hatoum OA and Binion DG. The vasculature and inflammatory bowel disease: contribution to pathogenesis and clinical pathology. *Inflamm Bowel Dis.* 2005;11:304-13.
8. Ha C, Magowan S, Accortt NA, Chen J and Stone CD. Risk of arterial thrombotic events in inflammatory bowel disease. *Am J Gastroenterol.* 2009;104:1445-51.
9. Barnes EL, Beery RM, Schulman AR, McCarthy EP, Korzenik JR and Winter RW. Hospitalizations for Acute Myocardial Infarction Are Decreased Among Patients with Inflammatory Bowel Disease Using a Nationwide Inpatient Database. *Inflamm Bowel Dis.* 2016;22:2229-37.
10. Panhwar MS, Mansoor E, Al-Kindi SG, Sinh P, Katz J, Oliveira GH, Cooper GS and Ginwalla M. Risk of Myocardial Infarction in Inflammatory Bowel Disease: A Population-based National Study. *Inflamm Bowel Dis.* 2019;25:1080-1087.
11. Aniwani S, Pardi DS, Tremaine WJ and Loftus EV, Jr. Increased Risk of Acute Myocardial Infarction and Heart Failure in Patients With Inflammatory Bowel Diseases. *Clin Gastroenterol Hepatol.* 2018;16:1607-1615 e1.
12. Chantry AA, Deneux-Tharaux C, Cans C, Ego A, Quantin C, Bouvier-Colle MH and group Gs. Hospital discharge data can be used for monitoring procedures and intensive care related to severe maternal morbidity. *J Clin Epidemiol.* 2011;64:1014-22.
13. Couris CM, Polazzi S, Olive F, Remontet L, Bossard N, Gomez F, Schott AM, Mitton N, Colonna M and Trombert B. Breast cancer incidence using administrative data: correction with sensitivity and specificity. *J Clin Epidemiol.* 2009;62:660-6.
14. Herzog D, Fournier N, Buehr P, Rueger V, Koller R, Heyland K, Nydegger A, Spalinger J, Schibli S, Petit LM, Braegger CP and Swiss IBDCSG. Age at disease onset of inflammatory bowel disease is associated with later extraintestinal manifestations and complications. *Eur J Gastroenterol Hepatol.* 2018;30:598-607.
15. Beaugerie L, Seksik P, Nion-Larmurier I, Gendre JP and Cosnes J. Predictors of Crohn's disease. *Gastroenterology.* 2006;130:650-6.

16. Hoie O, Schouten LJ, Wolters FL, Solberg IC, Riis L, Mouzas IA, Politi P, Odes S, Langholz E, Vatn M, Stockbrugger RW, Mowm B and European Collaborative Study Group of Inflammatory Bowel D. Ulcerative colitis: no rise in mortality in a European-wide population based cohort 10 years after diagnosis. *Gut*. 2007;56:497-503
17. Nic Suibhne T, Raftery TC, McMahon O, Walsh C, O'Morain C and O'Sullivan M. High prevalence of overweight and obesity in adults with Crohn's disease: associations with disease and lifestyle factors. *J Crohns Colitis*. 2013;7:e241-8.
18. Gandhi S, Narula N, Marshall JK and Farkouh M. Are patients with inflammatory bowel disease at increased risk of coronary artery disease? *Am J Med*. 2012;125:956-62.
19. Andersohn F, Waring M and Garbe E. Risk of ischemic stroke in patients with Crohn's disease: a population-based nested case-control study. *Inflamm Bowel Dis*. 2010;16:1387-92.
20. Kristensen SL, Ahlehoff O, Lindhardtsen J, Erichsen R, Jensen GV, Torp-Pedersen C, Nielsen OH, Gislason GH and Hansen PR. Disease activity in inflammatory bowel disease is associated with increased risk of myocardial infarction, stroke and cardiovascular death--a Danish nationwide cohort study. *PLoS One*. 2013;8:e56944.
21. Kristensen SL, Ahlehoff O, Lindhardtsen J, Erichsen R, Lamberts M, Khalid U, Nielsen OH, Torp-Pedersen C, Gislason GH and Hansen PR. Prognosis after first-time myocardial infarction in patients with inflammatory bowel disease according to disease activity: nationwide cohort study. *Circ Cardiovasc Qual Outcomes*. 2014;7:857-62.
22. Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV and Go AS. Population trends in the incidence and outcomes of acute myocardial infarction. *N Engl J Med*. 2010;362:2155-65.
23. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM, Jr., Kastelein JJ, Koenig W, Libby P, Lorenzatti AJ, MacFadyen JG, Nordestgaard BG, Shepherd J, Willerson JT, Glynn RJ and Group JS. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008;359:2195-207.
24. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, Fonseca F, Nicolau J, Koenig W, Anker SD, Kastelein JJP, Cornel JH, Pais P, Pella D, Genest J, Cifkova R, Lorenzatti A, Forster T, Kobalava Z, Vida-Simiti L, Flather M, Shimokawa H, Ogawa H, Dellborg M, Rossi PRF, Troquay RPT, Libby P, Glynn RJ and Group CT. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med*. 2017;377:1119-1131.
25. Danese S, Papa A, Saibeni S, Repici A, Malesci A and Vecchi M. Inflammation and coagulation in inflammatory bowel disease: The clot thickens. *Am J Gastroenterol*. 2007;102:174-86.
26. Angiolillo DJ. The evolution of antiplatelet therapy in the treatment of acute coronary syndromes: from aspirin to the present day. *Drugs*. 2012;72:2087-116.
27. Secemsky EA, Yeh RW, Kereiakes DJ, Cutlip DE, Cohen DJ, Steg PG, Cannon CP, Apruzzese PK, D'Agostino RB, Sr., Massaro JM, Mauri L and Dual Antiplatelet Therapy Study I. Mortality Following Cardiovascular and Bleeding Events Occurring Beyond 1 Year After Coronary Stenting: A Secondary Analysis of the Dual Antiplatelet Therapy (DAPT) Study. *JAMA Cardiol*. 2017;2:478-487.
28. Costa F, van Klaveren D, James S, Heg D, Raber L, Feres F, Pilgrim T, Hong MK, Kim HS, Colombo A, Steg PG, Zanchin T, Palmerini T, Wallentin L, Bhatt DL, Stone GW, Windecker S, Steyerberg EW, Valgimigli M and Investigators P-DS. Derivation and validation of the predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet

therapy (PRECISE-DAPT) score: a pooled analysis of individual-patient datasets from clinical trials. *Lancet*. 2017;389:1025-1034.

29. Valgimigli M, Bueno H, Byrne RA, Collet JP, Costa F, Jeppsson A, Juni P, Kastrati A, Kolh P, Mauri L, Montalescot G, Neumann FJ, Petricevic M, Roffi M, Steg PG, Windecker S, Zamorano JL, Levine GN, Group ESCSD, Guidelines ESCCfP and Societies ESCNC. 2017 ESC focused update on dual antiplatelet therapy in coronary artery disease developed in collaboration with EACTS: The Task Force for dual antiplatelet therapy in coronary artery disease of the European Society of Cardiology (ESC) and of the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J*. 2018;39:213-260.

30. Rao SV, Eikelboom JA, Granger CB, Harrington RA, Califf RM and Bassand JP. Bleeding and blood transfusion issues in patients with non-ST-segment elevation acute coronary syndromes. *Eur Heart J*. 2007;28:1193-204.

31. Hebert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, Tweeddale M, Schweitzer I and Yetisir E. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. *N Engl J Med*. 1999;340:409-17.

32. Stucchi M, Cantoni S, Piccinelli E, Savonitto S and Morici N. Anemia and acute coronary syndrome: current perspectives. *Vasc Health Risk Manag*. 2018;14:109-118.

33. Ducrocq G, Puymirat E, Steg PG, Henry P, Martelet M, Karam C, Schiele F, Simon T and Danchin N. Blood transfusion, bleeding, anemia, and survival in patients with acute myocardial infarction: FAST-MI registry. *Am Heart J*. 2015;170:726-734 e2.

34. Silvain J, Pena A, Cayla G, Brieger D, Bellemain-Appaix A, Chastre T, Vignalou JB, Beygui F, Barthelemy O, Collet JP and Montalescot G. Impact of red blood cell transfusion on platelet activation and aggregation in healthy volunteers: results of the TRANSFUSION study. *Eur Heart J*. 2010;31:2816-21.

35. Fransen E, Maessen J, Dentener M, Senden N and Buurman W. Impact of blood transfusions on inflammatory mediator release in patients undergoing cardiac surgery. *Chest*. 1999;116:1233-9.

36. Peyrin-Biroulet L, Reinisch W, Colombel JF, Mantzaris GJ, Kornbluth A, Diamond R, Rutgeerts P, Tang LK, Cornillie FJ and Sandborn WJ. Clinical disease activity, C-reactive protein normalisation and mucosal healing in Crohn's disease in the SONIC trial. *Gut*. 2014;63:88-95.

37. Torres J, Mehandru S, Colombel JF and Peyrin-Biroulet L. Crohn's disease. *Lancet*. 2017;389:1741-1755.

Table 1. Baseline Characteristics in the overall and matched cohorts of IBD (inflammatory bowel diseases) and non IBD patients

	Overall cohort			Matched cohort		
	Patients with IBD (n=246)	Patients without IBD (n=59585)	P value	Patients with IBD (n=245)	Patients without IBD (n=1 225)	P value
Mean age	69.0 (13.3)	70.8 (14.6)	0.041	69.2 (13.0)	69.2 (13.0)	1
<18	0 (0.0)	7 (0.01)		0 (0)	0 (0)	
18-55	38 (15.2)	10 186 (17.1)		37 (15.1)	185 (15.1)	
56-65	68 (27.8)	11 321 (19.0)	0.017	68 (27.8)	340 (27.8)	1
66-75	50 (20.3)	12 291 (20.6)		50 (20.4)	250 (20.4)	
76-85	59 (24.0)	15 243 (25.6)		59 (24.1)	295 (24.1)	
>85	31 (12.6)	10 537 (17.7)		31 (12.6)	155 (12.6)	
Men	162 (65.9)	38 470 (64.6)	0.673	83 (33.9)	415 (33.9)	1
Risks factors						
Smoker (current or past)	48 (19.5)	11 119 (18.7)	0.732	48 (19.6)	263 (21.5)	0.474
Hypertension	131 (53.3)	33 248 (55.8)	0.422	131 (53.5)	666 (54.4)	0.790
Dyslipidemia	45 (18.3)	9 050 (15.2)	0.176	45 (18.4)	198 (16.2)	0.393
Diabetes mellitus	67 (27.2)	15 874 (26.6)	0.833	67 (27.3)	313 (25.6)	0.550
Increased body mass index	53 (21.5)	8 914 (15.0)	0.004	53 (21.6)	181 (14.8)	0.007
Medical history						
Previous myocardial infarction	12 (4.9)	1 788 (3.0)	0.085	12 (4.9)	40 (3.3)	0.206
Previous ischemic cardiomyopathy	45 (18.3)	7 532 (12.6)	0.008	45 (18.4)	160 (13.1)	0.029
Previous Heart Failure	17 (6.9)	4 349 (7.3)	0.815	17 (6.9)	80 (6.5)	0.815
Atrial fibrillation	9 (3.7)	1 871 (3.1)	0.642	9 (3.7)	32 (2.6)	0.349
Peripheral artery disease	15 (6.1)	2 926 (4.9)	0.390	15 (6.1)	62 (5.1)	0.483
Chronic Renal Failure	22 (8.9)	3 353 (5.6)	0.02	22 (9.0)	73 (6.0)	0.076
Clinical presentation: STEMI	144 (58.5)	35 054 (58.8)	0.926	143 (58.4)	716 (58.4)	0.980
Heart failure during index hospitalisation	62 (25.2)	15 341 (25.8)	0.846	62 (25.3)	311 (25.4)	0.978
Revascularisation during index hospitalisation						
Patients undergoing percutaneous coronary interventions	112 (45.5)	29 387 (49.3)	0.235	112 (45.7)	626 (51.1)	0.116
Patients undergoing Coronary surgical revascularization	21 (8.5)	5 157 (8.7)	0.948	21 (8.6)	117 (9.6)	0.628
In-hospital mortality	14 (5.7)	4094 (6.9)	0.755/0.465	14 (5.7)	79 (6.5)	0.194
30-day mortality	12 (4.9)	4 109 (6.9)	0.212	12 (4.9)	76 (6.2)	0.419
Need for red blood cell transfusion within 30 days after admission	18 (7.3)	2 802 (4.7)	0.053	18 (7.4)	69 (5.6)	0.298

Table 2. Prognosis at 1 year in the overall and matched Cohorts

	Overall cohort			Matched population		
	Patients with IBD (n=246)	Patients without IBD (n=59585)	P value	Patients with IBD (n=245)	Patients without IBD (n=1225)	P value
1 year death	22 (8.9)	5 291 (8.9)	0.972	22 (9.0)	102 (8.3)	0.729
Recurrent MI	7 (2.8)	1 148 (1.9)	0.296	7 (2.9)	23 (1.9)	0.328
Stroke	2 (0.8)	339 (0.6)	0.612	2 (0.8)	7 (0.6)	0.656
Hospitalisation for Heart Failure	8 (3.3)	2 081 (3.5)	0.838	8 (3.3)	43 (3.5)	0.846
Need for red blood cell transfusion during 1 year after admission	37 (15.0)	4 929 (8.3)	0.0001	37 (15.1)	115 (9.4)	0.0083
CABG at 1 year	22 (8.9)	5 901 (9.9)	0.615	22 (9.0)	136 (11.1)	0.325
Composite criteria: 1 year Death + recurrent MI + stroke	30 (12.2)	6 603 (11.1)	0.579	30 (12.2)	129 (10.5)	0.415
1 year Death + recurrent MI + stroke + hospitalization for heart failure	38 (15.5)	8 222 (13.8)	0.455	38 (15.5)	165 (13.5)	0.381

Table 3. Multivariable analysis of short- and long-term prognosis in the overall and matched cohorts

	Overall cohort			Matched cohort		
	OR/HR ¹	95% CI	p-value	OR/HR ²	95% CI	p-value
In-hospital mortality	0.87	0.50-1.50	0.623	0.76	0.40-1.42	0.387
30-day mortality	0.74	0.41-1.34	0.323	0.72	0.36-1.41	0.334
1-year mortality	1.08	0.71-1.65	0.142	0.96	0.59-1.55	0.869
1-year recurrent MI	1.43	0.68-3.00	0.348	1.18	0.43-3.24	0.751
1-year stroke	1.48	0.37-5.95	0.581	2.15	0.26-17.62	0.475
1-year heart failure	0.93	0.46-1.86	0.838	0.82	0.37-1.82	0.631
1-year mortality or 1-year recurrent MI or 1-year stroke	1.17	0.82-1.68	0.380	1.03	0.68-1.57	0.880
1-year mortality or 1-year recurrent MI or 1-year stroke or 1-year heart failure	1.18	0.86-1.63	0.303	1.03	0.71-1.50	0.870

¹adjusted for age, gender, history of obesity, history of ischemic cardiomyopathy, history of chronic renal failure, number of hospital stay during last year , percutaneous coronary intervention

²adjusted for history of obesity, history of ischemic cardiomyopathy, history of chronic renal failure, number of hospital stay during last year , percutaneous coronary intervention

4.2.4 Synthèse de l'étude 5

Cette étude s'est intéressée à l'impact des MICI dans le pronostic des sujets atteints d'un syndrome coronarien aigu.

Le recrutement exhaustif à partir des bases nationales du PMSI a permis d'inclure 59 831 sujets à partir des bases nationales du PMSI. Les sujets ont pu être classés selon leur statut par rapport aux MICI au travers de leurs hospitalisations dans l'année précédente.

Toutefois, le nombre de sujets atteints de MICI et ayant présenté un syndrome coronarien aigu durant 6 ans n'était que de 245. De ce fait, la puissance de l'étude restait très faible (entre 6 et 13 %). Aucune différence sur la mortalité ou les critères combinant mortalité et réhospitalisation n'ont pu être mis en évidence.

Suite à des suggestions de relecteurs lors de soumission dans une revue de cardiologie, ce travail fait l'objet de nouvelles analyses afin d'augmenter la durée de suivi pour examiner le pronostic à 2 ans.

4.3 Effet de l'indice de masse corporelle sur la mortalité de patients atteints de COVID-19 pris en charge en réanimation : Etude 6

4.3.1 Contexte

Dès le début de la pandémie de COVID-19, l'obésité a été identifiée comme facteur de risque de survenue de formes sévères. Cependant, son impact sur la mortalité restait encore controversé (72–75), en particulier, la linéarité de la relation entre l'indice de masse corporelle (IMC) et la mortalité.

Dans les formes les plus graves prises en charge en réanimation, peu de données étaient disponibles dans la littérature au décours de la première vague. L'objectif de cette étude était d'évaluer l'association entre l'IMC et la mortalité intra-hospitalière chez des patients avec un syndrome de détresse respiratoire aigu lié à la COVID-19.

4.3.2 Méthode

Cette étude a inclus tous les patients avec un diagnostic de COVID-19 admis en réanimation au CHRU de Nancy entre le 27/02/2020 et le 18/04/2020 pour un syndrome de détresse respiratoire aiguë.

Le recueil de données a mobilisé 2 sources : les bases médico-administratives du CHRU et les dossiers médicaux. Les sujets ont été ciblés au travers des données du PMSI : les RUM des patients COVID-19 ont été marqués par des codes propres de diagnostic (diagnostic positif ou probable). Un bordereau spécifique de données a permis de vérifier les critères d'inclusion et d'exclusion et de collecter les données cliniques et biologiques par retour aux dossiers médicaux. Les dates de séjour en réanimation et totale ainsi que le statut vital ont été extraits des données PMSI. Un modèle de régression logistique a évalué l'association entre le décès intra-hospitalier et l'IMC en fonction de différents seuils.

4.3.3 Résultats

Parmi les 222 patients inclus, 43,6 % avaient un IMC supérieur ou égal à 30 et 68 % présentaient au moins une comorbidité. Le taux de mortalité intra-hospitalière était de 20,3 % soit 45 sujets dont 41 sont décédés en réanimation. Dans les classes d'IMC, le taux de mortalité le plus faible (13,8 %) était observé dans la modalité 30-39,9kg/m² suggérant une relation non linéaire entre IMC et mortalité.

Certaines comorbidités étaient associées, en analyse bivariée, au décès intra-hospitalier telles que l'hypertension, les dyslipidémies la BPCO, l'insuffisance cardiaque et rénale.

Le modèle de régression logistique ajusté sur l'âge, le nombre de comorbidités, le stade de la détresse respiratoire et le score IGSII a retrouvé un sur-risque de décès significatif chez les sujets avec un IMC inférieur ou égal à 29kg/m² (OR=3,09 ; IC_{95%}=[1,19-7,72]) et chez les sujets avec un IMC supérieur à 39 kg/m² (OR=6,74 ; IC_{95%}=[1,75-25,82]) par rapport aux sujets avec un IMC entre 29,1 et 39 kg/m².



Clinical Research

Obesity and mortality in critically ill COVID-19 patients with respiratory failure

Richard Dana¹ · Aurélie Bannay² · Pauline Bourst¹ · Caroline Ziegler³ · Marie-Reine Losser¹ · Sébastien Gibot⁴ · Bruno Levy⁵ · Gérard Audibert¹ · Olivier Ziegler⁶

Received: 5 October 2020 / Revised: 5 May 2021 / Accepted: 24 May 2021
© The Author(s), under exclusive licence to Springer Nature Limited 2021

Abstract

Background Patients with obesity are at increased risk of severe COVID-19, requiring mechanical ventilation due to acute respiratory failure. However, conflicting data are obtained for intensive care unit (ICU) mortality.

Objective To analyze the relationship between obesity and in-hospital mortality of ICU patients with COVID-19.

Subjects/methods Patients admitted to the ICU for COVID-19 acute respiratory distress syndrome (ARDS) were included retrospectively. The following data were collected: comorbidities, body mass index (BMI), the severity of ARDS assessed with PaO₂/FiO₂ (P/F) ratios, disease severity measured by the Simplified Acute Physiology Score II (SAPS II), management and outcomes.

Results For a total of 222 patients, there were 34 patients (15.3%) with normal BMI, 92 patients (41.4%) who were overweight, 80 patients (36%) with moderate obesity (BMI:30–39.9 kg/m²), and 16 patients (7.2%) with severe obesity (BMI ≥ 40 kg/m²). Overall in-hospital mortality was 20.3%. Patients with moderate obesity had a lower mortality rate (13.8%) than patients with normal weight, overweight or severe obesity (17.6%, 21.7%, and 50%, respectively; *P* = 0.011. Logistic regression showed that patients with a BMI ≤ 29 kg/m² (odds ratio [OR] 3.64, 95% CI 1.38–9.60) and those with a BMI > 39 kg/m² (OR 10.04, 95% CI 2.45–41.09) had a higher risk of mortality than those with a BMI from 29 to 39 kg/m². The number of comorbidities (≥2), SAPS II score, and P/F < 100 mmHg were also independent predictors for in-hospital mortality.

Conclusions COVID-19 patients admitted to the ICU with moderate obesity had a lower risk of death than the other patients, suggesting a possible obesity paradox.

Introduction

The disproportionate impact of COVID-19 on patients with obesity is now well established [1–5]. Several observational

studies have clearly demonstrated that obesity leads to the severe form of COVID-19, and is associated with a greater risk of advanced levels of treatment such as admission to intensive or critical care, invasive mechanical ventilation, and death [3, 5–19]. The association between BMI and in-hospital mortality has been explored in hospitalized patients with COVID-19. Overall, the results are conflicting. Some studies reported that obesity has no effect on in-hospital mortality [12, 20, 21], whereas recent cohort studies found a significant association between obesity and in-hospital

These authors contributed equally: Richard Dana, Aurélie Bannay

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41366-021-00872-9>.

✉ Olivier Ziegler
o.ziegler@chru-nancy.fr

¹ Department of Anesthesiology and Intensive Care Medicine, CHRU-Nancy, Université de Lorraine, Nancy, France

² Medical Information Department, CHRU-Nancy, CNRS, Inria, LORIA, Université de Lorraine, Nancy, France

³ Department of Hematology, CHRU-Nancy, Brabois Hospital, Nancy, France

⁴ Medical Intensive Care Unit, CHRU-Nancy, Central Hospital, Université de Lorraine, Nancy, France

⁵ Medical Intensive Care Unit, CHRU-Nancy, Brabois Hospital, Université de Lorraine, Nancy, France

⁶ Specialized Obesity Center and Department of Endocrinology, Diabetology, Nutrition, CHRU-Nancy, Brabois Hospital, Université de Lorraine, Nancy, France

mortality, after adjustment for confounding factors. Williamson et al. [22] reported a dose-response independent relationship between BMI and mortality with an increase in the fully adjusted hazard ratio from 1.05 for patients with a BMI between 30 and 34.9 kg/m² to 1.92 for patients with a BMI \geq 40 kg/m². Anderson et al. [6], found that a BMI \geq 35 kg/m² was associated with an increased risk of death or intubation, although there was an interaction between BMI and age of admission with a significant impact only for patients aged under 65 years. A French cohort study including 5795 patients reported that obesity doubles mortality independently of known chronic comorbidities in patients hospitalized with Covid-19 [9]. Consequently, the hypothesis of a survival obesity paradox with COVID-19 is controversial [2, 6, 23, 24].

In critically ill adult COVID-19 patients, the relationship between obesity and mortality is less clear. Results assessing if obesity is an independent risk factor for intensive care unit (ICU) mortality are contradictory. No definitive conclusion on excess weight and the COVID-19 mortality rate adjusted for key confounding factors has been published by the long-running ICNARC case-mix registry [25]. In the large New York cohort study of Cummings et al. [26], obesity was not identified as an independent risk factor for mortality. Lower BMI was associated with mortality in the study of Auld et al. [27]. On the contrary, severe obesity was independently associated with death in a recently published US multicenter cohort study [28]. The role of obesity was not analyzed in the first systematic review and meta-analysis of the outcome of patients admitted to ICU with COVID-19, probably due to a lack of data [29]. Moreover, the relative importance of different underlying health conditions and patient status on admission for prognosis in the ICU is unclear.

The aim of our study was to evaluate the relationship between BMI and in-hospital mortality in critically ill patients with Acute Respiratory Distress Syndrome (ARDS) due to COVID-19. We attempted to determine a BMI threshold for increased mortality.

Methods

Study design and participants

Our cohort study enrolled all consecutive adult patients with severe SARS-CoV-2 pneumonia, admitted to 4 ICUs at Nancy University Hospital during the early epidemic phase from February 27th to April 18th, 2020. The diagnosis of SARS-Cov-2 infection was defined according to World Health Organization guidance either by real-time reverse transcription-polymerase chain reaction (RT-PCR) for nasal or pharyngeal samples or the presence of hallmark COVID-

19 lesions on CT scans. Patients in the ICU with a positive COVID-PCR without pneumonia and hospitalized for another reason were excluded.

Clinical data were collected by manual review of electronic medical records, using a standardized case report form. The recorded data included both administrative data (age, sex, and length of ICU and hospital stay) and clinical data (height, weight, comorbidities). Variables were derived from these collected data. In order to measure the burden of any underlying conditions, we prespecified the following 7 variables for the assessment of the number of comorbidities, based on published studies and completeness of data: diabetes, hypertension, dyslipidemia, coronary heart disease (CHD), chronic obstructive pulmonary disease (COPD), immunodeficiency (diseases causing immunosuppression or use of immunosuppressant medications) and chronic kidney disease (CKD). We also collected data on management including mode of ventilator support, mechanical ventilation, the fraction of inspired oxygen (FiO₂), arterial partial pressure in oxygen (PaO₂), PaO₂/FiO₂ ratio, specific treatment (corticosteroids, hydroxychloroquine, antiviral agents). Severity was assessed using Sequential Organ Failure Assessment (SOFA) score [30] and Simplified Acute Physiology Score II (SAPS II) [31] at admission. ARDS severity was defined as a PaO₂/FiO₂ ratio of 300 to 201, 200 to 101, and \leq 100 mm Hg for mild, moderate, and severe ARDS, respectively. The BMI was calculated using weight and height measurements at admission and patients were classed into five groups: underweight (BMI < 18.5 kg/m²), normal (18.5–24.9 kg/m²), overweight (25–29.9 kg/m²), mild to moderate obesity (30–39.9 kg/m²) and severe obesity (\geq 40 kg/m²), respectively.

According to French regulations, this study was declared to the French National Data Protection Authority (CNIL: Commission nationale de l'informatique et des libertés) and registered on ClinicalTrials.gov (No: 2020P1082). It was also approved by the Nancy University Hospital ethics committee.

The primary outcome was in-hospital death. Secondary outcomes included 30-day mortality, clinical characteristics, duration of mechanical ventilation, and length of ICU and hospital stays, according to BMI categories.

Statistical analysis

The descriptive analysis of characteristics was carried out using frequency distribution for categorical variables, the mean and standard deviation for normally distributed continuous variables, or median and interquartile range (IQR) for continuous variables with skewed distribution. Characteristics were compared by groups using the two-tailed *t*-test or Chi² test, as appropriate. Assumptions for tests were tested and checked.

Logistic regression models were built to analyze the associations between BMI and all-cause in-hospital/30-day

mortality, using the group with the lowest mortality rate as a reference. Potential confounders were selected by clinical relevance and/or a P -value < 0.10 in the univariate analysis. Results are given with their odds ratio (OR) and 95% confidence interval (CI 95%). Statistical analysis was performed with R 3.6.3 statistical software.

Results

From February 28th to April 18th, 2020, for 226 eligible patients, we included 222 critically ill patients with acute respiratory failure admitted to ICUs with laboratory-confirmed ($n = 212$) or highly probable ($n = 10$) COVID-19 infection, during the first pandemic wave at Nancy University Hospital. Two patients withdrew their consent and two patients had missing BMI data.

Patients had a mean age of 62.4 ± 12 years, median BMI of 29.1 kg/m^2 (IQR, 25.8–32.8), and 72% were male (Table 1). Obesity ($\text{BMI} \geq 30 \text{ kg/m}^2$) was present in 96 patients (43.24%) and severe obesity ($\text{BMI} \geq 40 \text{ kg/m}^2$) in 16 patients (7.2%). Diabetes (26.6%), hypertension (48.6%), dyslipidemia (35.6%), cardiovascular disease (24.3%) including CHD (14%), pre-existent respiratory diseases (18.9%), including COPD (6.3%), CKD (6.8%) and immunosuppression (10.8%) were the most frequent comorbid conditions. Thirty-two percent of patients had no identified comorbidities, 25.2% had one, 21.2% two, and 21.6% three or more comorbidities. The median number of comorbid conditions was 1 (IQR, 0–2).

Mortality

Forty-one patients (18.5%) died in the ICU, while 4 others died after ICU discharge to a medical ward. The overall in-hospital mortality was 20.3%. The characteristics of patients at ICU admission according to survival or death ($n = 45$) are shown in Table 1.

Older patients were more likely to die. There were no statistically significant differences in sex distribution. Patients with moderate obesity ($\text{BMI} 30\text{--}39.9 \text{ kg/m}^2$) had a lower in-hospital death rate (13.8%) compared to normal weight (17.6%) and overweight (21.7%) patients, while patients with severe obesity had the highest mortality rate (50%, $P = 0.011$) (Table 1 and Fig. 1). Univariate analysis showed that hypertension, dyslipidemia, COPD, CHD, and CKD were also significantly associated with increased hospital mortality. A preexisting disease could not be identified in 37.3% of survivors compared to 11.1% for non-survivors ($P < 0.001$). About 31% of non-survivors had 2 co-morbidities compared to 18.6% for survivors and the same was true for 3 comorbidities or more (37.8 vs. 17.5%; $P < 0.001$).

Patients who died were more likely to have presented with more severe illness and organ dysfunction. In non-survivors, the incidence of severe ARDS was higher than in survivors 77.7% vs. 36.7%, $P < 0.001$). Death was associated with lower worst $\text{PaO}_2/\text{FiO}_2$ values ($\text{PaO}_2/\text{FiO}_2$ 85 mm Hg [IQR, 70.7–97.7] in non-survivors vs. 118.5 mm Hg [IQR, 82.0–153], in survivors, $P < 0.001$). The SAPS II and day 1 SOFA scores were significantly associated with overall in-hospital mortality (both $P < 0.001$).

One hundred and eighty-three (82.4%) of all patients required invasive mechanical ventilation, 93.3% in non-survivors vs. 79.7% in survivors ($P = 0.053$). The median duration of ventilation was about 13 days for both survivors and non-survivors. Compared with survivors, non-survivors were more likely to receive corticosteroids ($P = 0.008$).

Role of obesity

The distribution of age and sex was relatively homogenous across BMI categories (Table 2). Patients with severe ($\text{BMI} \geq 40 \text{ kg/m}^2$) and moderate obesity ($\text{BMI} 30$ to $< 40 \text{ kg/m}^2$) had higher rates of diabetes (62.5% and 38.8%, respectively) compared to patients with normal weight or overweight (23.5% and 10.9%, respectively; $P < 0.001$). The prevalence of hypertension ($P < 0.004$) and obstructive sleep apnea ($P < 0.001$) progressively increased with the BMI category from normal weight to severe obesity. The proportion of patients with no comorbidities was highest in the normal weight group (44.1%). The proportion of patients with 3 or more comorbidities ranged from 12% in overweight patients to 43.8% for those with $\text{BMI} \geq 40 \text{ kg/m}^2$.

There was a trend, albeit not significant, for patients with moderate or severe obesity (50.0% and 68.7%, respectively) to have a higher risk of severe ARDS compared to patients with normal BMI (32.3%) or were overweight (41.3%) ($P = 0.106$) (Table 2 and Fig. 1). Conversely, $\text{PaO}_2/\text{FiO}_2$ and the prevalence of minor ARDS progressively decreased with BMI category from the normal BMI group to those with severe obesity, despite the fact that the proportion of patients receiving invasive mechanical ventilation was the same in all groups. SAPS II and SOFA scores were comparable in patients with or without obesity. In-hospital mortality was significantly lower for patients with moderate obesity ($\text{BMI} 30\text{--}39.9 \text{ kg/m}^2$) (13.8%) compared to patients with normal weight, overweight or severe obesity (17.6%, 21.7%, and 50%, respectively; $P = 0.011$). The results for 30-day mortality were similar.

Multivariable logistic regression

In order to discriminate the role of BMI categories, we performed an exploratory analysis with 4 categories (Supplementary Table 1). This model showed a similar risk for

Table 1 Characteristics of patients according to their vital status.

	Overall (n = 222)	Survivors (n = 177)	Non-survivors (n = 45)	p-value
Age at admission	64.0 (55.3–72.0)	62.0 (54.0–70.0)	72 (62.0–75.0)	<0.001
18–45 years	19 (8.6)	19 (10.8)	0 (0.0)	0.007
45–54 years	34 (15.5)	28 (15.9)	6 (13.6)	
5–64 years	63 (28.6)	53 (30.1)	10 (22.7)	
65–74 years	72 (32.7)	57 (32.4)	15 (34.1)	
75 years and more	32 (14.5)	19 (10.8)	13 (29.5)	
Male sex	160 (72.1)	128 (72.3)	32 (71.1)	1.000
BMI (kg/m ²)	29.1 (25.8–32.8)	29.3 (25.9–32.8)	28.7 (25.7–32.3)	0.251
BMI categories				0.024
Normal weight	34 (15.3)	28 (15.8)	6 (13.3)	
Overweight	92 (41.4)	72 (40.7)	20 (44.4)	
Moderate obesity	80 (36.1)	69 (39.0)	11 (24.4)	
Severe obesity	16 (7.2)	8 (4.5)	8 (17.9)	
Diabetes mellitus	59 (26.6)	44 (24.9)	15 (33.3)	0.337
Hypertension	108 (48.6)	78 (44.1)	30 (66.7)	0.011
Dyslipidemia	79 (35.6)	56 (31.6)	23 (51.1)	0.024
Chronic pulmonary disease	42 (18.9)	29 (16.4)	13 (28.9)	0.089
Including COPD	14 (6.3)	5 (2.8)	9 (20.0)	<0.001
Including OSAS	22 (9.9)	16 (9.0)	6 (13.3)	0.561
Cardiovascular disease	54 (24.3)	39 (22.0)	15 (33.3)	0.167
Including CHD	31 (14.0)	20 (11.3)	11 (24.4)	0.042
including chronic heart failure	5 (2.3)	3 (1.7)	2 (4.5)	0.568
Immunodeficiency	24 (10.8)	17 (9.6)	7 (15.6)	0.379
Including immunosuppressant medication	9 (4.1)	7 (4.0)	2 (4.4)	1.000
Including active cancer	9 (4.1)	5 (2.8)	4 (8.9)	0.156
Chronic kidney disease	15 (6.8)	8 (4.5)	7 (15.6)	0.021
Neurological disorder	10 (4.5)	7 (4.0)	3 (6.7)	0.703
Number of comorbidities				<0.001
0	71 (32.0)	66 (37.3)	5 (11.1)	
1	56 (25.2)	47 (26.6)	9 (20.0)	
2	47 (21.2)	33 (18.6)	14 (31.1)	
3 and more	48 (21.6)	31 (17.5)	17 (37.8)	
ARDS severity ^a				<0.001
Mild	13 (5.9)	13 (7.3)	0 (0)	
Moderate	109 (49.1)	98 (55.4)	11 (24.4)	
Severe	100 (45)	66 (37.3)	34 (75.6)	
PaO ₂ /FiO ₂ (mmHg)	109 (79.0–142.0)	118.5 (84.0–153.0)	85 (70.0–97.0)	<0.001
Mechanical ventilation	183 (82.4)	141 (79.7)	42 (93.3)	0.053
Mechanical ventilation duration (days)	12.5 (8.0–18.0)	12.0 (7.0–18.0)	13.0 (8.0–19.0)	0.518
Hydroxychloroquine	31 (14.3)	25 (14.5)	6 (13.3)	1.000
Antiviral therapy	42 (19.0)	32 (18.2)	10 (22.2)	0.686
Corticosteroids	53 (23.9)	35 (19.8)	18 (40.0)	0.008
SAPS II	44.5 (32.3–62.0)	42.0 (28.0–59.0)	62.0 (44.0–75.0)	<0.001
SOFA score	7.0 (4.0–8.0)	7.0 (4.0–8.0)	8.0 (7.0–11.0)	<0.001
ICU length of stay (days)	13.0 (5.0–22.0)	13.0 (5.0–22.0)	12.0 (6.0–19.0)	0.360
Length of hospital stay (days)	18.0 (10.0–28.0)	20.0 (12.0–29.0)	13.0 (7.0–22.0)	0.001

Data are expressed as median (IQR), n (%).

Boldface p-value is statistically significant ($p < 0.05$).

ARDS acute respiratory distress syndrome, CHD coronary heart disease, COPD chronic obstructive pulmonary disease, ICU intensive care unit, OSAS obstructive sleep apnea syndrome, PaO₂/FiO₂ arterial partial pressure in oxygen/ fraction of inspired oxygen, SAPS II simplified acute physiology score II, SOFA sequential organ failure assessment.2 2 arterial partial pressure in oxygen/ fraction of inspired oxygen, SAPS II simplified acute physiology score II, SOFA sequential organ failure assessment.

^aARDS severity was defined as a PaO₂/FiO₂ ratio of 300 to 201, 200 to 101, and ≤100 mm Hg for mild, moderate, and severe ARDS, respectively.

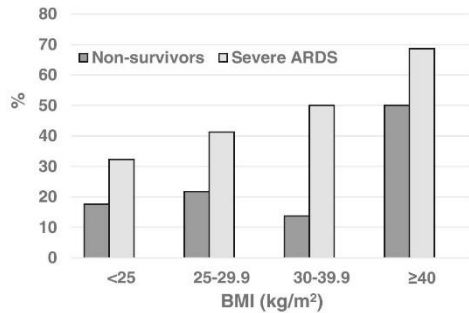


Fig. 1 Mortality rate (%) and prevalence of severe ARDS (%) according to BMI (kg/m²) categories. Severe ARDS was defined based on a PaO₂/FiO₂ ≤ 100 mmHg.

overweight and normal BMI patients but both were non-significant. These results supported pooling the normal and overweight BMI categories.

Multivariable analysis was performed for models A and B using these 3 BMI categories (normal weight or overweight, moderate obesity (BMI 30 to <40 kg/m²), and severe obesity (BMI ≥ 40 kg/m²). In model A, logistic regression analysis showed that age (OR 1.03, 95%CI 1.01–1.12), COPD chronic obstructive pulmonary disease (OR 13.01, 95%CI 3.24–52.28), the presence of severe hypoxemia (PaO₂/FiO₂ ≤ 100 vs. >100 mm Hg: OR 3.73, 95%CI 1.55–8.95), and SAPS II score (OR 1.03, 95%CI 1.01–1.06) were variables independently associated with in-hospital mortality. Due to the small number of patients with some co-morbidities, the number of co-morbidities was introduced in model B. Logistic regression analysis identified BMI ≥ 40 kg/m², number of co-morbidities (≥2), lowest PaO₂/FiO₂, and SAPS II score, as independent predictors of in-hospital mortality (Table 3). As BMI categories were related to in-hospital mortality in a J-shaped manner with a nadir at BMI > 29 to 39 kg/m², these cut-off points for the definition of moderate obesity were used in model C. The mortality rate decreased to 13% for patients with BMI > 29 to 39 kg/m², whereas it was 21.8% for patients with BMI ≤ 29 and 45% for those with BMI > 39 kg/m² ($P = 0.002$). This specific BMI classification was added to model C as a covariate with age, the number of comorbidity categories, SAPS II, and PaO₂/FiO₂. The logistic regression model revealed a paradoxical association between BMI categories and mortality, with a significantly higher risk among those with a BMI ≤ 29 (OR 3.64, 95% CI 1.38–9.60) and those with a BMI > 39 kg/m² (OR 10.45, 95% CI 2.45–41.09), compared to those with BMI > 29 to 39 kg/m². The 2 or more comorbidities categories, SAPS II score, and PaO₂/FiO₂ < 100 mm Hg remained significant predictors for in-hospital mortality.

The same paradoxical association remained for 30-day mortality, but the effect was significant only for BMI > 39 Category (Supplementary Table 2).

Discussion

To our knowledge, this is the first study reporting the absence of an increase in mortality in COVID-19 patients with moderate obesity admitted to the ICU and even a significant survival benefit in the optimal BMI range (BMI 29–39 kg/m²) for this population, after full adjustment for confounding factors. Patients with severe obesity (BMI ≥ 40 kg/m²) and patients with normal BMI or who were overweight, had a higher mortality rate, suggesting that the BMI-associated mortality curve is J-shaped. This unexpected finding has been termed the Obesity–ARDS paradox [32]. Older age, the high number of comorbidities, in particular COPD, the severity of ARDS, and SAPS II severity score were other independent predictors of in-hospital mortality.

Patients with obesity and a severe form of COVID 19 were over-represented in the ICU [5, 33]. We noted a high prevalence of obesity (BMI ≥ 30 kg/m²) of 43.3%, and severe obesity (7.3%), whereas patients with obesity usually account for 20% of intensive care inpatients [34, 35]. In another French study, the prevalence of obesity in critically ill COVID-19 patients was almost 3 times higher than in critical patients without COVID-19, after standardization for age and sex [8]. According to the ICNARC report, 31.3% and 7.9% of patients admitted to ICUs with confirmed COVID-19 had a BMI ≥ 30 kg/m² or a BMI ≥ 40 kg/m² compared with 28.9% and 2.9% of the general population, respectively, after adjusting for age and sex [3, 25].

Obesity may impact COVID-19 severity according to several different mechanisms including abnormal ventilation, low-grade chronic inflammation with adipose-lung cell crosstalk, impaired immune response to viral infection, endothelial dysfunction, local biological effects of ectopic fat deposition (in visceral, mediastinal, and epicardial tissues and even in the alveolar space) and extensive coagulopathy [2, 4, 18, 32, 36, 37]. Moreover, obesity increases the risk of numerous comorbidities considered to be risk factors for severe complications of COVID-19, in particular, hypertension, cardiovascular disease, type 2 diabetes, non-alcoholic fatty liver disease, and obstructive sleep apnea, [3, 4]. Some were overrepresented in our critically ill patients and were included in the regression analysis. Diabetes has been identified as a risk factor for worse outcomes in COVID-19 patients [38]. However, this was not the case in our ICU study, probably because of regular blood glucose monitoring and the use of insulin. In our hands, only one pre-existing disease, COPD was an independent predictor of in-hospital mortality after adjustment for

Table 2 Characteristics of patients according to their BMI status.

	Normal weight (n = 34)	Overweight (n = 92)	Moderate obesity (n = 80)	Severe obesity (n = 16)	p-value
Age at admission	62.0 (54–73.5)	66.0(58.8–72.0)	62.0 (54.0–71.0)	60.5 (50.0–66.8)	0.255
18–45 years	3 (9.1)	7 (7.6)	7 (8.9)	2 (12.5)	0.296
45–54 years	8 (24.2)	9 (9.8)	14 (17.7)	3 (18.8)	
55–64 years	10 (30.3)	21 (22.8)	28 (35.4)	4 (25.0)	
65–74 years	6 (18.2)	39 (42.4)	22 (27.8)	5 (31.2)	
75 years and more	6 (18.2)	16 (17.4)	8 (10.1)	2 (12.5)	
Male sex	25 (73.5)	67 (72.8)	58 (72.5)	10 (62.5)	0.850
Diabetes mellitus	8 (23.5)	10 (10.9)	31 (38.8)	10 (62.5)	<0.001
Hypertension	8 (23.5)	43 (46.7)	47 (58.8)	10 (62.5)	0.004
Dyslipidemia	9 (26.5)	28 (30.4)	34 (42.5)	8 (50.0)	0.144
Chronic pulmonary disease	4 (11.8)	10 (10.9)	21 (26.2)	7 (43.8)	0.003
including COPD	0 (0.0)	5 (5.4)	8 (10.0)	1 (6.2)	0.235
including OSAS	1 (2.9)	4 (4.3)	12 (15.0)	5 (31.2)	0.001
Coronary heart disease	5 (14.7)	10 (10.9)	14 (17.5)	2 (12.5)	0.657
Chronic kidney disease	2 (5.9)	6 (6.5)	4 (5.0)	3 (18.8)	0.251
Number of comorbidities					0.011
0	15 (44.1)	30 (32.6)	22 (27.5)	4 (25.0)	
1	11 (32.4)	28 (30.4)	15 (18.8)	2 (12.5)	
2	2 (5.9)	23 (25.0)	19 (23.8)	3 (18.8)	
3 and more	6 (17.6)	11 (12.0)	24 (30.0)	7 (43.8)	
ARDS severity ^a					0.106
Mild	4 (11.8)	3 (3.3)	5 (6.2)	1 (6.2)	
Moderate	19 (55.9)	51 (55.4)	35 (43.8)	4 (25.0)	
Severe	11 (32.3)	38 (41.3)	40 (50.0)	11 (68.8)	
PaO ₂ /FiO ₂ (mmHg)	116.0 (86.0–170.0)	111.0 (82.0–142.8)	101.5 (74.8–141.0)	87.0 (69.8–108.0)	0.032
Mechanical ventilation	25 (73.5)	77 (83.7)	66 (82.5)	15 (93.8)	0.337
Mechanical ventilation duration (days)	13.0 (8.0–22.0)	11.5 (6.0–17.0)	13.0 (9.0–18.0)	11.0 (8.0–21.0)	0.289
Mechanical ventilation duration for survivor only (days) ^b	9.0 (7.5–20.0)	11.0 (6.0–16.0)	14.0 (9.0–19.2)	21.0 (10.5–24.0)	0.233
Hydroxychloroquine	2 (5.9)	14 (15.6)	15 (18.8)	0 (0.0)	0.136
Antiviral therapy	9 (26.5)	14 (15.4)	13 (16.2)	6 (37.5)	0.114
Corticosteroids	6 (17.6)	23 (25.0)	19 (23.8)	5 (31.2)	0.736
SAPS II	50.0 (25.5–62.0)	45.0 (31.0–61.3)	43.0 (33.5–62.0)	46.5 (40.3–62.0)	0.989
SOFA score	6.5 (3.3–9.8)	7.0 (4.0–8.0)	7.0 (4.0–8.03.91)	7.0 (5.8–12.0)	0.139
ICU length of stay (days)	14.5 (8.3–21.8)	12.0 (4.0–20.3)	13.5 (6.8–23.0)	13.0 (7.0–22.5)	0.394
ICU length of stay for survivors only (days) ^c	11.0 (7.5–18.0)	12.0 (4.0–20.3)	15.0 (7.0–23.0)	22.50 (13.0–27.8)	0.196
30-day death ^d	4 (11.8)	20 (21.7)	12 (15.6)	8 (50.0)	0.009
In-hospital death	6 (17.6)	20 (21.7)	11 (13.8)	8 (50.0)	0.011
Length of hospital stay (days)	22.0 (14.5–34.8)	16.0 (8.0–26.0)	18.0 (12.5–27.0)	16.5 (7.0–27.8)	0.085
Length of hospital stay (days) for survivors ^e	21.0 (12.8–35.0)	18.0 (9.0–27.3)	23.0 (13.0–29.0)	29.5 (19.3–43.0)	0.070

Data are expressed as median (IQR) or n (%).

Boldface p-value is statistically significant ($p < 0.05$).

ARDS acute respiratory distress syndrome, COPD chronic obstructive pulmonary disease, ICU intensive care unit, OSAS: obstructive sleep apnea syndrome, PaO₂/FiO₂ arterial partial pressure in oxygen/ fraction of inspired oxygen, SAPS II simplified acute physiology score II, SOFA sequential organ failure assessment.

^aARDS severity was defined as a PaO₂/FiO₂ ratio of 300 to 201, 200 to 101, and ≤ 100 mm Hg for mild, moderate, and severe ARDS, respectively.

^bSurvivors, $n = 141$.

^cSurvivors $n = 177$.

^dData available $n = 219$.

Table 3 Factors associated with in-hospital mortality.

Model A			Model B			Model C		
	OR	95%CI		OR	95%CI		OR	95%CI
<i>BMI categories</i>			<i>BMI categories</i>			<i>BMI categories</i>		
Normal/overweight	2.37	(0.84–6.73)	Normal/overweight	2.38	(0.94–6.06)	29 kg/m ² and less	3.03	(1.19–7.72)
Moderate obesity	1		Moderate obesity	1		29.1–39 kg/m ²	1	
Severe obesity	12.25	(2.53–59.45)	Severe obesity	7.67	(1.78–33.03)	More than 39 kg/m ²	6.74	(1.75–25.82)
Age at admission	1.06	(1.01–1.12)	Age at admission	1.05	(0.97–1.10)	Age at admission	1.05	(0.99–1.10)
Diabetes mellitus	0.56	(0.18–1.72)	<i>Number of comorbidities</i>			<i>Number of comorbidities</i>		
Hypertension	1.16	(0.45–3.01)	0	1		0	1	
Dyslipidemia	1.84	(0.70–4.84)	1	2.17	(0.61–7.78)	1	2.33	(0.65–8.37)
Coronary heart disease	0.92	(0.26–3.28)	2	3.71	(1.05–13.04)	2	4.24	(1.18–15.28)
COPD	12.94	(3.22–51.97)	3 or more	4.45	(1.22–16.17)	3 or more	4.89	(1.33–18.05)
Chronic kidney disease	3.14	(0.61–16.08)						
PaO ₂ /FiO ₂ ≤100 mmHg	3.75	(1.56–8.99)	PaO ₂ /FiO ₂ ≤100 mmHg	3.99	(1.74–9.18)	PaO ₂ /FiO ₂ ≤100 mmHg	4.26	(1.81–9.96)
SAPS II	1.03	(1.01–1.06)	SAPS II	1.04	(1.01–1.06)	SAPS II	1.04	(1.01–1.06)

OR odds ratio, 95%CI 95% confidence interval, COPD chronic obstructive pulmonary disease, PaO₂/FiO₂ arterial partial pressure in oxygen/fraction of inspired oxygen, SAPS II simplified acute physiology score II.

confounding factors. The role of COPD [26, 39], coronary artery disease [28], active cancer [28], hypercholesterolemia, and diabetes [39] has been demonstrated in previous ICU studies.

Most of our patients admitted to the ICU had moderate (49.1%) to severe ARDS (45%). The proportion of severe ARDS reached 68.7% for patients with severe obesity and 50% for those with moderate obesity. As in previous COVID-19 ICU studies, a low PaO₂/FiO₂ was associated with a higher death rate [27, 28, 39]. We reported a high rate of mechanical ventilation (82.4%) compared to others studies [Simmonet 68.6% [16], Gupta 67.4% [28], Auld 76% [27], Cummings 79% [26] because of the severity of our patient's illness: the median PaO₂/FiO₂ of 109 mm Hg (IQR, 79–142) was lower than in other Covid-19 studies: 124 mm Hg (IQR, 86–166) in the Gupta study [28], 129 mm Hg (IQR, 80–203) in the Cummings study [26] and 132 mm Hg (IQR, 100–178) in the Auld study [27]. Despite these figures, the in-hospital and 30-day mortality rates in this study (20.3% and 20.1%, respectively) were lower than that reported in a US cohort (35.4% for 28-day mortality) [28] and close to that for critically ill patients with COVID-19 in Lombardy (26%) [39]. An in-ICU mortality rate of 41.6% was found in a systematic review and meta-analysis of ICU outcomes in patients with COVID-19 across 24 international studies [29]. The authors underlined that reported mortality rates fell from above 50% in March 2020 to close to 40% at the end of May 2020. This change is probably due to differences in the case-mix or treatment and care protocol [28, 29, 39].

In this study, severe obesity was associated with a higher risk of death, supporting a recently published US

multicenter cohort study [28]. Patients with BMI ≥ 40 kg/m² had a higher ICU mortality rate than their normal-weight counterparts, after adjusting for confounding factors, in particular the PaO₂/FiO₂ and liver or renal component of the SOFA score [28]. On the contrary, in a cohort of 257 ICU patients admitted for SARS-CoV-2, severe obesity was found to be associated with mortality by univariate analysis, but this relation disappeared after multivariate analysis [26]. A meta-analysis found that patients with H1N1 influenza A (H1N1) infection affected by severe obesity were twice as likely to be admitted to the ICU or die, then H1N1 patients with BMI < 40 kg/m² [40].

Although our patients with mild to moderate obesity tend to develop more frequently severe ARDS, they were also more likely to recover once the acute lung insult resolved. These findings support the “obesity paradox” theory mentioned above, by which obesity increases the risk of obesity-related diseases and long-term mortality but is paradoxically associated with a survival benefit during some types of acute illness [23, 34, 41–43]. This was the case in our study in which the number of comorbidities was increased in patients with obesity. Validation of the obesity paradox requires that the inverse correlation between BMI and mortality persists, even when the analysis is corrected for confounding variables [23, 34]. The PaO₂/FiO₂ was the main confounding variable in our study, whereas severity scores and the number of comorbidities were less important. Using slightly different cut-offs for BMI reference category (>29 to 39 kg/m²) the multivariable analysis demonstrated that BMI ≥ 39 kg/m² and BMI < 29 kg/m² were independently associated with higher in-hospital mortality after adjusting for other key factors (high number of co-morbidities (≥2), higher SAPS II score,

lower PaO₂/FiO₂). This suggests that mild to moderate obesity may have a protective effect in critical illness and in particular ARDS.

Many experts consider that COVID-19 challenges the obesity paradox [2, 6, 23, 24] because observational cohort study studies demonstrated a strong association between obesity and COVID-19 in-hospital morbidity and mortality. However, most clinical studies or meta-analyses published to date have analyzed data for hospitalized COVID-19 patients [6, 10, 12–14, 22]. To our knowledge, this is the first study to address the role of obesity specifically in critically ill patients with ARDS due to COVID-19. Our findings are consistent both with the study of Goyal et al. [12] and with previous studies in patients with ARDS outside COVID-19 [44]. Goyal et al. [12] reported a J-shaped pattern for in-hospital mortality among 1687 hospitalized patients in 2 New York City hospitals, although the statistically significant difference in death rate according to BMI classes disappeared in those with respiratory failure; the fully adjusted hazard ratio (HR) for death (normal BMI as reference) was lowest in patients with a BMI of 25–40 kg/m²: overweight HR 0.76, 95% CI 0.52–1.12; mild to moderate obesity HR 0.82, 95% CI 0.53–1.27; vs. severe obesity HR 1.29, 95% CI 0.58–2.86. A J-shaped distribution in mortality risk according to BMI has previously been observed in patients with ARDS [44]. A large meta-analysis evaluated the effect of obesity on ICU morbidity and mortality and demonstrated that there was no change in mortality for patients with obesity, despite their increased morbidity [41, 42]. A meta-analysis of 6 studies (*n* = 7165) found that patients with overweight or obesity admitted to the ICU with septic shock, excluding patients with BMI ≥ 40 kg/m², had a reduced risk of mortality compared to patients with normal BMI [43]. This reduced mortality for patients with obesity was also reported in the subgroup of ARDS patients [45, 46].

This survival paradox's actual existence and mechanisms in critical patients with obesity are controversial [23, 32, 34, 43]. The J or U-shaped relationship between BMI and outcome during acute disease may simply be due to increased mortality at the extremes (i.e., underweight or severe obesity). Moderate obesity may therefore reflect relatively good health. Moreover, the paradoxical association may be mediated by other confounding factors such as deprivation, malnutrition, smoking status, and by unmeasured variables such as the effect of fluid retention on BMI classification that are responsible for a poorer or a better outcome in some BMI groups. Surprisingly, there were no underweight patients in this study. Patients with obesity might be admitted to the ICU at a different level of severity or because of different admission criteria to those for patients without obesity. This could be due to a perceived difference in risk for the subjects with obesity or due to real

difficulties in providing adequate care in the ward setting [44]. Disparities in care processes may lead to triage of patients with obesity to higher levels and standards of care due to their higher physical care requirements, which could have a favorable impact on outcomes [41]. There may also be a “selective survivor” effect, in which highest-risk patients with obesity and many comorbidities die before hospital admission.

Despite multiple chronic morbidities, this unexpected survival of critically ill patients with obesity may be explained by several reasons [23, 32, 34]. Excess adipose tissue and increased fat-free mass provide additional metabolic reserves during the catabolic state. Obesity is associated with increased activity of the renin-angiotensin system contributing to hypertension, but it may also have protective hemodynamic effects in the ICU due to a decreased need for fluid or vasopressor support which in excess can adversely impact outcomes [43]. Finally, obesity may constitute a preconditioned chronic pro-inflammatory status, creating a protective environment that limits the detrimental effects of a more aggressive second hit during critical illness [32].

Our study has both methodological strengths and limitations. By focusing on critically ill COVID-19 patients with ARDS, we reduced selection bias by restricting the study population. All medical data were reviewed twice to guarantee their quality. Only 2 patients were excluded for missing data. The admission weight ignored any recent weight change, in particular unintentional weight loss before ICU admission. Height measurements are often inaccurate in the ICU and this can result in misclassifications in the BMI. As this bias is the same for survivors and non-survivors, it should not affect the interpretation of our results. Our small sample size limited the significance of the association between in-hospital and 30-day mortality and standard BMI categories, therefore our results need to be confirmed in a larger study. However, the same J-shaped curve was observed for both in-hospital and 30-day mortality. Smoking status, a healthy or unhealthy metabolic profile [21], and the potential role of excess visceral adipose tissue [47] could not be taken into account due to the lack of data.

Conclusions

Our data showed that patients treated at the ICU with moderate obesity survived more often than patients with normal BMI, overweight or severe obesity and suggest that the ARDS obesity paradox is not broken by COVID-19. Moderate obesity, although potentially associated with many chronic morbidities, may have a protective effect during critical illness and in particular ARDS. This unexpected increase in survival indicates that the management of

severe COVID-19 is not compromised by challenges in diagnosis and treatment caused by the physical effects of obesity, at least for a BMI below 40 kg/m². ICU admission and aggressive treatment should therefore be encouraged because of their good potential outcome.

Obesity must still be considered as a potential risk factor for severe adverse outcomes, including mortality. Our findings should not be interpreted to suggest that obesity per se has a protective effect on COVID-19 as this could give a false safety message to people with moderate obesity, especially during the reopening phase. The association between obesity and COVID-19 disease severity underlines the need for sustained long-term actions to address obesity [1, 3, 5].

Acknowledgements We would like to thank all members of Nancy Department of Anesthesiology and Intensive Care Medicine for helpful suggestions and discussions.

Author contributions RD: acquisition, analysis, or interpretation of data, drafting of the manuscript, and administrative, technical, or material support. RD and AB contributed equally to this work. AB: full access to all of the data in the study and responsibility for the integrity of the data and the accuracy of the data analysis, concept and design, methodology, acquisition, analysis, or interpretation of data, statistical analysis, drafting of the manuscript. PB and CZ: acquisition, analysis, or interpretation of data, administrative, technical, or material support, and critical revision of the manuscript for important intellectual content. M-RL, BL, and SG: concept and design, critical revision of the manuscript for important intellectual content. GA: Full access to all of the data in the study and responsibility for the integrity of the data and the accuracy of the data analysis, concept and design, acquisition, analysis, or interpretation of data, drafting of the manuscript, obtained funding, study supervision, administrative, technical, or material support. OZ: Full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis, concept and design, acquisition, analysis, or interpretation of data, statistical analysis, drafting of the manuscript, obtained funding, study supervision, and administrative, technical, or material support.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Clément K, Coupaye M, Laville M, Oppert JM, Ziegler O. COVID-19: a lever for the recognition of obesity as a disease? The French experience. *Obesity*. 2020. <https://doi.org/10.1002/oby.22924>.
- Lavie CJ, Sanchis-Gomar F, Henry BM, Lippi G. COVID-19 and obesity: links and risks. *Expert Rev Endocrinol Metab*. 2020;15:215–6.
- Public Health England. Excess weight and COVID-19: insights from new evidence 2020 July. <https://www.gov.uk/government/publications/excess-weight-and-covid-19-insights-from-new-evidence>.
- Sanchis-Gomar F, Lavie CJ, Mehra MR, Henry BM, Lippi G. Obesity and outcomes in COVID-19: when an epidemic and pandemic collide. *Mayo Clin Proc*. 2020;95:1445–53.
- Popkin BM, Du S, Green WD, Beck MA, Algaith T, Herbst CH, et al. Individuals with obesity and COVID-19: a global perspective on the epidemiology and biological relationships. *Obes Rev*. 2020. <https://doi.org/10.1111/obr.13128>.
- Anderson MR, Geleris J, Anderson DR, Zucker J, Nobel YR, Freedberg D, et al. Body mass index and risk for intubation or death in SARS-CoV-2 infection: a retrospective cohort study. *Ann Intern Med*. 2020. <https://doi.org/10.7326/m20-3214>.
- Cai Q, Chen F, Wang T, Luo F, Liu X, Wu Q, et al. Obesity and COVID-19 severity in a designated hospital in Shenzhen, China. *Diabetes Care*. 2020;43:1392–8.
- Caussy C, Pattou F, Wallet F, Simon C, Chalopin S, Telliam C, et al. Prevalence of obesity among adult inpatients with COVID-19 in France. *Lancet Diabetes Endocrinol*. 2020;8:562–4.
- Czernichow S, Beeker N, Rives-Lange C, Guerot E, Diehl JL, Katsahian S, et al. Obesity doubles mortality in patients hospitalized for SARS-CoV-2 in Paris hospitals, France: a cohort study on 5795 patients. *Obesity*. 2020. <https://doi.org/10.1002/oby.23014>.
- Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ*. 2020;369:m1985.
- Gao F, Zheng KI, Wang XB, Sun QF, Pan KH, Wang TY, et al. Obesity is a risk factor for greater COVID-19 severity. *Diabetes Care*. 2020;43:e72–e74.
- Goyal P, Ringel JB, Rajan M, Choi JJ, Pinheiro LC, Li HA, et al. Obesity and COVID-19 in New York City: a retrospective cohort study. *Ann Intern Med*. 2020. <https://doi.org/10.7326/m20-2730>.
- Hajifathalian K, Kumar S, Newberry C, Shah S, Fortune B, Krisko T, et al. Obesity is associated with worse outcomes in COVID-19: analysis of early data from New York City. *Obesity*. 2020. <https://doi.org/10.1002/oby.22923>.
- Klang E, Kassim G, Soffer S, Freeman R, Levin MA, Reich DL. Morbid obesity as an independent risk factor for COVID-19 mortality in hospitalized patients younger than 50. *Obesity*. 2020. <https://doi.org/10.1002/oby.22913>.
- Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ*. 2020;369:m1966.
- Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, et al. High prevalence of obesity in severe acute respiratory syndrome Coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity*. 2020;28:1195–9.
- Földi M, Farkas N, Kiss S, Zádori N, Váncsa S, Szakó L, et al. Obesity is a risk factor for developing critical condition in COVID-19 patients: a systematic review and meta-analysis. *Obes Rev*. 2020;21:e13095.
- Kang Z, Luo S, Gui Y, Zhou H, Zhang Z, Tian C, et al. Obesity is a potential risk factor contributing to clinical manifestations of COVID-19. *Int J Obes*. 2020. <https://doi.org/10.1038/s41366-020-00677-2>.
- Nakeshbandi M, Maini R, Daniel P, Rosengarten S, Parmar P, Wilson C, et al. The impact of obesity on COVID-19 complications: a retrospective cohort study. *Int J Obes*. 2020;44:1832–7.
- Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and mortality among black patients and white patients with Covid-19. *N Engl J Med*. 2020;382:2534–43.
- Xie J, Zu Y, Alkhatib A, Pham TT, Gill F, Jang A, et al. Metabolic syndrome and COVID-19 mortality among adult Black patients in New Orleans. *Diabetes Care*. 2020. <https://doi.org/10.2337/dc20-1714>.

22. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature*. 2020. <https://doi.org/10.1038/s41586-020-2521-4>.
23. Ades PA, Savage PD. The obesity paradox: perception vs knowledge. *Mayo Clin Proc*. 2010;85:112–4.
24. Jose RJ, Manuel A. Does Coronavirus disease 2019 disprove the obesity paradox in acute respiratory distress syndrome? *Obesity*. 2020;28:1007.
25. ICNARC. ICNARC report on COVID-19 in critical care 2020. 2020. <https://www.icnarc.org/DataServices/Attachments/Download/a9875849-f16c-ea11-9124-00505601089b>.
26. Cummings MJ, Baldwin MR, Abrams D, Jacobson SD, Meyer BJ, Balough EM, et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet*. 2020;395:1763–70.
27. Auld SC, Caridi-Scheible M, Blum JM, Robichaux C, Kraft C, Jacob JT, et al. ICU and ventilator mortality among critically ill adults with Coronavirus disease 2019. *Crit Care Med*. 2020. <https://doi.org/10.1097/ccm.0000000000004457>.
28. Gupta S, Hayek SS, Wang W, Chan L, Mathews KS, Melamed ML, et al. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. *JAMA Intern Med*. 2020. <https://doi.org/10.1001/jamainternmed.2020.3596>.
29. Armstrong RA, Kane AD, Cook TM. Outcomes from intensive care in patients with COVID-19: a systematic review and meta-analysis of observational studies. *Anaesthesia*. 2020. <https://doi.org/10.1111/anae.15201>.
30. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996;22:707–10.
31. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993;270:2957–63.
32. Bustamante A, Repine J. Adipose-lung cell crosstalk in the obesity-ARDS paradox. *J Pulmon Resp Med*. 2013;3:144.
33. Földi M, Farkas N, Kiss S, Zádori N, Váncsa S, Szakó L, et al. Obesity is a risk factor for developing critical condition in COVID-19 patients: A systematic review and meta-analysis. *Obes Rev*. 2020 <https://doi.org/10.1111/obr.13095>.
34. Ball L, Serpa Neto A, Pelosi P. Obesity and survival in critically ill patients with acute respiratory distress syndrome: a paradox within the paradox. *Crit Care*. 2017;21:114.
35. De Jong A, Verzilli D, Jaber S. ARDS in obese patients: specificities and management. *Crit Care*. 2019;23:74.
36. Sattar N, McInnes IB, McMurray JJV. Obesity is a risk factor for severe COVID-19 infection: multiple potential mechanisms. *Circulation*. 2020;142:4–6.
37. Rebello CJ, Kirwan JP, Greenway FLObesity. the most common comorbidity in SARS-CoV-2: is leptin the link? *Int J Obes*. 2020;44:1810–7.
38. Zhu L, She ZG, Cheng X, Qin JJ, Zhang XJ, Cai J, et al. Association of blood glucose control and outcomes in patients with COVID-19 and pre-existing type 2 diabetes. *Cell Metab*. 2020;31:1068–e3.
39. Grasselli G, Greco M, Zanella A, Albano G, Antonelli M, Bellani G, et al. Risk factors associated with mortality among patients with COVID-19 in intensive care units in Lombardy, Italy. *JAMA Intern Med*. 2020. <https://doi.org/10.1001/jamainternmed.2020.3539>.
40. Fezeu L, Julia C, Henegar A, Bitu J, Hu FB, Grobbee DE, et al. Obesity is associated with higher risk of intensive care unit admission and death in influenza A (H1N1) patients: a systematic review and meta-analysis. *Obes Rev*. 2011;12:653–9.
41. Acharya P, Upadhyay L, Qavi A, Naaraayan A, Jesmajian S, Acharya S, et al. The paradox prevails: outcomes are better in critically ill obese patients regardless of the comorbidity burden. *J Crit Care*. 2019;53:25–31.
42. Akinusi ME, Pineda LA, El, Solh AA. Effect of obesity on intensive care morbidity and mortality: a meta-analysis. *Crit Care Med*. 2008;36:151–8.
43. Pepper DJ, Sun J, Welsh J, Cui X, Suffredini AF, Eichacker PQ. Increased body mass index and adjusted mortality in ICU patients with sepsis or septic shock: a systematic review and meta-analysis. *Crit Care*. 2016;20:181.
44. O'Brien JM Jr, Phillips GS, Ali NA, Lucarelli M, Marsh CB, Lemeshow S. Body mass index is independently associated with hospital mortality in mechanically ventilated adults with acute lung injury. *Crit Care Med*. 2006;34:738–44.
45. Ni YN, Luo J, Yu H, Wang YW, Hu YH, Liu D, et al. Can body mass index predict clinical outcomes for patients with acute lung injury/acute respiratory distress syndrome? A meta-analysis. *Crit Care*. 2017;21:36.
46. Zhi G, Xin W, Ying W, Guohong X, Shuying L. "Obesity Paradox" in acute respiratory distress syndrome: a systematic review and meta-analysis. *PLoS ONE*. 2016;11:e0163677.
47. Battisti S, Pedone C, Napoli N, Russo E, Agnoletti V, Nigra SG, et al. Computed tomography highlights increased visceral adiposity associated with critical illness in COVID-19. *Diabetes Care*. 2020. <https://doi.org/10.2337/dc20-1333>.

4.3.4 Synthèse de l'étude 6

En termes cliniques, cette étude a montré une association significative entre l'IMC et la mortalité intrahospitalière des patients atteints de COVID-19 pris en charge en réanimation pour une détresse respiratoire aiguë. La définition des seuils d'IMC à considérer n'est toujours pas tranchée dans la littérature en 2021. Dichotomiser l'IMC à partir du seuil d'obésité classique ($\geq 30 \text{kg/m}^2$) n'est probablement pas une approche satisfaisante (76,77). Nous avons proposé d'autres seuils témoignant d'une probable relation en J de l'IMC sur la mortalité.

En termes de méthodes, le ciblage des patients et le recueil de données ont été grandement facilités par le recours aux données médico-administratives. Les unités de réanimation ont été ciblées via leurs autorisations évoluant chaque jour pour répondre à l'urgence de la crise. Les séjours de patients pris en charge pour COVID—19 ont été identifiés dans les données médico-administratives locales grâce au marquage quotidien de leur RUM par des codes propres documentaires. Les données sociodémographiques (âge, sexe), la durée de séjour selon les unités et le critère de jugement principal sur la mortalité hospitalière ont pu être directement extraits des données médico-administratives. Une approche complémentaire par un recueil spécifique restreint a permis de réaliser cette étude dans des délais de temps réduit et compatibles avec la réponse à une crise sanitaire. Cette étude a montré un autre usage des BMA qui peuvent accélérer le recrutement des sujets en facilitant le processus de ciblage et le recueil partiel de données.

4.4 Synthèse de l'axe 2

Dans cet axe de recherche, nos travaux avaient pour objectif d'illustrer les apports que pouvaient constituer les données issues des bases médico-administratives au travers de différents cas d'usage en pharmacovigilance et en épidémiologie.

Nous avons ainsi pu montrer la possibilité de réaliser des études de cohorte retraçant les parcours de soins des patients. Les données issues des BMA permettaient à ces cohortes de bénéficier de fenêtres d'exposition étendue (étude sur l'impact des comorbidités maternelles pré-existantes à la grossesse) ou encore d'un recrutement exhaustif de patients particulièrement intéressant dans le cas de pathologies rares (étude sur les patients atteints de MICI et d'un syndrome coronarien). La disponibilité des données issues de BMA nous a permis de souligner leur intérêt pour faciliter le recrutement de patients et le recueil de données (étude sur l'impact de l'IMC sur la mortalité intra-hospitalière des patients COVID-19 en réanimation).

Au travers de l'étude INSHARE nous avons pu mettre en évidence l'intérêt d'enrichir les données issues des bases médico-administratives par des données cliniques et biologiques issues d'un entrepôt de données hospitalières.

5 Discussion générale

5.1 Principaux résultats

Les bases médico-administratives ont une finalité première qui repose sur le financement des activités de soins. De ce fait les données qu'elles renferment souffrent de deux écueils majeurs :

- ✓ le point de vue du producteur de ces données qui conditionne le périmètre des données recueillies,
- ✓ le prisme du financement qui conditionne le type, la qualité, et la complétude des données recueillies (40).

Dans le premier axe de ce travail, nous avons pu montrer que des outils méthodologiques pouvaient apporter une réponse aux contraintes liées à la nature des données issues de bases médico-administratives et ainsi les enrichir pour répondre de façon plus pertinente à des questions de recherche. Ces outils méthodologiques convergent vers un même objectif : apporter une information complémentaire à celle pouvant être extraite des BMA, soit par l'apport de données supplémentaires, soit par des algorithmes permettant d'extraire une information utile qui n'était pas directement disponible. Nous l'avons montré dans différents travaux : l'adaptation et la validation du score de comorbidités de Charlson aux données du SNDS, la construction d'un proxy pour identifier une donnée absente des BMA tel que le caractère programmé ou non des hospitalisations, le développement et l'évaluation d'un algorithme d'appariement semi-déterministe permettant de chaîner des données de consommation de ville et hospitalière.

Dans la seconde partie de ce travail, nous avons illustré au travers de cas d'usage en pharmacovigilance et en épidémiologie la possibilité de reconstituer des parcours de soins avec une fenêtre d'exposition étendue, un recrutement exhaustif de larges cohortes et la complémentarité des données issues des dossiers médicaux avec celles des bases médico-administratives. Nous avons montré que, partant de questions de recherche concrètes, une réponse était possible via l'exploitation des données des BMA,

si une expertise appropriée par rapport aux données et aux méthodes d'exploitation des données pouvait être mobilisée par rapport aux questions soulevées.

5.2 Place des données issues des BMA en recherche

Parmi les arguments légitimant la réutilisation des données issues de bases médico-administratives, l'exhaustivité des données recueillies sur une population donnée et un moindre coût du recueil de ces données, effectué habituellement en routine à visée financière, est retrouvé dans la littérature (39).

L'exhaustivité de l'enregistrement des données d'hospitalisation dans le PMSI et des remboursements dans le DCIR permet de garantir un recrutement des sujets sans biais lié à la participation des sujets. Dans l'étude 1 sur l'adaptation du score de Charlson aux données du SNDS et dans l'étude 4 sur l'impact des comorbidités maternelles sur l'issue de grossesse, l'ensemble des sujets répondant à nos critères d'inclusion ont pu être inclus dans nos cohortes soit respectivement 6,6 millions de sujets hospitalisés en 2010 et 1,8 millions de femmes enceintes entre 2016 et 2018. La contrepartie de ce volume important de données peut nous placer dans une situation difficile en termes de temps d'analyse qui, dans le cas de l'étude 1 nous a poussé à réaliser des échantillonnages ou encore en termes de surpuissance statistique. En travaillant sur l'ensemble d'une population donnée, nous sortons du cadre d'une situation inférentielle habituelle nécessitant de mesurer la vraisemblance d'un nombre limité d'observations puisque toutes sont finalement observées. Toutefois le diktat des p-value est si profondément ancré dans la littérature médicale qu'il en devient pratiquement impossible de passer outre si on veut pouvoir publier, ce qui passe par la satisfaction des « exigences » des relecteurs.

Un autre avantage classiquement attribué aux données issues des BMA est la limitation des coûts liés à la collecte des données. Au travers du cas d'usage sur l'impact de l'IMC sur le pronostic des patients atteints de COVID-19 en réanimation, nous avons montré que les BMA pouvaient servir à cibler des sujets et extraire des données de façon accélérée en situation de crise sanitaire. Toutefois, deux bémols doivent être formulés sur le raccourcissement de ces délais de collecte. Les délais d'accès à ces

données sont aussi à prendre en compte. Concernant les données du SNDS, de la demande d'accès réglementaire au le Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) et à la CNIL à la mise à disposition des données, ce délai peut varier entre 12 et 24 mois. L'accès limité aux données du PMSI via le portail de l'ATIH est plus aisé, notamment pour les établissements de santé. Ceci nous a conduit à faire le choix de nous limiter aux données du PMSI dans le cadre de l'étude 4 sur l'impact des comorbidités maternelles sur les issues de grossesse. Il en a résulté une sous-estimation de la prévalence des pathologies maternelles et une sélection des pathologies les plus sévères requérant une hospitalisation. Par ailleurs, traiter des données issues de BMA nécessite un fort investissement en temps et en infrastructure. Il est nécessaire de se reposer sur une équipe formée (connaissance et accès aux données PMSI (78), formations et habilitation aux données SNDS (79)) évoluant au sein d'une infrastructure sécurisée répondant aux exigences d'un référentiel de sécurité (80) comme dans le cas d'usage que nous avons présenté dans l'étude 3 sur la plateforme INSHARE illustrée par la description des trajectoires de soins de patients consommateurs de statines. En effet, un degré élevé de compréhension de ces données est essentiel afin de mettre en œuvre une étude sur BMA. La première étape concerne le ciblage des sujets correspondant aux critères d'inclusion et d'exclusion. Il s'agit par exemple d'identifier les sujets nouvellement atteints d'une pathologie ou nouvellement utilisateurs d'un traitement. Pour ce faire, il faut s'interroger sur la visibilité de la pathologie out du traitement au travers du prisme des données SNDS (pathologies avec hospitalisation, et/ou avec un acte spécifique, et/ou avec un traitement spécifique) comme nous avons pu le proposer pour chacune des pathologies du score de comorbidité de Charlson. Dans la plupart des études, il est préférable de se focaliser sur des sujets incidents de la pathologie donnée ou des nouveaux utilisateurs pour éviter certains biais de sélection tels que le biais de survie sélective (81) ou de temps immortel (82). Il faut donc pouvoir exclure de façon certaine les cas prévalents, notamment en s'assurant d'avoir un recul suffisant dans les données disponibles. La seconde étape concerne l'extraction des données des sujets inclus. Une bonne connaissance de l'historique du recueil des données issues des BMA est indispensable. Si un certain nombre

d'informations sont disponibles dans la documentation, une part de la connaissance de ces données reste du domaine des producteurs de ces données. Par exemple, même si la latéralité des actes CCAM est recueillie de manière obligatoire depuis 2015, cette donnée n'est pas fiable car un certain nombre d'éditeurs ne l'ont pas tous intégré dans leurs solutions logicielles du dossier patients informatisé. Il est à noter que, dans le cadre de la mise en place de la Plateforme des données de santé, un effort a été réalisé sur la centralisation et la mise à disposition d'une documentation collaborative sur les données disponibles dans le SNDS (36) avec, en particulier, des fiches thématiques sur des points méthodologiques et un dictionnaire interactif des tables et variables disponibles.

Face à aux critiques communément formulées sur l'absence de données cliniques, nous avons montré l'intérêt de l'appariement d'une autre source de données aux données issues de BMA dans l'étude 3 sur les trajectoires de soins de patients consommateurs de statines et dans l'étude 6 sur l'impact de l'IMC sur le pronostic de patients atteints de COVID-19 en réanimation. L'apport de données cliniques issues de dossiers médicaux permet de combler l'angle mort des administrations médicamenteuses hospitalières. Toutefois, certaines données ne sont pas capturées par ces 2 sources de données, on peut ainsi citer les indications de traitements ou la réelle observance des prescriptions délivrées en ville par exemple. L'enrichissement par d'autres sources et le développement de proxy pourrait contribuer à diminuer ces manques dans les données disponibles. L'utilisation de données issues des objets connectés, comme les applications de suivi des sujets diabétiques, pourraient permettre de mieux évaluer l'observance des traitements.

Un autre reproche adressé aux données issues de BMA concerne leur recueil conditionné par leur finalité financière. Si le paradigme de n'utiliser les données que dans leur but de collecte (1^{ère} loi d'informatique médicale (83)) est, de nos jours, battu en brèche, le prisme du financement teinte les données disponibles dans les BMA (40,84). Dans l'étude 2 sur le développement d'un algorithme d'identification des hospitalisations non programmées chez les patients atteints de MICI, nous avons pu montrer que nous pouvions approcher la notion d'hospitalisation non programmée bien qu'elle soit

absente des données collectées dans le PMSI. Ceci nécessite toutefois le développement d'un algorithme qui peut être spécifique d'une pathologie donnée et variable dans le temps. Plusieurs travaux ont montré la nécessité d'outils méthodologiques pour approcher les pathologies dans les bases du PMSI (54,85) et plus largement du SNDS (86–88). Dans ce cadre, il faut citer l'effort du Réseau pour mieux utiliser les Données du Système National des Données de Santé (ReDSiam) (89) pour centraliser et développer ces algorithmes d'identification de pathologies alors que des critiques sur l'absence de transparence des algorithmes utilisés dans le SNDS ont été émises (90). Cependant, au-delà de la publication de ces algorithmes, la question de leur validité dans le temps se pose, en effet l'évolution des règles de codage des pathologies et leur conséquence sur le financement peuvent impacter le repérage de ces pathologies dans les BMA (40,91). C'est pourquoi la volonté d'ouverture de ces données à une large communauté d'utilisateurs (chercheurs, décideurs) ne devrait pas se faire au prix d'une déconnexion avec les producteurs de ces données (établissements de santé, assurance maladie) qui en détiennent les clefs de lecture au travers de leur historique.

6 Conclusion et perspectives

En conclusion, nos travaux ont permis de montrer l'apport des données issues des BMA françaises en épidémiologie et en pharmacovigilance. L'exploitation de ces données requiert différentes conditions : leur disponibilité, des outils méthodologiques adaptés et une connaissance poussée des conditions de leur recueil. Cette thèse a présenté à la fois des outils méthodologiques pour prendre en compte les spécificités des données issues des BMA françaises, mais également des cas d'usage en épidémiologie et en pharmacovigilance répondant à des questions de recherche clinique.

Ces travaux ont pu faire ressortir l'intérêt de la réutilisation des données issue de bases médico-administratives en recherche tout en soulignant les écueils qui pouvaient en résulter. L'utilisation secondaire de ce type de données nécessite en premier lieu de savoir ce qu'on peut en faire, mais également et surtout ce qu'on ne peut pas en faire.

Des perspectives en santé publique peuvent être évoquées concernant la mise en place d'indicateurs basés sur le SNDS que ce soit dans l'évaluation de la qualité des soins, le suivi des politiques de santé, ou la surveillance sanitaire. Un récent appel à projets de la Haute Autorité de Santé ouvre la voie au financement du développement d'indicateurs basés sur les données médico-administratives (92). Toutefois, la portée en est réduite par le champ limité aux indicateurs de résultat, et le montant proposé. La plateforme ScanSanté (93) qui met déjà à disposition un certain nombre d'indicateurs développés à partir des données du PMSI pourrait être étendue aux données du SNDS pour élargir le périmètre des indicateurs .

Dans les perspectives en termes de recherche, nous pouvons mentionner le développement de plateformes d'intégration de données de santé multisource. Nos travaux ont montré la complémentarité des données issues d'entrepôt hospitalier avec des données médico-administratives. En allant plus loin, ces plateformes pourraient être à la convergence des données issues de différents producteurs de données incluant le sujet lui-même grâce aux données issues des objets connectés. Multiplier les sources de données permettrait ainsi de s'affranchir de 2 des principaux obstacles des

données issues des bases médico-administratives : leur périmètre restreint et le prisme du financement.

7 Références

1. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. *J Epidemiol Community Health*. 1 mars 2014;68(3):283-7.
2. Peek N, Holmes JH, Sun J. Technical Challenges for Big Data in Biomedicine and Health: Data Sources, Infrastructure, and Analytics. *Yearb Med Inform*. 15 août 2014;9(1):42-7.
3. HCSP. Pour une meilleure utilisation des bases de données administratives et médico-administratives nationales pour la santé publique et la recherche [Internet]. Paris: Haut Conseil de la Santé Publique; 2012 mars [cité 14 mars 2021]. Disponible sur: <https://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=258>
4. Medicare | CMS [Internet]. [cité 12 juill 2021]. Disponible sur: <https://www.cms.gov/Medicare/Medicare>
5. Mues KE, Liede A, Liu J, Wetmore JB, Zaha R, Bradbury BD, et al. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled populations in the US. *Clin Epidemiol*. 9 mai 2017;9:267-77.
6. Medicaid.gov: the official U.S. government site for Medicare | Medicaid [Internet]. [cité 12 juill 2021]. Disponible sur: <https://www.medicaid.gov/>
7. Bright RA, Avorn J, Everitt DE. Medicaid data as a resource for epidemiologic studies: Strengths and limitations. *J Clin Epidemiol*. 1 janv 1989;42(10):937-45.
8. Williams BC, Mehr DR, Fries BE. Use of administrative records to describe longitudinal patterns of health services use among veterans. *J Am Med Inform Assoc JAMIA*. août 1994;1(4):325-7.
9. Department of Veterans Affairs Open Data Portal [Internet]. [cité 12 juill 2021]. Disponible sur: <https://www.data.va.gov/>
10. Maynard C, Chapko MK. Data resources in the Department of Veterans Affairs. *Diabetes Care*. mai 2004;27 Suppl 2:B22-26.
11. Terdiman J. The Kaiser Permanente National Research Database. [cité 12 juill 2021]; Disponible sur: <https://grantome.com/grant/NIH/RC2-LM010797-01>
12. Information & data resources [Internet]. Kaiser Permanente Department of Research & Evaluation. [cité 12 juill 2021]. Disponible sur: <https://www.kp-scalresearch.org/aboutus/resources-capabilities/information-and-data-resources/>
13. Doyle CM, Lix LM, Hemmelgarn BR, Paterson JM, Renoux C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf*. janv 2020;29 Suppl 1:68-77.
14. Clinical Practice Research Datalink | CPRD [Internet]. [cité 12 juill 2021]. Disponible sur: <https://www.cprd.com/>
15. Khan NF, Booth HP, Myles P, Mullett D, Gallagher A, Evans C, et al. Use of prescribing safety quality improvement reports in UK general practices: a qualitative assessment. *BMC Health Serv Res*. 27 avr 2021;21:394.

16. Padmanabhan S, Carty L, Cameron E, Ghosh RE, Williams R, Strongman H. Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. *Eur J Epidemiol.* 1 janv 2019;34(1):91-9.
17. Thygesen LC, Daasnes C, Thaulow I, Brønnum-Hansen H. Introduction to Danish (nationwide) registers on health and social issues: Structure, access, legislation, and archiving. *Scand J Public Health.* 1 juill 2011;39(7_suppl):12-6.
18. Holcman R. Le programme de médicalisation du système d'information (PMSI). In: *Management hospitalier* [Internet]. Paris: Dunod; 2015. p. 553-68. (Guides Santé Social; vol. 2e ed.). Disponible sur: <https://www.cairn.info/management-hospitalier--9782100724291-p-553.htm>
19. Loi n° 91-748 du 31 juillet 1991 portant réforme hospitalière. 91-748 juill 31, 1991.
20. Arrêté du 22 juillet 1996 relatif au recueil et au traitement des données d'activité médicale, visées à l'article L. 710-6 du code de la santé publique, par les établissements de santé visés à l'article L. 710-16-2 du même code et à la transmission, visée à l'article L. 710-7 du code de la santé publique, aux agences régionales de l'hospitalisation, aux organismes d'assurance maladie et à l'Etat d'informations issues de ce traitement - Légifrance [Internet]. Disponible sur: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000549743/>
21. Loi n° 2003-1199 du 18 décembre 2003 de financement de la sécurité sociale pour 2004 (1).
22. Loi n° 2007-1786 du 19 décembre 2007 de financement de la sécurité sociale pour 2008. 2007-1786 déc 19, 2007.
23. Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. *Med Care.* févr 1980;18(2 Suppl):iii, 1-53.
24. Arrêté du 30 novembre 2009 modifiant l'arrêté du 22 février 2008 modifié relatif au recueil et au traitement des données d'activité médicale et des données de facturation correspondantes, produites par les établissements de santé publics ou privés ayant une activité en médecine, chirurgie, obstétrique et odontologie, et à la transmission d'informations issues de ce traitement dans les conditions définies à l'article L. 6113-8 du code de la santé publique - Légifrance [Internet]. Disponible sur: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000021410850?r=7vRE1UxMfe>
25. Bienvenue sur les ressources documentaires de la plateforme des données hospitalières [Internet]. 2017 [cité 8 juill 2021]. Disponible sur: <https://portail-acces-secure.atih.sante.fr/>
26. Portail SNIIRAM [Internet]. [cité 8 juill 2021]. Disponible sur: <https://portail.sniiram.ameli.fr>
27. Circulaire DHOS/PMSI n° 2001-106 du 22 février 2001 relative au chaînage des séjours en établissements de santé dans le cadre du programme de médicalisation des systèmes d'information (PMSI) [Internet]. Disponible sur: <https://solidarites-sante.gouv.fr/fichiers/bo/2001/01-13/a0130896.htm>
28. Arrêté du 7 février 2011 modifiant l'arrêté du 22 février 2008 modifié relatif au recueil et au traitement des données d'activité médicale et des données de facturation correspondantes, produites par les établissements de santé publics ou privés ayant une activité en médecine, chirurgie, obstétrique et odontologie, et à la transmission d'informations issues de ce traitement dans les conditions définies à l'article L. 6113-8 du code de la santé publique -

Légifrance [Internet]. Disponible sur:
<https://www.legifrance.gouv.fr/loda/id/JORFTEXT000023603083/>

29. Arrêté du 24 décembre 2020 modifiant l'arrêté du 23 décembre 2016 relatif au recueil et au traitement des données d'activité médicale des établissements de santé publics ou privés ayant une activité en psychiatrie et à la transmission d'informations issues de ce traitement dans les conditions définies aux articles L. 6113-7 et L. 6113-8 du code de la santé publique et l'arrêté du 23 décembre 2016 relatif au recueil et au traitement des données d'activité médicale des établissements de santé publics ou privés ayant une activité en soins de suite et de réadaptation et à la transmission d'informations issues de ce traitement, dans les conditions définies aux articles L. 6113-7 et L. 6113-8 du code de la santé publique - Légifrance [Internet]. [cité 26 août 2021]. Disponible sur:
<https://www.legifrance.gouv.fr/loda/id/LEGIARTI000043047747/2021-01-07/>
30. Loi n° 98-1194 du 23 décembre 1998 de financement de la sécurité sociale pour 1999 (1).
31. Délibération de la Commission Nationale de l'Informatique et des Libertés. 01-054 oct 18, 2001.
32. Arrêté du 11 avril 2002 relatif à la mise en oeuvre du système national d'information interrégimes de l'assurance maladie.
33. Arrêté du 20 juin 2005 relatif à la mise en oeuvre du système national d'information interrégimes de l'assurance maladie.
34. Décret n° 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « système national des données de santé ». 2016-1871 déc 26, 2016.
35. Loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé. 2019-774 juill 24, 2019.
36. Documentation collaborative du SNDS | Documentation du SNDS [Internet]. [cité 8 juill 2021]. Disponible sur: <https://documentation-snds.health-data-hub.fr/>
37. Mitchell JB, Bubolz T, Paul JE, Pashos CL, Escarce JJ, Muhlbaier LH, et al. Using Medicare claims for outcomes research. *Med Care*. juill 1994;32(7 Suppl):JS38-51.
38. Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. *Clin Ther*. avr 1997;19(2):346-66.
39. Burgun A, Bernal-Delgado E, Kuchinke W, van Staa T, Cunningham J, Lettieri E, et al. Health data for public health: towards new ways of combining data sources to support research efforts in Europe. *Yearb Med Inform*. 2017;26(01):235-240.
40. Sarrazin MSV, Rosenthal GE. Finding pure and simple truths with administrative data. *JAMA*. 4 avr 2012;307(13):1433-5.
41. Goldberg M, Quantin C, Guéguen A. Bases de données médico-administratives et épidémiologie : intérêts et limites. 1 janv 2008;
42. Dussaucy A, Viel JF, Mulin B, Euvrard J. The framework Prospective Payment Information Systems: bias, sources of errors and consequences. *Rev Epidemiol Sante Publique*. 1994;42(4):345-58.

43. Collonnaz M, Bethune B, Weisslinger C, Faulon M, Fiore P, Goetz C. Determinants of time required by medical information technicians for quality control of hospital activity coding, in French medico-administrative system. *Rev Epidemiol Sante Publique*. juill 2019;67(4):213-21.
44. Bosco-Lévy P, Duret S, Picard F, Dos Santos P, Puymirat E, Gilleron V, et al. Diagnostic accuracy of the International Classification of Diseases, Tenth Revision, codes of heart failure in an administrative database. *Pharmacoepidemiol Drug Saf*. févr 2019;28(2):194-200.
45. Quantin C, Benzenine E, Hägi M, Auverlot B, Abrahamowicz M, Cottenet J, et al. Evaluation of medical information systems as a mean of identification of incident cases of colorectal cancer. *Sante Publique Vandoeuve--Nancy Fr*. févr 2014;26(1):55-63.
46. Pierron A, Revert M, Goueslard K, Vuagnat A, Cottenet J, Benzenine E, et al. Evaluation of the metrological quality of the medico-administrative data for perinatal indicators: A pilot study in 3 university hospitals. *Rev Epidemiol Sante Publique*. août 2015;63(4):237-46.
47. Avillach P, Joubert M, Fieschi M. Improving the quality of the coding of primary diagnosis in standardized discharge summaries. *Health Care Manag Sci*. juin 2008;11(2):147-51.
48. Djennaoui M, Ficheur G, Beuscart R, Chazard E. Improvement of the quality of medical databases: data-mining-based prediction of diagnostic codes from previous patient codes. *Stud Health Technol Inform*. 2015;210:419-23.
49. 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*. oct 2017;65 Suppl 4:S149-67.
50. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. *J Clin Epidemiol*. mars 2003;56(3):221-9.
51. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-83.
52. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol*. oct 1993;46(10):1075-9; discussion 1081-1090.
53. Chae J-W, Song CS, Kim H, Lee K-B, Seo B-S, Kim D-I. Prediction of mortality in patients undergoing maintenance hemodialysis by Charlson Comorbidity Index using ICD-10 database. *Nephron Clin Pract*. 2011;117(4):c379-384.
54. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. nov 2005;43(11):1130-9.
55. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol*. déc 2004;57(12):1288-94.

56. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol*. nov 1994;47(11):1245-51.
57. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *Am J Epidemiol*. 15 mars 2011;173(6):676-82.
58. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 28 févr 1996;15(4):361-87.
59. Ashton CM, Del Junco DJ, Soucek J, Wray NP, Mansur CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. *Med Care*. oct 1997;35(10):1044-59.
60. Gusmano M, Rodwin V, Weisz D, Cottenet J, Quantin C. Comparison of rehospitalization rates in France and the United States. *J Health Serv Res Policy*. janv 2015;20(1):18-25.
61. 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*. juill 2012;55(7):1953-62.
62. Soeiro T, Pradel V, Lacroix C, Lapeyre-Mestre M, Micallef J. Tendances nationales du comportement de doctor shopping pour l'oxycodone comme proxy de mésusage et d'abus potentiel. *Toxicol Anal Clin*. 1 déc 2020;32(4, Supplement):S36.
63. Bannay A, Chaignot C, Blotière P-O, Basson M, Weill A, Ricordeau P, et al. The Best Use of the Charlson Comorbidity Index With Electronic Health Care Database to Predict Mortality. *Med Care*. févr 2016;54(2):188-94.
64. Frostegård J. Atherosclerosis in patients with autoimmune disorders. *Arterioscler Thromb Vasc Biol*. sept 2005;25(9):1776-85.
65. Roifman I, Beck PL, Anderson TJ, Eisenberg MJ, Genest J. Chronic inflammatory diseases and cardiovascular risk: a systematic review. *Can J Cardiol*. avr 2011;27(2):174-82.
66. Wu G-C, Leng R-X, Lu Q, Fan Y-G, Wang D-G, Ye D-Q. Subclinical Atherosclerosis in Patients With Inflammatory Bowel Diseases: A Systematic Review and Meta-Analysis. *Angiology*. mai 2017;68(5):447-61.
67. Hatoum OA, Binion DG. The vasculature and inflammatory bowel disease: contribution to pathogenesis and clinical pathology. *Inflamm Bowel Dis*. mars 2005;11(3):304-13.
68. Ha C, Magowan S, Accortt NA, Chen J, Stone CD. Risk of arterial thrombotic events in inflammatory bowel disease. *Am J Gastroenterol*. juin 2009;104(6):1445-51.
69. Barnes EL, Beery RM, Schulman AR, McCarthy EP, Korzenik JR, Winter RW. Hospitalizations for Acute Myocardial Infarction Are Decreased Among Patients with Inflammatory Bowel Disease Using a Nationwide Inpatient Database. *Inflamm Bowel Dis*. sept 2016;22(9):2229-37.
70. Panhwar MS, Mansoor E, Al-Kindi SG, Sinh P, Katz J, Oliveira GH, et al. Risk of Myocardial Infarction in Inflammatory Bowel Disease: A Population-based National Study. *Inflamm Bowel Dis*. 4 mai 2019;25(6):1080-7.

71. Aniwaniwan S, Pardi DS, Tremaine WJ, Loftus EV. Increased Risk of Acute Myocardial Infarction and Heart Failure in Patients With Inflammatory Bowel Diseases. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc.* oct 2018;16(10):1607-1615.e1.
72. Anderson MR, Geleris J, Anderson DR, Zucker J, Nobel YR, Freedberg D, et al. Body Mass Index and Risk for Intubation or Death in SARS-CoV-2 Infection. *Ann Intern Med.* 17 nov 2020;173(10):782-90.
73. Popkin BM, Du S, Green WD, Beck MA, Algaith T, Herbst CH, et al. Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. *Obes Rev.* 2020;21(11):e13128.
74. Goyal P, Ringel JB, Rajan M, Choi JJ, Pinheiro LC, Li HA, et al. Obesity and COVID-19 in New York City: A Retrospective Cohort Study. *Ann Intern Med.* 17 nov 2020;173(10):855-8.
75. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *N Engl J Med.* 25 juin 2020;382(26):2534-43.
76. Friedman AN, Guirguis J, Kapoor R, Gupta S, Leaf DE, Timsina LR, et al. Obesity, Inflammatory and Thrombotic Markers, and Major Clinical Outcomes in Critically Ill Patients with COVID-19 in the US. *Obes Silver Spring Md.* 9 juin 2021;
77. Pouwels S, Ramnarain D, Aupers E, Rutjes-Weurding L, van Oers J. Obesity May Not Be Associated with 28-Day Mortality, Duration of Invasive Mechanical Ventilation and Length of Intensive Care Unit and Hospital Stay in Critically Ill Patients with Severe Acute Respiratory Syndrome Coronavirus-2: A Retrospective Cohort Study. *Med Kaunas Lith.* 29 juin 2021;57(7):674.
78. Accès aux données pour les établissements de santé, les chercheurs et les institutionnels | Publication ATIH [Internet]. [cité 12 juill 2021]. Disponible sur: <https://www.atih.sante.fr/acces-aux-donnees-pour-les-etablissements-de-sante-les-chercheurs-et-les-institutionnels>
79. Comment se former au traitement des données de santé ? [Internet]. Health Data Hub. [cité 12 juill 2021]. Disponible sur: <https://www.health-data-hub.fr/formations>
80. Arrêté du 22 mars 2017 relatif au référentiel de sécurité applicable au Système national des données de santé - Légifrance [Internet]. [cité 12 juill 2021]. Disponible sur: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000034265125/>
81. Neyman J. Statistics; servant of all sciences. *Science.* 2 sept 1955;122(3166):401-6.
82. Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: immortal time bias in observational studies. *Am J Respir Crit Care Med.* 1 juill 2003;168(1):49-53.
83. van der Lei J. Use and abuse of computer-stored medical records. *Methods Inf Med.* avr 1991;30(2):79-80.
84. Peek N, Rodrigues PP. Three controversies in health data science. *Int J Data Sci Anal.* 2018;6(3):261-9.

85. Couris CM, Forêt-Dodelin C, Rabilloud M, Colin C, Bobin J-Y, Dargent D, et al. Sensitivity and specificity of two methods used to identify incident breast cancer in specialized units using claims databases. *Rev Epidemiol Sante Publique*. avr 2004;52(2):151-60.
86. Bousquet P-J, Caillet P, Coeuret-Pellicer M, Goulard H, Kudjawu YC, Le Bihan C, et al. Using cancer case identification algorithms in medico-administrative databases: Literature review and first results from the REDSIAM Tumors group based on breast, colon, and lung cancer. *Rev Epidemiol Sante Publique*. oct 2017;65 Suppl 4:S236-42.
87. Fosse-Edorh S, Rigou A, Morin S, Fezeu L, Mandereau-Bruno L, Fagot-Campagna A. Algorithms based on medico-administrative data in the field of endocrine, nutritional and metabolic diseases, especially diabetes. *Rev Epidemiol Sante Publique*. oct 2017;65 Suppl 4:S168-73.
88. Quantin C, Collin C, Frérot M, Besson J, Cottenet J, Corneloup M, et al. Study of algorithms to identify schizophrenia in the SNIIRAM database conducted by the REDSIAM network. *Rev Epidemiol Sante Publique*. oct 2017;65 Suppl 4:S226-35.
89. Site officiel du réseau ReDSiam | Redsiam [Internet]. [cité 13 juill 2021]. Disponible sur: <https://www.redsiam.fr/>
90. Looten V. Les études portant sur les bases de données médico-administratives sont-elles reproductibles ? - L'hypothèse d'une inconduite éthique en santé publique. *médecine/sciences*. 1 août 2019;35(8-9):689-92.
91. Lindenauer PK, Lagu T, Shieh M-S, Pekow PS, Rothberg MB. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. *JAMA*. 4 avr 2012;307(13):1405-13.
92. Mise en œuvre d'indicateurs de résultat mesurés à partir des bases de données médico-administratives en ville ou en établissement de santé - APPEL A PROJETS [Internet]. Haute Autorité de Santé. [cité 13 juill 2021]. Disponible sur: https://www.has-sante.fr/jcms/p_3266150/fr/mise-en-oeuvre-d-indicateurs-de-resultat-mesures-a-partir-des-bases-de-donnees-medico-administratives-en-ville-ou-en-etablissement-de-sante-appel-a-projets
93. Stats ATIH [Internet]. [cité 13 juill 2021]. Disponible sur: <https://www.scansante.fr/>

Résumé : La disponibilité et le volume des données issues de bases médico-administratives (BMA) en font un outil de plus en plus utilisé en recherche. Cependant, la nature et la finalité financière de ces données rendent nécessaire une adaptation des méthodes d'épidémiologie classique. En effet, la réutilisation de ces données a soulevé des objections qu'il convient de prendre en compte. L'objectif de cette thèse était de montrer quel pouvait être l'apport des données issues de BMA en épidémiologie et en pharmacovigilance.

Dans une première partie introductive sur les BMA, nous nous sommes attachés à en rechercher une définition, à illustrer leur périmètre, au travers d'exemples internationaux, et, à décrire le contenu des bases disponibles en France : les données hospitalières (Programme de médicalisation des systèmes d'information, PMSI) et les données de remboursement extra-hospitalières (datamart de consommation inter-régime, DCIR). Dans une deuxième partie, nous avons présenté des travaux méthodologiques prenant en compte les spécificités des données issues des BMA : la définition et l'évaluation d'un score pronostique (score de comorbidités de Charlson) calculé à partir des données disponibles dans les BMA, la gestion des données absentes des BMA. Deux approches ont été explorées : i) le développement d'un proxy appliqué à la notion d'hospitalisation programmée, ii) l'enrichissement de données de BMA par un appariement à des données issues d'un entrepôt de données hospitalières. Dans une troisième partie, nous avons illustré, au travers de cas d'usage, l'intérêt des données médico-administratives en épidémiologie et en pharmacovigilance : une fenêtre d'exposition étendue (recueil de comorbidités maternelles dans les 6 ans précédant une grossesse), un recrutement quasi-exhaustif en particulier dans le cadre de pathologies rares (sujets atteints de MICI et d'un syndrome coronarien), l'accélération du recrutement de sujets et du recueil de données (sujets atteints de COVID-19 pris en charge en réanimation) notamment en situation de crise sanitaire.

Nous avons ensuite discuté les limites qui subsistent dans la réutilisation de ces données : leur volumétrie qui peut conduire à une situation de surpuissance statistique, leurs délais d'accès réglementaire et technique, l'investissement nécessaire à la compréhension de ces données, incomplètement mesuré par les utilisateurs de ces données qui n'en sont pas forcément les producteurs.

En conclusion, nos travaux ont fait ressortir l'intérêt de la réutilisation des données issues de bases médico-administratives en recherche tout en soulignant les écueils qui pouvaient en résulter. L'intégration et la réutilisation de données multisources en santé permettrait de pallier un certain nombre de ces difficultés.

Mots-clefs : base médico-administrative, utilisation secondaire de données, épidémiologie, pharmacovigilance, Système National des Données de Santé

Abstract: The availability and the volume of data in administrative claims databases could offer a powerful tool for research. However, the origin and the financial purpose of these data require an adaptation of the classical epidemiological approaches. Indeed, the secondary use of these data raises several issues that need to be carefully addressed. The aim of this thesis work was to assess the contribution of claims databases in epidemiology and in pharmacovigilance.

The first section discusses the definition of the administrative claims databases and their scope through international examples and then focuses on French claims databases that collect inpatient care data (Programme de médicalisation des systèmes d'information, PMSI) and outpatient care data (datamart de consommation inter-régime, DCIR). The second section discusses methodological issues to take into account the specificities of administrative claims databases: definition and assessment of a prognostic score (Charlson's comorbidity index) applied to the SNDS database, and the management of missing data. Two approaches were used: i) the development of a proxy applied to the concept of planned hospital stay, and ii) the enhancement of claims data with clinical data from a hospital warehouse. The third section illustrates, through use cases, how claims data could be useful in epidemiology and pharmacovigilance: wide exposure window (maternal comorbidities collected in the 6 years before the conception date), exhaustive recruitment especially for rare pathologies (patients with inflammatory bowel disease and myocardial infarction), and improved screening of eligible patients and data collection especially during a health crisis (patients admitted to the intensive care unit for COVID-19).

Finally, the limitations due to the secondary use of these data are discussed: their volume that might lead to statistical overpowered studies, their delayed access due to technical and legal restrictions, the effort required to understand these data that is often underestimated by users who are not always data producers.

To conclude, our work highlights the interest of the secondary use of data from claims database in research while underlining the potential pitfalls. The integration and the secondary use of multisource data in healthcare research could solve part of these issues.

Keywords: administrative claims database, secondary use of data, epidemiology, pharmacovigilance, Système National des Données de Santé