



HAL
open science

Contrôle des facteurs confondants pour étudier les résultats chirurgicaux à partir des bases hospitalières nationales

Cécile Payet

► **To cite this version:**

Cécile Payet. Contrôle des facteurs confondants pour étudier les résultats chirurgicaux à partir des bases hospitalières nationales. Santé publique et épidémiologie. Université de Lyon, 2020. Français. NNT : 2020LYSE1091 . tel-03506288

HAL Id: tel-03506288

<https://theses.hal.science/tel-03506288>

Submitted on 2 Jan 2022

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.

N°d'ordre NNT : xxx



THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale 205
Ecole Doctorale Interdisciplinaire Sciences-Santé

Spécialité de doctorat : Epidémiologie, Santé Publique, Recherche
sur les services de santé

Discipline : Biologie, médecine et santé

Soutenue publiquement le 03/06/2020, par :

Cécile Payet

**Contrôle des facteurs confondants lors
de l'étude de résultats chirurgicaux
dans les bases hospitalières nationales**

Devant le jury composé de :

Farges, Olivier, PU-PH, Assistance Hôpitaux Publique de Paris, Rapporteur
Grammatico-Guillon, Leslie, MCU-PH, CHU de Tours, Examinatrice
Katsahian, Sandrine, PU-PH, Assistance Hôpitaux Publique de Paris, Rapporteur
Lifante, Jean-Christophe, PU-PH, Hospices Civils de Lyon, Examineur

Duclos, Antoine, PU-PH, Hospices Civils de Lyon, Directeur de thèse
Rabilloud, Muriel, MCU-PH Hospices Civils de Lyon Co-directrice de thèse

Université Claude Bernard – LYON 1

Président de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Damien VERHAEGHE

COMPOSANTES SANTE

Faculté de Médecine Lyon-Est – Claude Bernard	Doyen : M. Gilles RODE
Faculté de Médecine et Maïeutique Lyon Sud Charles. Mérieux	Doyenne : Mme Carole BURILLON
UFR d'Odontologie	Doyenne : Mme Dominique SEUX
Institut des Sciences Pharmaceutiques et Biologiques	Directrice : Mme Christine VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. Xavier PERROT
Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

UFR Biosciences	Directrice : Mme Kathrin GIESELER
Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
UFR - Faculté des Sciences	Administrateur provisoire : M. Bruno ANDRIOLETTI
UFR (STAPS)	Directeur : M. Yannick VANPOULLE
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Ecole Polytechnique Universitaire Lyon 1	Directeur : Emmanuel PERRIN
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Institut de Science Financière et d'Assurances	Directeur : M. Nicolas LEBOISNE
ESPE	Administrateur Provisoire : M. Pierre CHAREYRON

Remerciements

Je tiens à remercier tout particulièrement Antoine Duclos dans le cadre de la direction de ce travail et plus largement pour tous les projets que nous menons ensemble depuis maintenant sept ans. Grâce à ton soutien et ta confiance, j'apprends quotidiennement au sein de ton équipe.

Je remercie vivement Muriel Rabilloud d'avoir accepté de co-diriger ce travail. Merci pour nos échanges et tes conseils.

Je remercie Olivier Farges et Sandrine Katsahian d'avoir accepté de juger mon travail et de m'avoir fait bénéficier de votre expertise, ainsi que Leslie Grammatico-Guillon et Jean-Christophe Lifante pour avoir accepté de faire partie du jury.

Je voulais témoigner ma reconnaissance à l'ensemble des personnes qui ont rendu possible la réalisation de ce travail, Stéphanie pour l'élaboration des bases et son expertise sur les données ainsi que les cliniciens pour leur précieuse collaboration.

Je remercie les membres de mon équipe avec lesquels j'ai beaucoup de plaisir à travailler au quotidien. Je salue plus généralement l'ensemble des collègues de Lacassagne pour leur bonne humeur et les moments de convivialité partagés au travail et à l'extérieur.

Merci à tous ceux avec qui j'ai pu travailler jusqu'à présent, qui ont pris le temps de me transmettre une partie de leur savoir et de leur expérience.

Enfin comme l'équilibre de la vie professionnelle dépend de celui de la vie personnelle, j'adresse un clin d'œil à mes amis. Bien sûr, j'ai une pensée pour ma famille, pour Damien, Milo et Elise, pour tout ce que nous partageons et à nos projets futurs. Vous m'avez donné l'énergie nécessaire, du début à la fin.

Résumé

Les bases de données hospitalières nationales sont de plus en plus utilisées pour étudier des résultats chirurgicaux en vie réelle. Néanmoins, les données collectées dans ces bases reflètent l'activité de routine des hôpitaux et ne répondent pas à une question spécifique de recherche. Certaines données peuvent être manquantes ou de qualité insuffisante pour contrôler divers facteurs confondants ce qui peut conduire à une interprétation erronée des résultats. L'objectif de ce travail était d'explorer différentes méthodes permettant de mieux contrôler de potentiels facteurs confondants mesurés ou non, lors de l'étude de résultats chirurgicaux dans les grandes bases de données hospitalières nationales.

Dans un premier temps, nous avons réalisé une revue de la littérature afin de décrire la qualité méthodologique des études analysant la mortalité chirurgicale dans les bases de données hospitalières nationales et d'identifier des pistes d'amélioration pour mieux contrôler ces facteurs confondants. Dans un second temps, nous avons proposé trois travaux dans lesquels nous avons cherché à contrôler des facteurs confondants à diverses étapes de réalisation d'une étude incluant son schéma méthodologique, les stratégies de sélection de sa population, l'analyse statistique et les analyses de sensibilité. Nous avons en particulier étudié le lien entre le volume d'activité et le résultat chirurgical dans les hôpitaux et la comparaison directe d'efficacité entre différentes procédures. Ainsi, le large périmètre dans le temps et l'espace des bases hospitalières nationales permet d'envisager différentes méthodes pour contrôler de potentiels facteurs confondants.

Mots clés : bases de données hospitalières nationales, facteurs confondants, recherche sur les résultats chirurgicaux, chirurgie

Discipline : Epidémiologie, Santé Publique, Recherche sur les Services de Santé

Intitulé et adresse du laboratoire :

EA 7425 HESPER - Health Services and Performance Research

Université Claude Bernard Lyon 1

Domaine Rockefeller - 2ème étage (couloir CD)

8 avenue Rockefeller, 69373 Lyon Cedex 8

Abstract

Large hospital databases are frequently used to explore real-world surgical outcomes. However, hospital administrative data is typically generated for billing purposes and does not answer a specific research question. Insufficient data granularity or unrecorded variables for specific confounders may lead to misinterpretation of results. This is of special importance for interpreting outcome in surgical comparative effectiveness studies when certain types of patients are more likely to receive surgery owing to possible confounding by indication. The objective of this work was to explore different methods to better control potential observed and unobserved confounding factors to study surgical outcomes in large hospital databases.

First, we conducted a systematic review to describe the methodological quality of surgical mortality studies using large hospital databases and to identify ways to improve confounding control. Then, we proposed three studies in which we sought to control for confounding factors at various steps of the study, including study design, population selection, statistical analysis and sensitivity analyses. In particular, we performed a study investigating the relationship between hospital surgical procedures volume and outcome as well as two surgical comparative effectiveness studies. Different methodologies can thus be considered to control potential confounding factors due to the large temporal and spatial scope of these databases.

Keywords : large hospital databases, confounding factors, surgical outcome research, surgery

Discipline : Epidemiology, Public Health, Health Services Research

Intitulé et adresse du laboratoire :

EA 7425 HESPER - Health Services and Performance Research

Université Claude Bernard Lyon 1

Domaine Rockefeller - 2ème étage (couloir CD)

8 avenue Rockefeller, 69373 Lyon Cedex 8

Table des matières

I.	Contextualisation du travail	8
A.	Etudier les résultats chirurgicaux	8
1.	La place de la chirurgie dans les soins hospitaliers	8
2.	La recherche sur les résultats	8
3.	Le résultat comme mesure de la qualité	10
4.	Les déterminants du résultat chirurgical	15
B.	Les bases de données hospitalières nationales	17
1.	Intérêts et limites des bases de données hospitalières	17
2.	Le PMSI	19
C.	Les facteurs confondants	23
1.	Les différents types de facteurs confondants	23
2.	Méthodes de prise en compte des facteurs confondants	25
II.	Objectifs.....	37
A.	Evaluer la qualité méthodologique des études analysant la mortalité chirurgicale dans les bases de données hospitalières nationales	37
B.	Contrôler les facteurs confondants lors de l'étude du lien volume-résultat en chirurgie dans les bases de données hospitalières nationales.....	37
C.	Contrôler les facteurs confondants lors de la comparaison de l'efficacité réelle entre différentes interventions dans les bases hospitalières nationales	37
III.	Travaux scientifiques répondant aux objectifs	38
A.	Qualité méthodologique des études analysant la mortalité chirurgicale dans les bases de données hospitalières nationales : une revue de la littérature (article 1).....	38
B.	Reconsidérer le lien volume-résultat en chirurgie : étude TREVORS (article 2).....	47
C.	Reprise chirurgicale après une chirurgie bariatrique (article 3)	59
D.	Performance des scores de propension à haute dimension (article 4)	70
IV.	Synthèse	81
A.	Discussion des travaux réalisés	81

1. Contrôle des facteurs confondants lors de l'étude du lien volume résultat en chirurgie (Etude 2)	81
2. Contrôle des facteurs confondants lors de la comparaison de l'efficacité réelle de deux interventions chirurgicales (Etude 3 et 4).....	83
B. Perspectives	87
1. La recherche du bon groupe contrôle	87
2. Utiliser de nouvelles approches pour la sélection des variables	88
3. Apport de données additionnelles	89
V. Bibliographie.....	91
VI. Annexes	106
A. Annexes des articles présentés dans ce travail.....	106
1. Annexes de l'article 1	106
2. Annexes de l'article 2	114
3. Annexes de l'article 3	120
4. Annexes de l'article 4	122
B. Publications et communications de ce travail	124
1. Publications	124
2. Communications orales.....	124
3. Posters	125

I. Contextualisation du travail

A. Etudier les résultats chirurgicaux

1. La place de la chirurgie dans les soins hospitaliers

La chirurgie constitue une part prépondérante des soins hospitaliers. On estime que 313 millions d'opérations chirurgicales importantes sont effectuées chaque année dans le monde [1]. En France, 4,8 millions d'hospitalisations ont été enregistrées en 2018 pour la réalisation d'un acte chirurgical [2]. Ces estimations n'incluent que les procédures menées en salle d'opération, impliquant incision, excision, manipulation ou suture de tissus et nécessitant généralement une anesthésie locale ou générale ou une sédation profonde. Outre le nombre d'actes réalisés, les coûts engendrés font de la chirurgie un enjeu majeur de santé publique.

De plus, dans les pays industrialisés 3 à 16% des interventions chirurgicales entraînent des complications sérieuses, avec un taux d'incapacité permanente ou de mortalité allant de 0,4 à 0,8% [3]. Plus de 100 000 américains meurent chaque année suite à une chirurgie et 20 fois plus présentent des complications chirurgicales [4]. En France, une enquête nationale menée sur un échantillon de 81 établissements français a estimé la proportion de complications survenues au cours d'un séjour hospitalier avec chirurgie à 9,2%. Ces complications seraient évitables dans 2,9% des cas [5].

A l'heure où notre système de santé évolue en permanence sous la pression des enjeux épidémiologiques (vieillesse de la population, ...), technologiques (techniques robotiques, ...) et dans un contexte de limitation des ressources, l'intérêt pour la recherche sur les services de santé et les résultats en chirurgie prend tout son sens afin d'améliorer les soins et offrir une prise en charge de meilleure qualité et plus sûre pour le patient [6].

2. La recherche sur les résultats

a) *Qu'est-ce que la recherche sur les services de santé ?*

Il est utile de considérer les deux définitions les plus utilisées de la recherche sur les services de santé, ou « *Health Services Research* » (HSR) dans les pays anglophones, issues de deux organismes américains :

- La définition de l'*AcademyHealth* publiée en 2000 [7]: « *the multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and personal behaviors affect access to health care, the quality and cost of health care, and ultimately our health and well-being. Its research domains are individuals, families, organizations, institutions, communities, and populations* ».
- La définition de l'*Agency for Healthcare Research and Quality (AHRQ)* publiée en 2002 [8]: « *Health services research examines how people get access to health care, how much care costs, and what happens to patients as a result of this care. The main goals of health services research are to identify the most effective ways to organize, manage, finance, and deliver high quality care; reduce medical errors; and improve patient safety*».

Ces définitions résument ainsi les principaux aspects de la recherche sur les services de santé et font apparaître l'aspect pluridisciplinaire de ce domaine de recherche. La recherche sur les services de santé s'intéresse à l'organisation des services de santé au sens large et à leurs effets sur les soins délivrés aux patients. Elle étudie l'accès aux services de santé, le coût des soins et leurs impacts sur les patients. Les principaux objectifs de la recherche sur les services de santé sont d'identifier les facteurs individuels et organisationnels pouvant entraîner des variations sur la qualité du soin ainsi que les meilleures stratégies permettant d'organiser, gérer, financer et fournir des soins de haute qualité pour le patient. La recherche sur les services de santé vise à produire des éléments nouveaux de connaissance et de méthodologie mais également à fournir aux décideurs des éléments de preuve robustes pour évaluer les résultats de leurs actions et de leurs politiques [9]. La recherche sur les services de santé s'intéresse à la prise en charge du patient dans un contexte large incluant à la fois, le praticien, l'hôpital et le système de soins [10]. Néanmoins dans ce travail, nous étudierons les prises en charge des patients dans le contexte de l'hôpital de par les données manipulées.

b) Qu'est-ce que la Recherche sur les Résultats ?

La recherche sur les résultats ou *outcomes research* est une des disciplines principales de la recherche sur les services de santé [4]. L'AHRQ définit la recherche axée sur les résultats comme suit: "*Outcomes research seeks to understand the end results of particular health care*

practices and interventions. End results include effects that people experience and care about, such as change in the ability to function. In particular, for individuals with chronic conditions—where cure is not always possible—end results include quality of life as well as mortality. By linking the care people get to the outcomes they experience, outcomes research has become the key to developing better ways to monitor and improve the quality of care “ [11].

La recherche sur les résultats s'intéresse aux résultats finaux des pratiques et interventions en matière de soins de santé. Elle permet d'identifier les soins chirurgicaux qui sont inefficaces, s'assure que les évaluations des nouvelles technologies tiennent compte des coûts et des bénéfices pour le patient. Elle permet également de surveiller et d'améliorer la qualité des soins.

3. Le résultat comme mesure de la qualité

La communauté scientifique a perçu depuis longtemps l'importance de mesurer le résultat des soins délivrés aux patients. Au début du 20^{ème} siècle, Ernest Codman, chirurgien au Massachusetts General Hospital de Boston, notait que les hôpitaux reportaient le nombre de patients traités mais ne s'intéressaient pas au devenir de leurs patients et si le soin leur avait été bénéfique ou non. Ainsi, il insista sur la nécessité d'une collecte minutieuse de données sur les résultats des soins chirurgicaux et de la publication de ces derniers afin d'améliorer la qualité des soins délivrés aux patients [12].

Un soin de qualité « doit permettre de garantir à chaque patient l'assortiment d'actes diagnostiques et thérapeutiques qui leur assurera le meilleur résultat en termes de santé, conformément à l'état actuel de la science médicale, au meilleur coût pour un même résultat, au moindre risque iatrogène, et pour sa grande satisfaction, en termes de procédures, de résultats et de contacts humains à l'intérieur du système de soins » [13]. Ainsi, le soin doit être à la fois efficace (conformes aux connaissances scientifiques et aux recommandations basées sur des données probantes), sûr (réduire au minimum les risques de préjudice pour le patient) et centré sur la personne (respecter et tenir compte des préférences, des besoins et des valeurs du patient) [14,15] (Figure 1). De plus pour tirer pleinement parti des avantages de la qualité, les soins doivent être équitables (la qualité ne varie pas selon les caractéristiques personnelles telles que le genre, l'appartenance ethnique, la localité ou le niveau socio-économique par exemple), intégrés (être coordonnés entre l'hôpital et les soins de ville),

efficaces (éviter le gaspillage de ressources) et dispensés en temps opportun (éviter au maximum les retards dans la délivrance du soin) [14] (Figure 1).



Figure 1- Composantes de la qualité des soins [14].

Afin d'évaluer la qualité des soins, le modèle conceptuel de Donabedian, publié en 1966 est traditionnellement utilisé. Il permet de distinguer trois domaines clés dans l'évaluation de la qualité : la structure, le processus et le résultat [3] (Figure 2). La structure fait référence aux moyens mis en œuvre (gestion des ressources humaines, matérielles, financières, organisation des services) durant le processus de soins. Le processus est une notion plus complexe en rapport avec la qualité des pratiques d'une activité de soins et le respect des recommandations et bonnes pratiques professionnelles. Enfin le résultat montre directement, à l'issue des soins, les bénéfices ou les risques générés pour le patient.

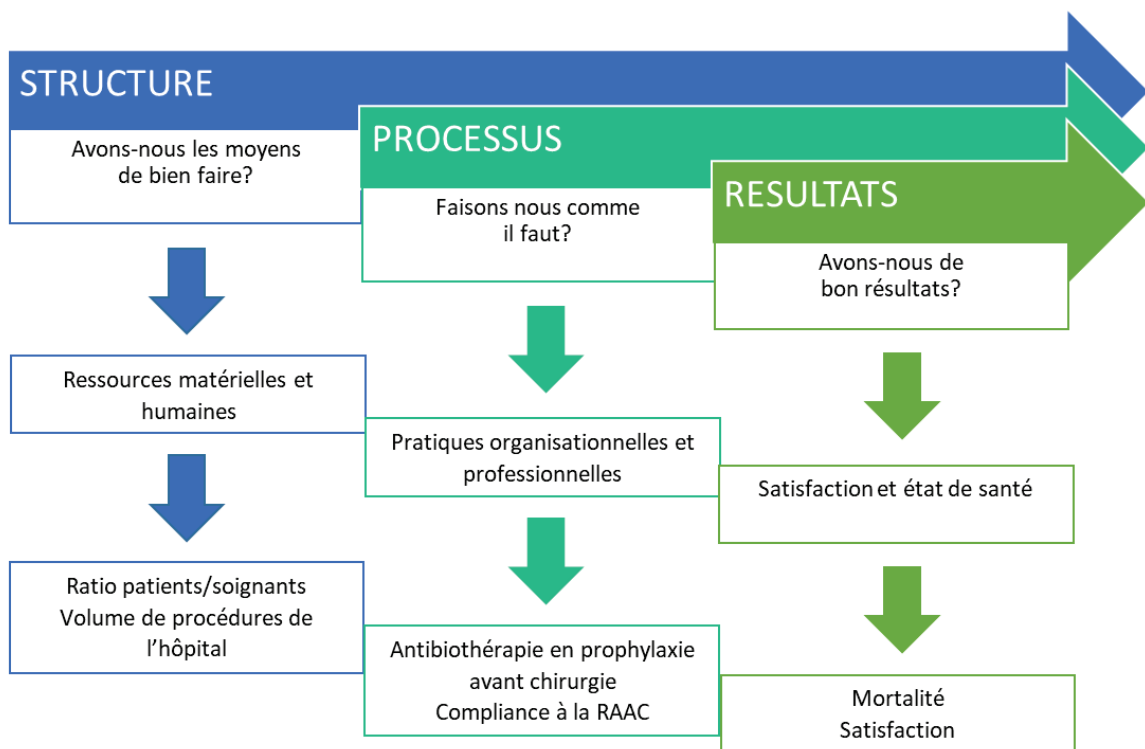


Figure 2 – Les trois niveaux du modèle théorique de l'évaluation de la qualité des soins selon Donabedian [16]. RAAC : Programme de récupération améliorée après chirurgie.

Pour mesurer chacun de ces domaines, l'utilisation d'indicateurs doit être privilégiée afin d'avoir des données chiffrées rigoureuses. Un indicateur est défini comme « un élément mesurable qui fournit des renseignements au sujet de phénomènes complexes, qui ne sont pas faciles à saisir » [17]. C'est donc un outil d'aide à la décision dont l'utilisation s'inscrit dans une démarche qui répond à un objectif et se situe dans un contexte donné. L'indicateur n'a d'intérêt que par le choix qu'il aide à faire dans ce cadre. Il doit être cliniquement pertinent et validé afin de refléter ce qu'il est censé mesurer. Il doit être fiable c'est-à-dire apte à une mesure précise et reproductible. Enfin, il doit être simple d'interprétation [18].

Ainsi, il existe des indicateurs de structure, de processus et de résultat chacun se rapportant aux trois domaines clés du modèle théorique de Donabedian [19]. Des associations statistiques existent entre ces trois indicateurs et notamment entre les indicateurs de structure et de résultat [10,20]. Par exemple, des études ont montré qu'une augmentation du ratio patients/soignants, c'est-à-dire du nombre de patients par soignant, était associée à une augmentation de la mortalité [21–24]. La relation volume-résultat est également un autre

exemple bien documenté. Dans la littérature, les patients pris en charge dans des établissements de soins à haut volume d'activité présenteraient un risque moins élevé de complications, principalement dans le cadre de chirurgies complexes [25–33].

En chirurgie, les indicateurs de résultats les plus étudiés sont la mortalité et les complications post-opératoires [34]. D'autres indicateurs de résultats sont couramment inclus dans les études telles que la durée du séjour, le taux de réhospitalisations ou de reprise chirurgicale. Les indicateurs de résultats peuvent être également plus subjectifs comme la satisfaction des patients ou leur qualité de vie [34,35].

Lorsqu'on étudie des résultats cliniques, si certains indicateurs sont faciles à définir (par ex. la mortalité), d'autres sont plus compliqués et peuvent avoir diverses définitions ce qui peut nuire à la généralisation des résultats [36]. Néanmoins, il existe des indicateurs dont la mesure est standardisée. Par exemple, pour les complications post-opératoires, il est intéressant d'utiliser des systèmes de classification validés. La classification de Clavien-Dindo est couramment utilisée pour grader les complications post-opératoires selon l'intensité de la prise en charge de ces complications [37–39] (Table 1). Ce système peut être ajusté afin de correspondre à une définition claire et consensuelle de la définition de complication postopératoire spécifique à la chirurgie étudiée [40].

Table 1 -Classification des complications chirurgicales selon Clavien-Dindo [37,38]

Grade	Définition
I	Tout évènement post-opératoire indésirable ne nécessitant pas de traitement chirurgical, endoscopique, radiologique ou médical. Traitements autorisés : antiémétiques, antipyrétiques, analgésiques, diurétiques, électrolytes et kinésithérapie.
II	Complication nécessitant un traitement médical n'étant pas autorisé au grade I.
III	Complication nécessitant un traitement chirurgical, endoscopique ou radiologique IIIa sans anesthésie générale IIIb sous anesthésie générale
IV	Complications engageant le pronostic vital et nécessitant des soins intensifs IVa défaillance d'un organe (y compris dialyse) IVb défaillance multi-viscérale
V	Décès

Une deuxième approche peut consister à utiliser des indicateurs déjà validés comme les *patient safety indicators* (PSI) dont certains permettent d'étudier le résultat d'une chirurgie (Table 2). Initialement développés aux Etats-Unis dans les années 2000 par l'AHRQ, les PSI sont produits à partir des données hospitalières et mesurent la prévalence de complications [41,42]. Par la suite, ils ont été adaptés au contexte français ainsi qu'à d'autres pays industrialisés. Ils permettent d'avoir une définition standardisée des complications afin de les suivre dans le temps et de les comparer, entre établissements ou pays par exemple, lorsque la qualité du codage le permet. En effet, leur mesure est tributaire des pratiques de codage ce qui constitue son principal inconvénient [43,44]. Dans la continuité des PSI, des indicateurs de qualité et de sécurité des soins (IQSS) ont été mis en place en France pour permettre de mesurer un état de santé, une pratique ou la survenue d'un évènement afin d'évaluer la qualité des soins et ses variations dans le temps [45]. Un de ces indicateurs est directement calculé à partir des bases hospitalières et permet d'étudier le résultat chirurgical (Table 2). L'IQSS ETE-ORTHO s'inspire directement du PSI 12 en ciblant plus particulièrement la chirurgie orthopédique.

Table 2- Liste des Patient Safety Indicators (PSI) développés initialement par l'AHRQ et de l'indicateur de qualité et de sécurité des soins (IQSS) en rapport avec la chirurgie

Indicateur	N°	Libellé
PSI	05.	Corps étranger oublié au cours d'une intervention chirurgicale
	06.	Pneumothorax iatrogène
	08.	Fracture de hanche post-opératoire
	09.	Hémorragie ou hématome post-opératoire
	10.	Désordre physiologique et métabolique post-opératoire
	11.	Détresse respiratoire post-opératoire
	12.	Embolie pulmonaire ou thrombose veineuse profonde post-opératoire
	13.	Septicémie post-opératoire
	14.	Désunion de plaie opératoire après chirurgie abdomino-pelvienne
	15.	Difficulté technique au cours d'une intervention chirurgicale - lacération ou piqûre accidentelle au cours d'un soin
IQSS	ETE-ORTHO	Événements thrombo-emboliques après pose de prothèse totale de hanche, hors fracture ou de genou

4. Les déterminants du résultat chirurgical

Pour mesurer correctement le résultat chirurgical, il est indispensable de tenir compte de l'ensemble des facteurs qui pourraient le faire varier lors de la prise en charge chirurgicale.

Le geste chirurgical nécessite une attention particulière car il est l'élément clé de la prise en charge chirurgicale. Sa complexité varie d'une intervention chirurgicale à une autre et selon son indication opératoire. Néanmoins, une intervention chirurgicale ne se limite pas au seul geste chirurgical. Elle a lieu dans un environnement de travail donné, pour un patient donné et allie des soins préopératoires, un temps d'anesthésie, un temps chirurgical et des soins postopératoires (Figure 3) [46]. Ainsi, la préparation du patient et sa surveillance ou la prescription adéquate de thérapeutiques avant et après le temps opératoire sont également des facteurs qui peuvent influencer le résultat chirurgical. Le résultat chirurgical varie donc en fonction d'un certain nombre de facteurs humains et organisationnels [47,48].

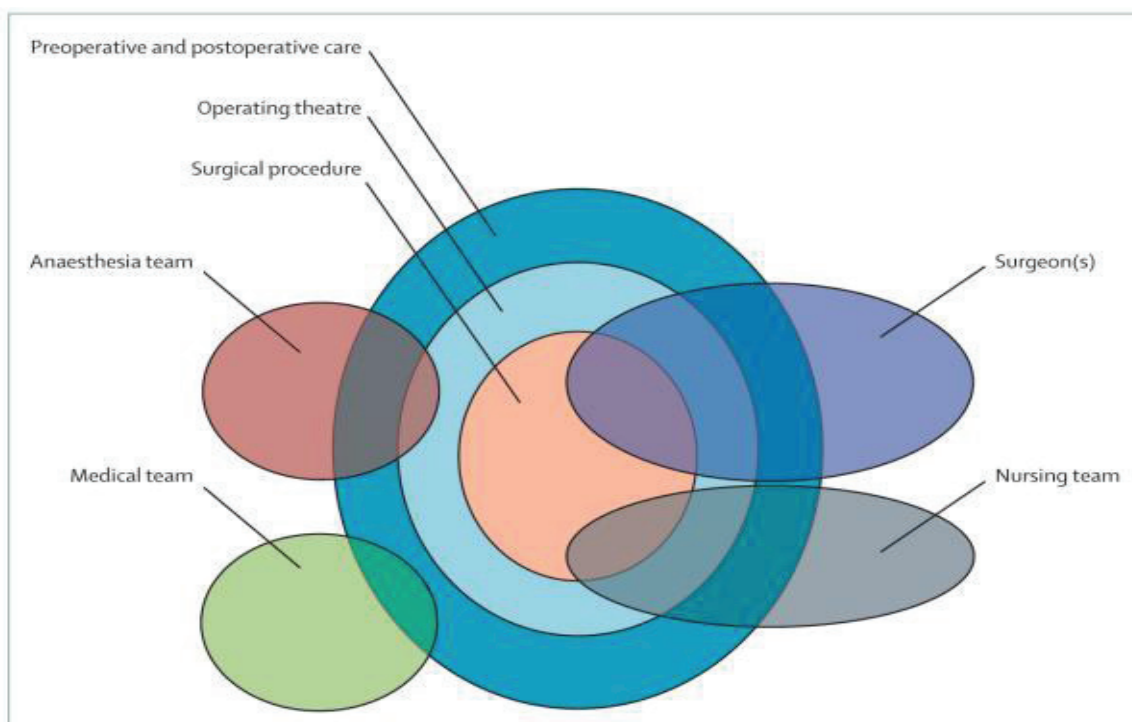


Figure 3 – Schéma de la complexité d'une intervention chirurgicale [46].

Certaines caractéristiques du patient telles que l'âge avancé, le sexe, la présence de comorbidités et/ou d'antécédents médicaux sont associées à de moins bons résultats et à un risque plus élevé de complications [49]. Le terme *case-mix* est régulièrement utilisé pour

décrire l'éventail des cas médicaux et chirurgicaux traités par un établissement de santé [50]. Des facteurs socio-économiques en rapport avec le lieu de résidence du patient peuvent également influencer le résultat chirurgical du fait des disparités de richesse, d'offre et d'accessibilité aux soins par exemple [51,52].

Le résultat chirurgical peut aussi varier en fonction du chirurgien et donc d'un chirurgien à un autre, selon son habilité chirurgicale, ses préférences, son expérience et son volume d'activité [46]. Il existe une courbe d'apprentissage initiale nécessaire à la maîtrise de la majorité des interventions chirurgicales [53] ainsi qu'une courbe de performance individuelle du chirurgien tout au long de sa carrière [53,54], qui amènent à une amélioration progressive des résultats obtenus grâce à la répétition de l'intervention [55]. L'aptitude du chirurgien à travailler en équipe est également reconnue comme un déterminant du résultat chirurgical [56–59].

L'organisation et les ressources de l'hôpital sont également des déterminants du résultat chirurgical. Les hôpitaux qui présentent des plateaux techniques performants pour l'imagerie, la radiologie interventionnelle et la réanimation par exemple, qui ont des équipes travaillant de manière coordonnée et des systèmes pour identifier précocement, puis réagir efficacement, à la survenue de complications, ou qui ont mis en place des programmes de réhabilitation destinés à améliorer la qualité des soins délivrés au patient présentent de meilleurs résultats chirurgicaux [60].

Néanmoins, la mesure du résultat chirurgical et son ajustement sont limités par les données utilisées. Ainsi, les sources de données et la manière dont les données sont collectées, sont deux considérations primordiales à prendre en compte lorsque l'on souhaite mesurer les résultats chirurgicaux. En effet, les données peuvent être imprécises du fait d'une variabilité importante dans la collecte des données. La granularité des données peut être insuffisante et des variables peuvent ne pas être collectées dans les bases ce qui va limiter l'ajustement des résultats. Cela est notamment le cas dans les bases de données hospitalières nationales.

B. Les bases de données hospitalières nationales

1. Intérêts et limites des bases de données hospitalières

La recherche sur les résultats chirurgicaux s'intéresse à des questions diverses, nous le reverrons par la suite, exigeant non seulement un échantillon de grande taille, mais aussi divers milieux de pratique. Les essais contrôlés randomisés étant difficilement envisageables dans ces cas-là au vu des coûts engendrés et de la taille d'échantillon nécessaire, il est particulièrement intéressant d'avoir recours aux bases de données hospitalières nationales [4].

Les bases de données hospitalières nationales sont disponibles dans la plupart des pays industrialisés. Les données collectées dans ces bases reflètent l'activité de routine des établissements de santé pour facturer les soins lors d'un séjour hospitalier. Elles contiennent des informations de facturation structurées, basées sur l'utilisation de classifications, incluant des informations médicales sur le patient, ses problèmes de santé et ses traitements. En revanche certaines données cliniques telles que les résultats des examens biologiques ne sont pas disponibles. Bien qu'elles ne soient pas initialement destinées à répondre à une question de recherche, ces bases présentent un certain nombre d'avantages puisqu'elles reflètent les soins délivrés aux patients en vie réelle.

Ces données englobent généralement de vastes populations, sont facilement accessibles, rapidement disponibles, relativement peu coûteuses à acquérir et se prêtent à une extraction informatisée des données [4]. Dans ces bases, il est également possible de relier les patients entre eux ce qui permet de suivre les différentes hospitalisations à travers les établissements. Par exemple, cela permet d'étudier les réhospitalisations même si elles ont lieu dans un autre établissement.

En revanche, ces avantages sont contrebalancés par des préoccupations relatives à l'exactitude et à l'exhaustivité du codage clinique. Le codage n'a pas pour but de décrire la réalité médicale, mais les éléments qui caractérisent le coût du séjour. Par conséquent, ces données sont souvent qualifiées d'imprécises en raison d'une variabilité importante dans la collecte de données liée à la motivation des codeurs [61–63]. Cette collecte peut varier d'un hôpital à l'autre ou d'une année à l'autre. De plus, la granularité des données est parfois

insuffisante et des variables peuvent ne pas être collectées dans les bases, ce qui peut limiter l'ajustement des résultats sur des facteurs confondants spécifiques [64,65]. Ainsi, les informations cliniques en rapport avec la facturation telles que la procédure chirurgicale réalisée, les informations démographiques, la durée de séjour, la mortalité sont enregistrées avec une bonne précision, ce qui n'est pas toujours le cas des informations qui ne participent pas à la facturation des soins comme certaines comorbidités du patient, le codage de certaines complications ou les informations sur le chirurgien qui a réalisé l'intervention par exemple.

Néanmoins, le large périmètre dans le temps et dans l'espace des bases hospitalières nationales sont particulièrement intéressantes pour investiguer des questions en rapport avec des interventions chirurgicales ou des complications rares du fait de la grande taille d'échantillon, pour explorer les évolutions temporelles des résultats chirurgicaux de par la profondeur des données et pour examiner des disparités dans la prise en charge chirurgicale au vu de la variété des pratiques représentées [4].

Les bases hospitalières nationales sont également intéressantes pour comparer l'efficacité réelle d'interventions chirurgicales. Bien que les essais cliniques randomisés soient considérés comme la référence pour comparer l'efficacité d'une prise en charge chirurgicale, en pratique, les conditions cliniques de l'essai en termes de suivi de patients et de ressources hospitalières disponibles sont rarement remplies et les patients ayant effectivement eu l'intervention ont des indications plus larges que les critères de sélection de l'essai ce qui limite la généralisation des résultats [4]. L'utilisation des essais cliniques randomisés pour l'évaluation d'interventions chirurgicales pose également des problèmes méthodologiques en lien notamment avec les préférences des patients, les préférences, l'expérience et la courbe d'apprentissage du chirurgien et la difficulté à mettre en place l'aveugle [46,66–69]. De plus, l'organisation globale de l'hôpital est difficilement contrôlable dans les essais cliniques.

Enfin, les bases hospitalières nationales sont particulièrement intéressantes pour évaluer l'effet de politiques de santé comme les paiements à la performance ou la centralisation des opérations dans des centres d'excellence [4].

Selon le pays, l'exhaustivité des patients hospitalisés inclus ainsi que les variables présentes dans ces bases hospitalières nationales varient. Par exemple, aux Etats-Unis, le *Healthcare Cost and Utilization Project National Inpatient Sample* (HCUP-NIS) correspond à un échantillon

représentatif de la population [70,71]. En revanche, en Angleterre (*Hospital Episodes Statistics*, HES) [72] ou en France (Programme de Médicalisation du Système d'Information, PMSI), l'ensemble des patients hospitalisés sont inclus.

2. Le PMSI

a) Historique

Importé des Etats-Unis dans les années 80 [73,74], le Programme de Médicalisation du Système d'Information (PMSI) consiste en un recueil exhaustif, systématique et standardisé d'informations administratives et médicales pour tout séjour d'un patient dans un établissement de santé public ou privé français. Depuis la loi du 31 juillet 1991 portant sur la réforme hospitalière, les établissements de santé publics et privés doivent procéder à l'analyse de leur activité médicale et transmettre aux services de l'Etat et à l'Assurance maladie « les informations relatives à leurs moyens de fonctionnement et à leurs activités [75,76]. A cette fin, ils doivent mettre en place un système d'information dédié, qui tient compte des pathologies et des modes de prise en charge de chaque séjour hospitalier. Le PMSI génère une base de données nationale, permanente et médicalisée. Depuis 2005 et la mise en application de la tarification à l'activité (T2A), Le PMSI sert également à assurer le financement des hôpitaux en fonction de l'activité renseignée dans la base. Le PMSI couvre plusieurs champs d'hospitalisations :

- Le PMSI – Médecine, Chirurgie, obstétrique et odontologie (MCO) dit « soins de courte durée »
- Le PMSI – Soins de Suite et Réadaptation (SSR) plus communément appelé « moyen-séjour »
- Le PMSI – Hospitalisation à Domicile (HAD)
- Le PMSI – Psychiatrie (PSY).

La majeure partie de l'activité est liée au PMSI Médecine, Chirurgie, Obstétrique (MCO) [77,78].

b) Hospitalisation : production d'un Résumé de Sortie Standardisé

Chaque séjour hospitalier MCO produit un Résumé de Sortie Standardisé (RSS), constitué d'un ou plusieurs Résumés d'Unité Médicale (RUM). Si le patient n'a été hospitalisé que dans une seule unité médicale durant son séjour, le RSS correspond au RUM généré par ce séjour (RSS « monoRUM »). Si le patient a fréquenté plusieurs unités médicales, le RSS est produit à partir des données des RUM générés pour chaque unité fréquentée par le patient au cours de son hospitalisation, ordonnés chronologiquement. Le RSS est alors constitué d'un ensemble de RUM (RSS « multiRUM ») identifié par le même numéro RSS.

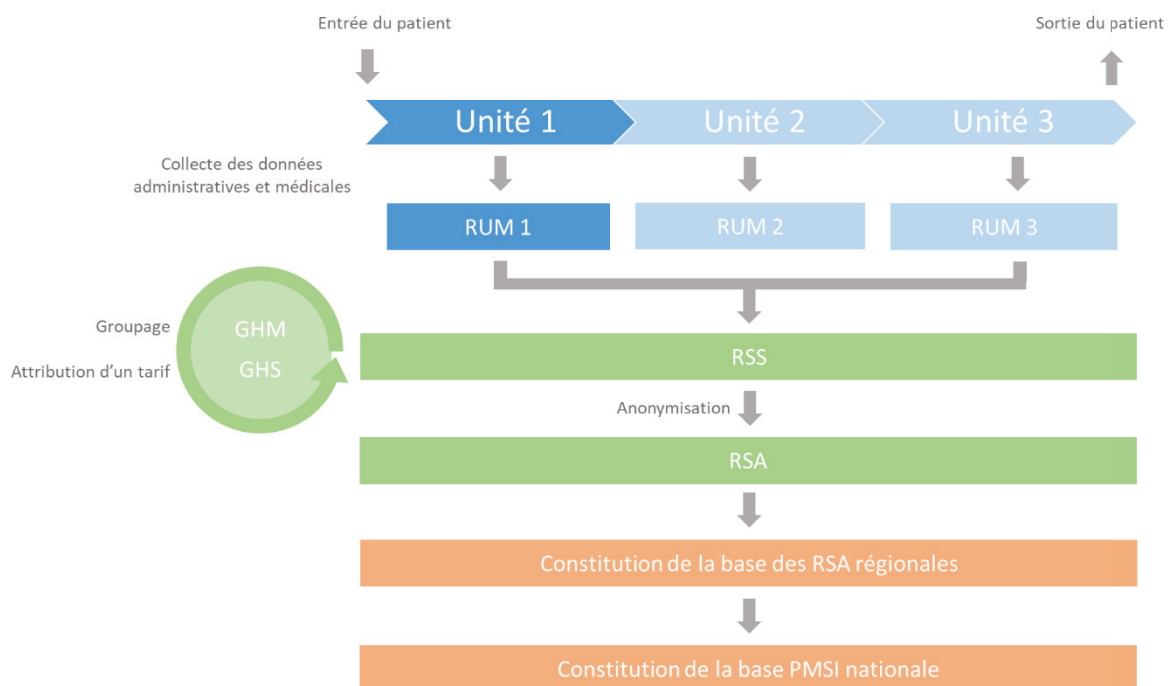


Figure 4 – De l'hospitalisation à la constitution de la base PMSI nationale. Unité : unité médicale d'hospitalisation du patient, RUM : Résumé d'Unité Médicale, RSS : Résumé de Sortie Standardisé, RSA : Résumé de sortie anonymisé, GHM : Groupement Homogène de Malade, GHS : Groupement Homogène de Séjour.

Le RUM contient un nombre limité d'informations administratives et médicales. Les informations administratives comprennent le numéro de l'établissement, les identifiants du patient et de son séjour, le sexe, la date de naissance et le code postal du lieu de résidence du patient, le numéro de l'unité médicale d'hospitalisation, les dates, les modes d'entrée et de sortie dans l'unité, la provenance et la destination. Les informations médicales contiennent

principalement les actes médicaux avec leur date de réalisation, les diagnostics réalisés pendant le séjour ainsi que certains consommables (certains dispositifs médicaux implantables dans la liste des produits et prestations et certaines molécules onéreuses). Les diagnostics se décomposent comme suit:

- Diagnostic principal (DP) : il s'agit du motif d'hospitalisation du patient ;
- Diagnostic relié (DR) : il est renseigné lorsque le DP est insuffisant et rend compte de la prise en charge du patient ;
- Diagnostics associés significatifs (DAS) : il s'agit de toutes les autres prises en charge du patient.

L'ensemble des informations contenues dans le RUM suivent des nomenclatures et classifications dédiées [78]. Les diagnostics sont codés avec la 10^{ème} révision de la Classification Internationale des Maladies (CIM-10) [79,80] et les actes médicaux sont codés avec la Classification Commune des Actes Médicaux (CCAM) [81]. Il ne figure dans les RUM que les problèmes de santé présents, actifs, diagnostiqués ou traités au moment de l'hospitalisation ou lors du passage par la structure d'accueil des urgences de l'établissement.

Les données collectées sont ensuite traitées automatiquement par un algorithme de groupage permettant de décrire la prise en charge d'un séjour sous la forme d'une Groupe Homogène de Malade (GHM).

c) Classification des séjours : les Groupes Homogènes de Malade

Chaque RSS est donc obligatoirement classé dans un seul GHM, regroupant des prises en charge de même nature médicale et économique. Cette classification est dérivée de la classification américaine des *Diagnosis Related Groups* (DRG) et elle est régulièrement actualisée ce qui induit une mise à jour de la fonction de groupage.

Le classement de chaque RSS dans un GHM est effectué par un algorithme de groupage et résulte d'une série de tests réalisés sur un nombre limité d'informations contenues dans le RSS notamment sur le DP, la présence d'actes opératoires classant et la présence de DAS.

A chaque GHM est associé un GHS (Groupe Homogène de Séjours) qui correspond au tarif associé au séjour. Il s'agit d'un tarif forfaitaire servant à la facturation lors d'une hospitalisation. Il est défini par l'Assurance Maladie.

d) Anonymisation : production du Résumé de Sortie Anonymisé

Le RSS est rendu anonyme et transformé en Résumé de Sortie Anonyme (RSA) avant la transmission à l'Agence Régionale de l'Hospitalisation (ARS) puis à l'Agence technique de l'Information Hospitalière (ATIH) pour la constitution de la base nationale. Dans les RSA, la date de naissance est remplacée par l'âge calculé à la date d'entrée, le code postal de résidence par le code géographique de l'Institut national de la statistique et des études économiques (Insee), les dates d'entrée et de sortie par la durée de séjour, le mois et l'année de sortie, la date de réalisation des actes par le délai en jours par rapport à la date d'entrée. Il est donc impossible d'identifier un patient à partir de la base régionale ou nationale, ce qui assure la confidentialité et le secret médical [82].

Depuis 2001, une procédure de chaînage des résumés de séjours permet de relier les différentes hospitalisations d'un même patient quel que soit l'établissement et le secteur d'activités grâce à un identifiant patient unique et anonyme (ANO). Il est alors possible de reconstituer le parcours de soins d'un patient quel que soit l'établissement et les secteurs d'activités dans lesquels il a été hospitalisé.

C. Les facteurs confondants

1. Les différents types de facteurs confondants

Comme nous l'avons vu précédemment, une prise en charge chirurgicale est complexe. Son indication dépend de nombreux facteurs en rapport avec le patient, le chirurgien et l'hôpital. Si ces facteurs influencent également les résultats chirurgicaux, ils sont considérés comme des facteurs confondants.

Un facteur confondant, ou facteur de confusion, est défini comme un facteur associé à la fois à l'exposition et à l'évènement d'intérêt. Il entraîne un biais ou une différence systématique entre les groupe de patients exposés et non exposés menant à une interprétation erronée des résultats [64] (Figure 5). Il conduit à un biais d'indication lorsqu'il influence le choix d'être exposé ou non [83,84]. Par exemple, l'âge peut être un facteur confondant lorsqu'on étudie des résultats chirurgicaux tels que la mortalité. Si les patients âgés sont préférentiellement dans le groupe exposé, et que la mortalité est plus élevée dans ce groupe, il est compliqué de déterminer si cette mortalité plus élevée est associée à l'exposition elle-même ou à l'âge des patients.

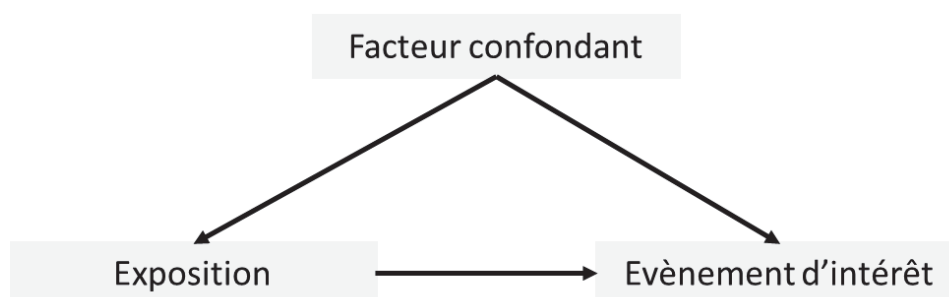


Figure 5 – Schéma représentant un facteur confondant.

Afin de faciliter la compréhension, dans la suite de cette partie, nous considérerons l'*outcome* comme un évènement binaire. Nous envisagerons également la variable d'exposition et les facteurs confondants comme des variables dichotomiques (présence ou absence).

Les facteurs confondants peuvent être mesurés, non mesurés ou inconnus dans les bases hospitalières.

Les facteurs confondants mesurés dans les bases, tels que l'âge et le sexe par exemple, peuvent être directement contrôlés dans le design d'étude ou dans l'analyse des résultats (Figure 6). Ces méthodes seront développées dans la partie suivante.

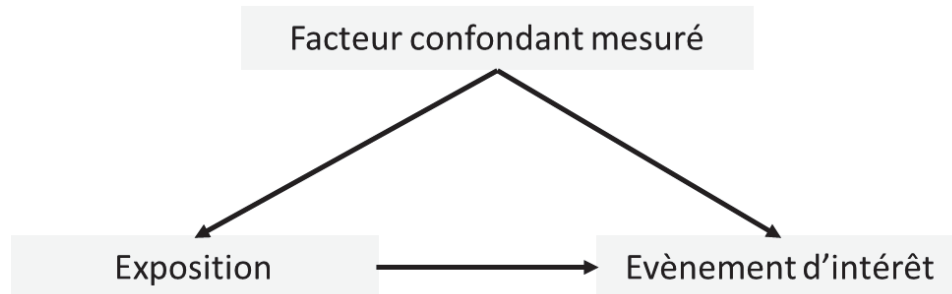


Figure 6 – Schéma représentant un facteur confondant mesuré.

En revanche, il est plus compliqué de contrôler un facteur confondant non mesuré tel que le statut tabagique ou la sévérité de la maladie qui sont des facteurs confondants connus par l'investigateur mais manquants ou collectés de manière imprécise dans les bases hospitalières. Comme ils sont absents ou mal renseignés de ces bases, il n'est pas possible de les contrôler directement par les méthodes classiques. Néanmoins, des variables mesurées dans les bases peuvent servir de proxy pour des facteurs confondants non mesurés (Figure 7) [85]. Par exemple, les codes de diagnostics présents dans ces bases peuvent indirectement décrire la sévérité de la maladie. De même, le statut de l'hôpital (centre hospitalo-universitaire, hôpital public, hôpital privée, ...) peut permettre de caractériser des différences structurelles et organisationnelles pour la dotation en personnel ou la disponibilité de plateaux techniques performants par exemple [86].

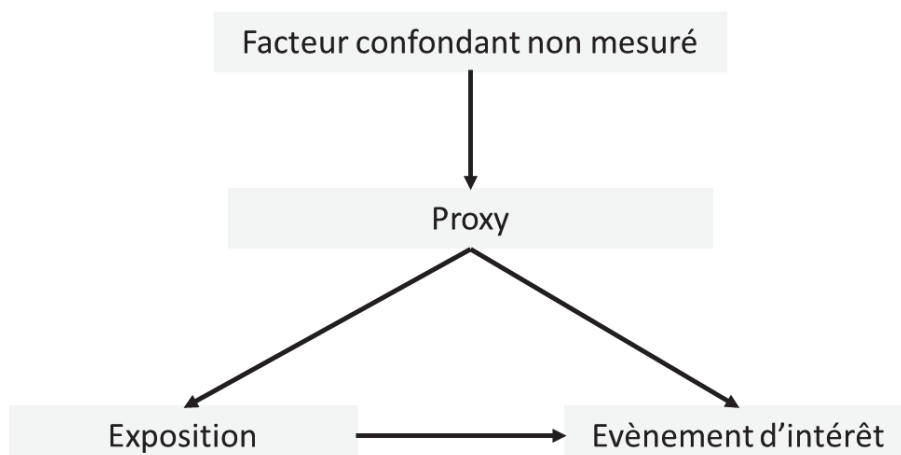


Figure 7 – Schéma représentant un facteur confondant non mesuré.

2. Méthodes de prise en compte des facteurs confondants

L'essai clinique randomisé est considéré comme la référence pour évaluer l'effet propre d'une exposition sur un évènement d'intérêt car les groupes ne diffèrent que par leur niveau d'exposition [4,69]. En l'absence de randomisation, la raison pour laquelle on cherche à contrôler les facteurs confondants lorsqu'on étudie l'effet d'une exposition sur un évènement d'intérêt dans les grandes bases de données hospitalières est de pouvoir mesurer cet effet propre. Différentes méthodes peuvent être utilisées pour minimiser le biais induit par ces facteurs confondants et faire en sorte, comme dans l'essai clinique, que les patients ne diffèrent entre les groupes que par leur niveau d'exposition (Table 3).

Table 3 – Méthodes de prise en compte des facteurs confondants mesurés ou non dans les études observationnelles

Méthodes	Facteur confondant	
	Mesuré	Non mesuré
Schéma d'étude	X	X
Restriction	X	
Stratification	X	
Appariement	X	
Analyse multivariée	X	
Score de propension	X	
Score de propension à haute dimension	X	X
Variable instrumentale	X	X

a) Le schéma d'étude

Contrairement à une étude observationnelle prospective, le chercheur qui réutilise secondairement les données hospitalières pour répondre à une question de recherche, ne peut pas contrôler la collecte des données pour minimiser certains biais potentiels. L'étude se limite donc aux variables disponibles dans la base au moment de la collecte, et des informations pertinentes peuvent être manquantes. En revanche, le large périmètre des

données hospitalières dans le temps et l'espace permet d'envisager divers schémas d'étude et divers groupes contrôles pour contrôler de potentiels facteurs confondants. De par la nature des données manipulées, l'étude de cohorte historique, est le schéma d'étude le plus utilisé. Ainsi, la survenue d'un évènement d'intérêt est comparée entre différents groupes définis en fonction de leur exposition à un facteur étudié [87].

b) La stratification et la restriction

La stratification peut être utilisée pour tenir compte d'un facteur confondant mesuré. Les patients sont regroupés au sein de strates en fonction des modalités du facteur confondant et l'effet de l'exposition est estimé en comparant les exposés et les non exposés au sein de chaque strate. Cette méthode permet d'obtenir les mesures de l'effet pour chaque strate, appelées mesures de l'effet conditionnel. Ces mesures n'indiquent pas l'effet moyen du traitement dans l'ensemble de la population même s'il existe des tests statistiques pour évaluer l'homogénéité de l'effet entre les strates [87].

La restriction consiste finalement à étudier l'effet de l'exposition uniquement dans certaines strates et donc à exclure les groupes de patients présentant une ou des caractéristiques données, à savoir une modalité d'un ou des facteurs confondants de l'étude.

Ainsi, ces deux techniques consistent à créer des sous-groupes de patients plus homogènes vis-à-vis du ou des facteurs confondants étudiés.

c) L'appariement

L'appariement consiste à contrôler a priori un ou plusieurs facteurs confondants en formant des paires homogènes quant à ce facteur confondant [87]. Il existe différents schémas d'appariement. Une des méthodes consiste à appairer les patients entre eux en fonction du facteur confondant qui nous intéresse. Par exemple, si on considère le sexe comme facteur confondant dans une étude, on appairera un homme exposé à un homme non exposé. Inversement, on appairera une femme exposée à une femme non exposée.

d) L'analyse multivariée

Il existe différents types de modèles de régression selon la distribution de l'*outcome*. Néanmoins, la méthode d'ajustement pour les facteurs confondants reste similaire quel que soit le modèle utilisé [4,87].

La régression multivariée permet de mesurer l'association entre l'évènement d'intérêt et l'exposition en tenant compte des facteurs confondants en tant que covariables du modèle. Elle permet d'obtenir l'effet de l'exposition lorsque le niveau des autres covariables est maintenu fixe [4,87].

Ainsi, la régression multivariée ne permet d'ajuster que sur des covariables disponibles dans la base et ne peut donc tenir compte que de facteurs confondants mesurés [4,87].

Néanmoins, il est également possible d'utiliser des modèles mixtes pour tenir compte de la structure des données [88] et indirectement de caractéristiques partagées par des individus présents dans une même entité. Par exemple, dans les bases hospitalières, les patients hospitalisés dans un même établissement partagent un environnement commun qui peut affecter l'évènement d'intérêt.

e) Les scores de propension

(1) Estimation du score de propension

En 1983, Rosenbaum et Rubin ont défini le score de propension (PS) comme la probabilité conditionnelle pour un patient d'être exposé en fonction de l'ensemble de ses caractéristiques observées [89]. Sa valeur varie donc entre 0 et 1, et les patients avec le même score de propension ont la même probabilité d'être exposés [89]. Ainsi, un patient exposé et un patient non exposé avec le même score de propension, peuvent avoir des caractéristiques individuelles différentes, mais en moyenne, la distribution des covariables observées entre les deux groupes est équilibrée afin de neutraliser les facteurs confondants [89,90].

Le score de propension est habituellement estimé à l'aide d'une régression logistique multivariée même si d'autres approches sont possibles telles que les réseaux neuronaux, les arbres de régression ou de classification [91]. Le choix des variables à inclure dans le modèle est principalement basé sur les connaissances cliniques en rapport avec le sujet de l'étude [92]. L'ensemble des facteurs confondants mesurés est habituellement inclus dans le modèle

ainsi que les variables pouvant influencer l'évènement d'intérêt [92–94]. En revanche, les variables liées uniquement à l'exposition ne doivent pas être incluses dans le modèle afin de ne pas augmenter le biais et diminuer la précision de l'estimation [92].

(2) Utilisation du score de propension

Le score de propension peut être utilisé selon quatre méthodes différentes : l'ajustement dans un modèle de régression, la stratification, l'appariement et la pondération par l'inverse de probabilité de traitement.

(a) Ajustement dans un modèle de régression

L'effet du traitement est estimé à l'aide d'un modèle multivarié incluant la variable permettant d'identifier le groupe d'exposition ainsi que le score de propension lui-même considéré en variable continue, en variable catégorielle ou avec toute autre transformation [95].

(b) Appariement

L'appariement sur le score de propension consiste à former des paires de patients exposés et non exposés ayant une valeur proche du score de propension [96]. Il existe plusieurs schémas d'appariement possibles [97,98] mais le plus utilisé consiste à appairer un patient exposé avec le patient non exposé le plus proche, en fixant une différence maximale entre deux patients (caliper) à 0.2 fois l'écart-type du logit du score de propension [99]. Une fois l'appariement réalisé, l'effet de l'exposition peut être estimé en comparant les critères de jugement entre les deux groupes en tenant compte du schéma d'appariement dans l'analyse [100].

(c) Stratification

Les patients sont regroupés au sein de différentes strates selon la valeur de leur score de propension. Traditionnellement, cinq strates sont constituées à partir des quintiles du PS. Il a été montré que la constitution de cinq strates permettait de réduire de 90% les biais liés aux facteurs confondants mesurés [101,102]. Cependant, les quintiles ne sont pas toujours appropriés, notamment lorsque la taille d'échantillon est importante comme c'est le cas dans les bases de données hospitalières nationales. D'autres méthodes pour calculer la taille et le nombre des strates ont été développées [103].

(d) Pondération par l'inverse de probabilité de traitement

La pondération par l'inverse de probabilité de traitement, plus connue sous le terme anglais d'*inverse probability of treatment weighting* (IPTW) consiste à appliquer à chaque patient une pondération qui est l'inverse de sa probabilité de traitement ($\frac{1}{PS}$ pour les patients exposés, $\frac{1}{1-PS}$ pour les patients non exposés) [95,104]. L'objectif de la méthode est de créer une pseudo-population dans laquelle les caractéristiques des patients sont équilibrées dans les deux groupes. Ainsi cette méthode augmente le poids des patients qui avaient peu de chance de recevoir le traitement d'après leurs caractéristiques observées et inversement [95,104]. L'effet de l'exposition est directement estimé dans cette pseudo-population à partir d'une régression logistique pondérée.

La pondération par l'inverse et l'appariement sont les méthodes qui équilibrent le mieux les groupes [100,105,106] et sont donc à privilégier. Dans certains cas l'appariement serait légèrement meilleur [100].

(3) Evaluation du score de propension

Il convient de vérifier que le score de propension équilibre les groupes sur les caractéristiques observées afin de s'assurer que le modèle est correctement spécifié. Il existe plusieurs méthodes selon le type d'utilisation du score de propension [100].

La différence standardisée est la référence pour évaluer l'équilibre entre les groupes [106]. Elle peut être calculée pour des variables quantitatives ou qualitatives et présente plusieurs avantages [107,108]. Elle est simple à calculer, indépendante de l'unité de mesure de la variable puisqu'il s'agit d'une quantité sans unité et ne dépend pas de la taille de l'échantillon [108]. Les covariables sont déséquilibrées entre les groupes lorsque la différence standardisée absolue est élevée. Même s'il n'existe pas de consensus, le seuil habituellement choisi pour définir l'existence d'un déséquilibre est de 10% [108].

Les différences standardisées n'analysant pas l'ensemble de la distribution de la variable, il est utile de représenter les diagrammes Quantiles-Quantiles du score de propension des deux groupes pour les variables continues afin de mettre en évidence un éventuel déséquilibre pour des zones spécifiques de la distribution de la variable malgré des valeurs de différences standardisées acceptables [108].

Il est également intéressant de représenter graphiquement la distribution des scores de propension dans chaque groupe [109] (Figure 8). Le degré de chevauchement permet d'identifier les patients qui ont un score de propension comparable entre les patients exposés et non exposés. A chaque extrémité des courbes, lorsque les distributions ne se chevauchent pas, les patients ont une probabilité d'être exposés si faible qu'aucun patient exposé n'a cette valeur de score de propension ou au contraire si élevée qu'aucun patient non exposé n'a cette valeur de score de propension. Il se pose alors la question d'inclure ces patients dans l'analyse puisqu'il n'existe pas de patients présentant des caractéristiques similaires dans les deux groupes d'exposition.

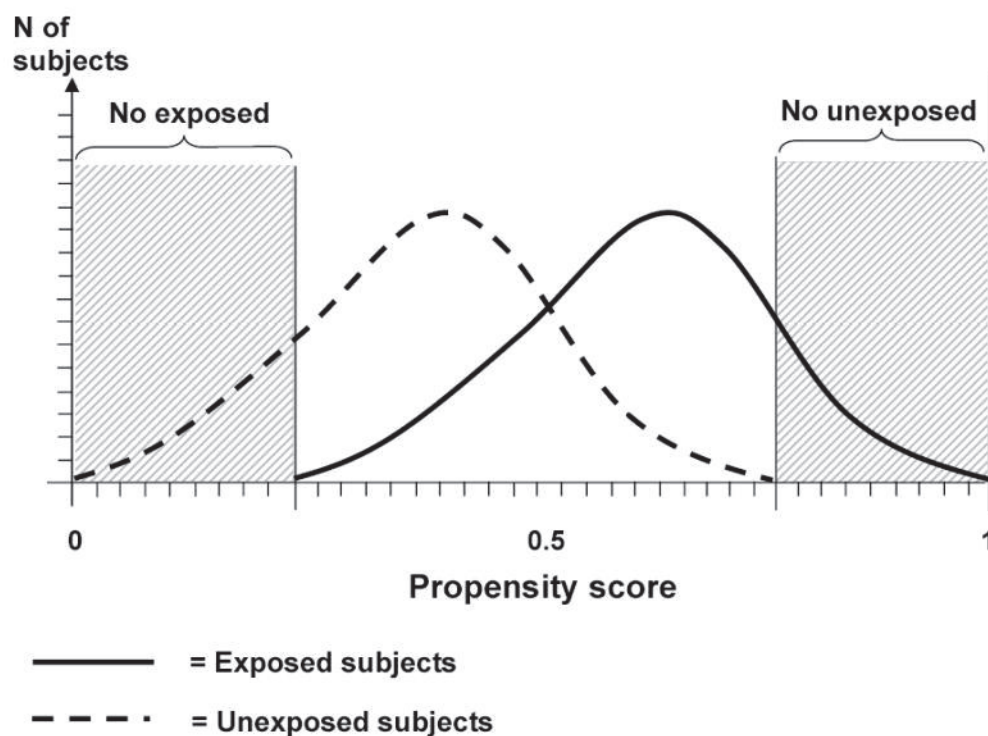


Figure 8 – Représentations graphiques des distributions de scores de propension pour les patients exposés et non exposés [109].

f) Les scores de propension à haute dimension

(1) Le concept

Les scores de propension ne tiennent compte que des facteurs confondants connus par l'investigateur et mesurés dans les bases de données. Néanmoins, les grandes bases de données contiennent des milliers de codes qui peuvent contenir de l'information sur de

potentiels facteurs confondants qui sont inconnus de l'investigateur. Même si ces facteurs ne sont pas directement mesurés dans les bases, ces codes peuvent servir de proxy et indirectement décrire les caractéristiques du patient. En 2009, Schneeweiss et ses collègues ont développé une nouvelle méthode, les scores de propension à haute dimension ou *high-dimensional propensity score* (hdPS), basés sur la méthode des scores de propension [110]. Elle vise à améliorer le contrôle du biais d'indication par rapport aux scores de propension classiques, en identifiant un nombre important de facteurs confondants normalement ignorés par l'investigateur, et à les inclure dans un score de propension. Cette nouvelle méthode est, à l'heure actuelle, principalement utilisée en pharmaco-épidémiologie pour étudier l'effet de traitements médicamenteux [110–116].

De la même manière que pour un score de propension classique, l'hdPS est estimé à partir d'un modèle de régression et correspond à la probabilité d'un patient d'être exposé compte tenu de ses caractéristiques. Cependant, à la différence du PS, la méthode de l'hdPS ne se limite pas aux variables sélectionnées par l'investigateur, mais inclut également des covariables identifiées grâce à un algorithme standardisé [110].

(2) L'algorithme

L'algorithme de l'hdPS est composé de 7 étapes distinctes qui peuvent être regroupées en trois parties (Figure 9) :

- Identification des covariables
- Priorisation et sélection des covariables
- Estimation et utilisation du hdPS.

Etape 1 : Sélection des données sources

Cette étape se décompose en deux sous-étapes. Premièrement, l'investigateur sélectionne les variables mesurées dans la base qu'il souhaite inclure dans l'estimation du hdPS. Deuxièmement, l'investigateur identifie p dimensions qui seront examinées par l'algorithme hdPS. Chaque dimension regroupe des codes de même type, et est identifiée selon un système de classification spécifique sur une période de temps considérée (par ex une dimension pour les codes de diagnostics, une dimension pour les codes de procédures, ...). Habituellement, les codes sont collectés sur les 6 ou 12 mois précédant l'exposition [117].

Etape 2 : Identification des covariables potentielles

Au sein de chaque p dimension, les codes sont triés selon leur prévalence. La prévalence est mesurée par la proportion de patients ayant le code étudié au moins une fois sur la période de temps considérée. Les prévalences au-dessus de 0.5 sont soustraites à 1 afin de tenir compte de la distribution binomiale des covariables. Il n'y a donc plus que des prévalences inférieures à 0.5. Ainsi, une prévalence de 30% du code A est équivalente à une prévalence de 70% du code B. Les n codes les plus prévalents sont identifiés comme des covariables potentielles (par défaut $n=200$).

Les codes ayant un nombre de patients inférieur à un seuil peuvent être exclus. Par défaut aucune exclusion n'est réalisée.

La prévalence de chaque code dépend de la granularité du codage. Par exemple, les codes de diagnostics sont hiérarchisés. Ainsi, lorsqu'on ajoute un chiffre au code on obtient plus de détails sur le diagnostic mais en revanche on diminue sa prévalence.

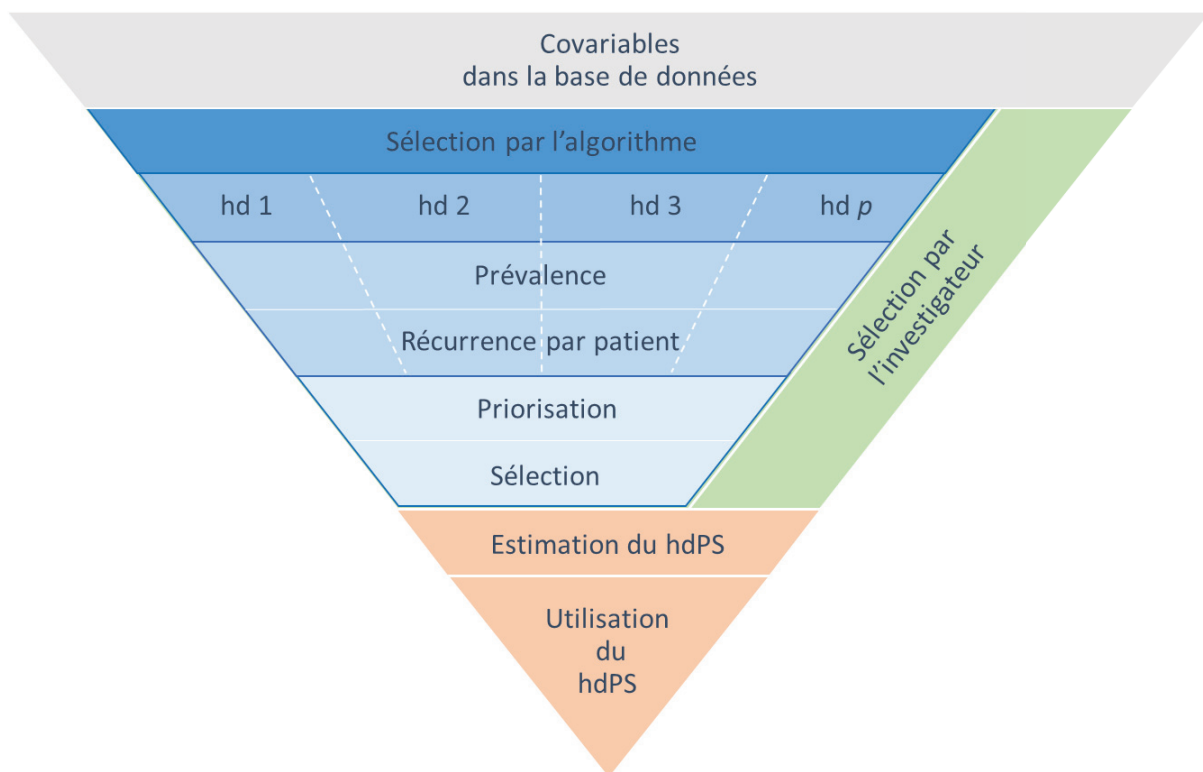


Figure 10 – Schéma représentant les différentes étapes de la méthode des scores de propension à haute dimension

Etape 3 : Identification de la récurrence des covariables

Pour les n codes sélectionnés à l'étape précédente, l'algorithme identifie combien de fois le code est présent pour chaque patient sur la période considérée.

Dans un premier temps, l'algorithme calcule le nombre de fois que chaque code n apparaît pour chacun des patients dans la période considérée. Dans un deuxième temps, la médiane et le troisième quartile de la distribution de chaque code sont calculés. Enfin, pour chaque patient, trois covariables sont créées en fonction du nombre de fois que le code apparaît chez le patient :

- CovX_1=1 si le code apparaît au moins une fois sur la période considérée
- CovX_median=1 si le code apparaît plus de fois que la valeur médiane
- CovX_Q3=1 si le code apparaît plus de fois que la valeur du troisième quartile.

Comme le nombre de fois qu'apparaît un code peut être faible, si la médiane est égale à 0 ou 1 alors CovX_median n'est pas créée. De même si le troisième quartile est égal e à 0, 1 ou la médiane, CovX_troiquart n'est pas créée (Table 4).

Ainsi, à cette étape, nous avons p dimensions avec chacune $n*3$ covariables au maximum.

Table 4- Exemple hypothétique de la création des covariables de récurrence d'une covariable sélectionnée à l'étape 2 :

Patient	Récurrence du code A chez le patient	CodeA_1	CodeA_median	CodeA_Q3
1	0	0	0	0
2	3	1	1	0
3	2	1	0	0
4	6	1	1	1
...

Dans, cet exemple, la médiane de la récurrence du code A dans la population étudiée est égale à 3 et le 3^{ème} quartile du code A dans la population est égal à 5. La variable CodeA_1, indique que la récurrence du code A chez le patient est ≥ 1 sur la période considérée. La variable CodeA_median, indique que la récurrence du code A chez le patient est ≥ 3 sur la période considérée. La variable CodeA_Q3, indique que la récurrence du code A chez le patient est ≥ 5 sur la période considérée.

Etape 4 : Priorisation des covariables

L'algorithme calcule pour chaque covariable r créée à l'étape précédente, le biais correspondant à cette variable, en utilisant la formule de Bross (paramètre par défaut) [110,118]. Il existe également deux méthodes mais qui sont peu utilisées [110].

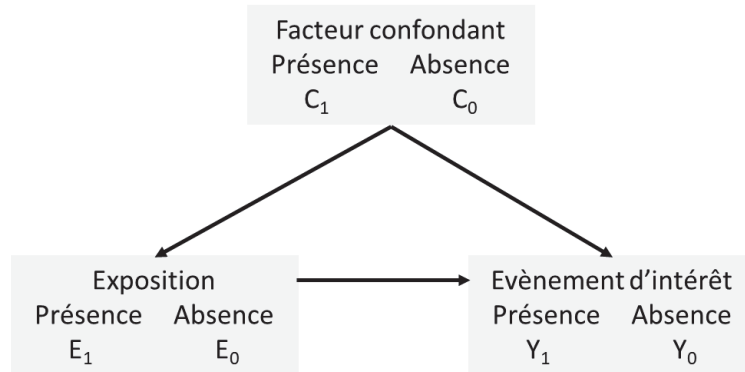


Figure 11 – Schéma représentant un facteur confondant et permettant d'introduire les notations de la formule de Bross.

Dans la formule de Bross, le biais dépend de la prévalence du facteur confondant dans les deux groupes étudiés ainsi que de l'association de ce facteur avec l'évènement étudié (Figure 12) :

$$BIAS = \frac{P_{C_1E_1}(RR_{CY} - 1) + 1}{P_{C_1E_0}(RR_{CY} - 1) + 1}$$

Où $P_{C_1E_1}$ correspond à la prévalence du facteur confondant dans le groupe exposé, $P_{C_1E_0}$ correspond à la prévalence du facteur confondant dans le groupe non exposé, et RR_{CY} correspond à l'association indépendante entre le facteur confondant et l'évènement étudié :

$$RR_{CY} = \frac{P_{Y_1C_1}}{P_{Y_1C_0}}$$

Où $P_{Y_1C_1}$ correspond à la prévalence de l'évènement d'intérêt dans le groupe présentant le facteur confondant et $P_{Y_1C_0}$ correspond à la prévalence de l'évènement d'intérêt dans le groupe ne présentant pas le facteur confondant.

Lors du calcul du RR_{CY} , il est possible d'appliquer une correction pour les cellules où il y aurait un zéro dans la table de contingence confondant croisant facteur et évènement d'intérêt par l'ajout de la constante 0.1.

Par la suite, c'est le $|\log(BIAS)|$ qui est utilisé. Il est égal à zéro lorsque :

- Les prévalences du facteur confondant dans les deux groupes sont les mêmes ($P_{C1E1} = P_{C1E0}$)
- Il n'existe pas d'association entre le facteur confondant et l'évènement étudié ($RR_{CY} = 1$)

Le $|\log(BIAS)|$ augmente :

- Plus l'écart entre les prévalences du facteur confondant dans les deux groupes est élevé
- Plus l'association entre le facteur confondant et l'évènement étudié est forte.

Etape 5 : Sélection des covariables

Les covariables avec les biais les plus élevés sont sélectionnées pour le modèle de score de propension. Ainsi, les covariables issues des différentes dimensions sont regroupées et triées dans l'ordre décroissant selon le $|\log(BIAS)|$. Les k premières covariables (par défaut $k=500$) sont sélectionnées.

Etape 6 : Estimation du hdPS

Le hdPS est estimé pour chaque patient en utilisant une régression logistique multivariée. Il correspond donc à la probabilité pour un patient d'être exposé compte tenu des variables sélectionnées par l'investigateur ainsi que des k variables sélectionnées par l'algorithme.

Etape 7 : utilisation du hdPS

La dernière étape de l'algorithme consiste à utiliser l'hdPS de la même manière qu'avec un score de propension classique : en variable d'appariement ou de stratification, en ajustement dans un modèle de régression ou pour la pondération inverse.

g) Les variables instrumentales

Largement utilisée par les économistes, la variable instrumentale permet de tenir compte à la fois des facteurs confondants mesurés et non mesurés [119–124]. L'utilisation de cette méthode se déroule en deux étapes : 1) l'identification d'une variable instrumentale spécifique à l'association étudiée, et 2) l'intégration de cette variable instrumentale dans un

modèle de régression décrivant la relation entre l'exposition et l'événement d'intérêt. Pour être valide, une variable instrumentale doit être fortement liée à l'exposition étudiée, ne pas avoir d'effet sur l'évènement d'intérêt excepté à travers l'exposition et ne pas être liée à un facteur confondant (Figure 13) [119–124].

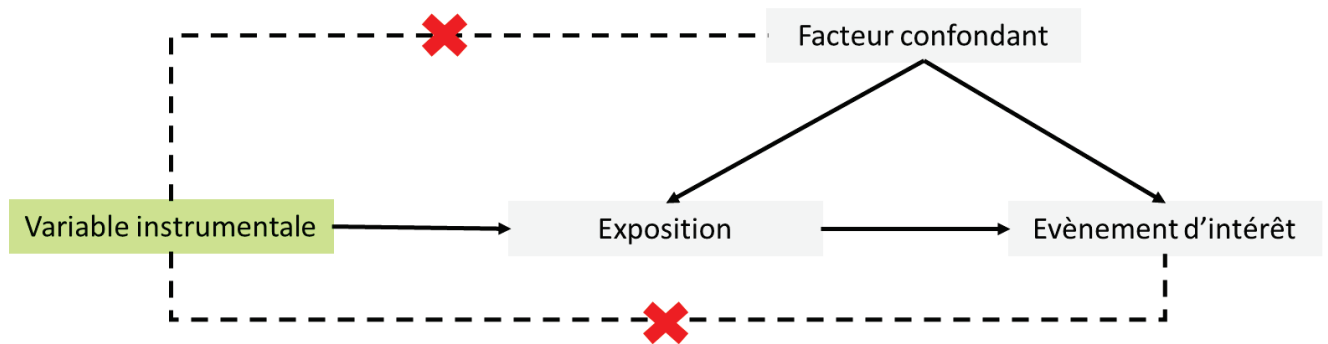


Figure 13 – Schéma représentant la méthode de la variable instrumentale

Ainsi, une variable instrumentale peut être considérée comme une variable observée qui est associée à une variation de l'exposition semblable à celle observée lors d'une randomisation dans les essais cliniques. Par conséquent, les distributions des caractéristiques mesurées et non mesurées sont équilibrées entre les groupes d'exposition [119–124].

Les variables instrumentales les plus utilisées en chirurgie sont en rapport avec la disponibilité des ressources clés (ex : réalisation de cathétérismes cardiaques par l'hôpital le plus proche [125]), les variations géographiques d'utilisation de techniques ou de pratiques (ex : taux régional de cathétérisme[126]), les incitations économiques et/ou les coûts des traitements alternatifs (ex : honoraires moyens des médecins pour la chirurgie de conservation du sein et la mastectomie [127]) et les tendances séculaires (ex : le jour de la semaine d'admission [128]) [4]. Dans ces exemples, les variables choisies comme variables instrumentales sont associées au choix d'être exposé ou non mais elles n'ont pas d'effet direct sur l'évènement d'intérêt.

Néanmoins ces exemples restent discutables et le choix ainsi que les conditions d'utilisation des variables instrumentales n'étant pas toujours faciles à mettre en œuvre, leur utilisation reste limitée [122,124].

II. Objectifs

L'objectif de ce travail était d'améliorer le contrôle des facteurs confondants lors de l'étude de résultats cliniques dans les bases hospitalières nationales en tenant compte de la finalité de l'étude.

Cet objectif a été décomposé en 3 sous-objectifs :

A. Evaluer la qualité méthodologique des études analysant la mortalité chirurgicale dans les bases de données hospitalières nationales

B. Contrôler les facteurs confondants lors de l'étude du lien volume-résultat en chirurgie dans les bases de données hospitalières nationales

C. Contrôler les facteurs confondants lors de la comparaison de l'efficacité réelle entre différentes interventions dans les bases hospitalières nationales

III. Travaux scientifiques répondant aux objectifs

A. Qualité méthodologique des études analysant la mortalité chirurgicale dans les bases de données hospitalières nationales : une revue de la littérature (article 1)

Les bases de données hospitalières nationales reflètent l'activité de routine des hôpitaux et ne répondent pas à une question spécifique de recherche. Certaines données peuvent être manquantes ou de qualité insuffisante pour contrôler divers facteurs confondants. C'est particulièrement le cas lorsqu'on souhaite étudier des résultats chirurgicaux puisque la prise en charge chirurgicale peut être influencée par des caractéristiques spécifiques au patient, à la procédure, au chirurgien, à l'équipe chirurgicale et à l'hôpital. De plus, lorsqu'on va comparer des résultats chirurgicaux entre des groupes prédéfinis, une différence systématique peut exister en l'absence de randomisation. L'amélioration des pratiques chirurgicales au cours du temps ainsi que la variabilité du codage entre les hôpitaux peuvent également influencer les résultats chirurgicaux. Idéalement, l'ensemble de ces facteurs devraient être considérés dans l'analyse.

Dans ce premier travail, nous avons souhaité faire un état des lieux des méthodes utilisées pour contrôler les facteurs confondants lors de l'analyse de résultats chirurgicaux dans les bases hospitalières nationales. Nous nous sommes donc intéressés à la qualité méthodologique des études analysant la mortalité chirurgicale dans ces bases de données [129]. Nous nous sommes focalisés sur la mortalité chirurgicale car c'est un des résultats chirurgicaux les plus étudiés pour décrire la qualité des soins dans les hôpitaux. Il s'agit d'un résultat critique pour les patients et sa mesure est valide et précise dans les bases hospitalières.

Nous avons donc réalisé une revue de la littérature en suivant les recommandations de la méthode PRISMA [130]. Notre stratégie de recherche a été réalisée dans Pubmed en ciblant six journaux (*New England Journal of Medicine, Lancet, British Medical Journal, Journal of the American Medical Association, Medical Care, Annals of Surgery*) et en sélectionnant

l'ensemble des citations disponibles jusqu'au 31 décembre 2014. A deux, nous avons revu indépendamment l'ensemble des citations. Puis, à partir d'une grille standardisée, nous avons extrait les informations relatives à la question de recherche et au design d'étude, à la source de données, à la sélection de la population, à la définition de la mortalité chirurgicale, aux facteurs confondants, à l'analyse statistique et aux analyses de sensibilités.

La qualité méthodologique des études a été estimée à partir de cinq critères et décrite au cours du temps.

Article 1: Payet C, Lifante JC, Carty MJ, Rabilloud M, Duclos A. Methodological Quality of Surgical Mortality Studies Using Large Hospital Databases: A Systematic Review. *Ann Surg.* 2017 Jun;265(6):1113-1118

Methodological Quality of Surgical Mortality Studies Using Large Hospital Databases

A Systematic Review

Cécile Payet, MPH,*† Jean-Christophe Lifante, MD, PhD,†‡ Matthew J. Carty, MD,§
Muriel Rabilloud, MD, PhD,¶|| and Antoine Duclos, MD, PhD*†§

Objective: To review the methodology employed in surgical mortality studies to control for potential confounders.

Summary Background Data: Nationwide hospital data are increasingly used to investigate surgical outcomes. However, poor data granularity and coding inaccuracies may lead to flawed findings.

Methods: We conducted a systematic review in accordance with the PRISMA statement in 6 major journals (*NEJM, Lancet, BMJ, JAMA, Medical Care, Annals of Surgery*) using PubMed from its inception until December 31, 2014. Two reviewers independently reviewed citations. Using a predesigned data collection form, we extracted information about study aim and design, data source, selected population, outcome definition, patient and hospital adjustment, statistics, and sensitivity analyses. The methodological quality of studies was assessed based on 5 criteria and explored over time.

Results: Among 89 included studies from 1987 to 2014, 54 explored surgical mortality determinants, 13 compared surgical procedure effectiveness, 13 evaluated the impact of healthcare policy, and 9 described outcome trends for specific procedures. A total of 89% (n = 79) of studies did not describe population selection criteria at patient and hospital level, 64% (n = 57) did not consider secular trends, 52% (n = 46) neglected hospital clustering or characteristics, 21% (n = 19) did not perform sensitivity analyses, and 4% did not adjust outcomes for patient risk (n = 4). The percentage of studies satisfying at least 3 of these criteria increased significantly from 44% before 1999 to 52% between 2000 and 2009 and 78% after 2010 (P = 0.008).

Conclusions: Although methodological quality of studies has improved over time, confounder control could be improved through better study design, homogeneous population selection, the consideration of hospital factors and secular trends influencing surgical mortality, and the systematic performance of sensitivity analyses.

Keywords: hospital databases, methodological quality, surgical mortality, systematic review

(*Ann Surg* 2016;xx:xxx–xxx)

From the *Department of Medical Information Evaluation and Research, Lyon University Hospital, Lyon, France; †Health Services and Performance Research Lab (EA 7425 HESPER), Lyon 1 Claude Bernard University, Lyon, France; ‡Department of General and Endocrine Surgery, Lyon University Hospital, Lyon, France; §Center for Surgery and Public Health, Brigham and Women's Hospital—Harvard Medical School, Boston, MA; ¶Department of Biostatistics, Lyon University Hospital, Lyon, France; and ||Biometry and Evolutionary Biology Laboratory (CNRS, UMR5558), Lyon 1 Claude Bernard University, Villeurbanne, France.

The authors report no conflicts of interest.

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 Web site (www.annalsofsurgery.com).

Reprints: Cécile Payet, MPH, Pôle Information Médicale Evaluation Recherche des Hospices Civils de Lyon, 162, avenue Lacassagne, Lyon, F-69424, France. E-mail: cecile.payet02@chu-lyon.fr

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

ISSN: 0003-4932/16/XXXX-0001

DOI: 10.1097/SLA.0000000000002119

Nationwide hospital databases are frequently used to explore time trends in surgical outcomes, identify factors associated with better care, compare procedural effectiveness, or measure the impact of health policies.¹ These large datasets are readily available in most industrialized countries, allowing population-based investigations of routine surgical practice at low cost. However, there are several precautions to adopt when using these data for research. Hospital administrative data is typically generated for billing purposes. As a consequence, studies based on this data may be flawed by the lack of accuracy and high variability in data collection depending on coder motivation.^{2–4} Furthermore, insufficient data granularity or unrecorded variables may limit outcome adjustment for specific confounders.⁵

Surgical mortality has been widely utilized to depict quality of care in hospitals because it is a critical outcome for patients and a reliably accurate metric in administrative databases.⁶ However, several factors may influence mortality independent of surgical care delivery, which may limit a fair interpretation of its variations. When comparing surgical mortality between predefined groups, systematic differences can exist in the absence of randomization.⁵ Risk of death can be heavily influenced by the interplay of characteristics specific to patient, procedure, surgeon, operating room team, and hospital.⁷ Furthermore, patients treated and care provided within a particular hospital tend to be more similar than those in another hospital, with the result being that observed outcomes for different individuals cannot be regarded as independent due to clustering effect.⁸ Finally, the confluence of many improvements in surgical practice with coding variation across hospitals can also influence surgical mortality over time under a secular trend.⁹ Ideally, all these factors should be considered in analyses. However, many confounders are missing or poorly described in administrative databases, which may be partly resolved in research investigations through appropriate study design and control group choice.¹

We conducted a methodological evaluation of studies investigating surgical mortality through assessment of administrative databases. Specifically, we aimed to describe study design, mortality definitions, population selection strategies, statistical modeling, and sensitivity analyses employed to control for potential confounders.

METHODS

Data Sources and Search Strategy

The systematic review was conducted in compliance with PRISMA guidelines.¹⁰ A comprehensive search was undertaken using PubMed from its inception through December 31, 2014. The search strategy was devised with the assistance of a research librarian. The search included 3 domains of MeSH terms and keywords combined using “AND,” whereas each domain was created using “OR.” The first domain contained terms related to surgery and surgical procedures, the second contained terms related to mortality,

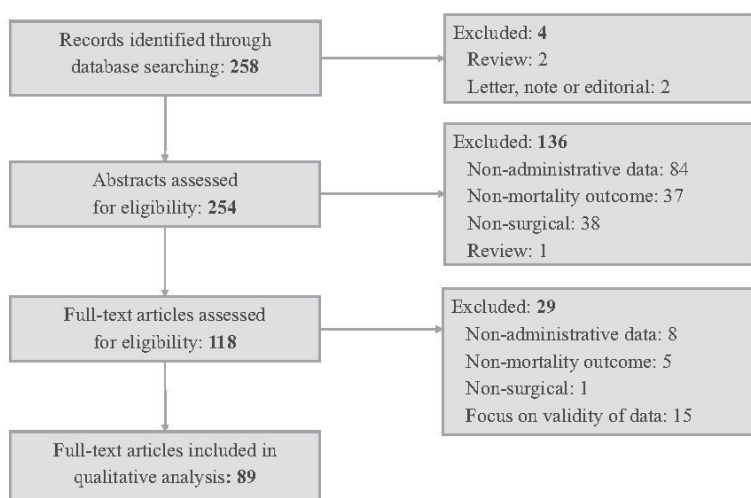


FIGURE 1. Flow chart of study selection.

and the third contained terms related to administrative databases (Appendix S1, <http://links.lww.com/SLA/B162>).

We focused our search strategy on the 4 generalist journals presenting the highest number of articles published in our theme (*New England Journal of Medicine*, *Lancet*, *Journal of the American Medical Association*, and *British Medical Journal*) within the category “Medicine, General and Internal” of the Web of Science Journal Citation Report (JCR). Based on the same rule, we also selected 2 specialist journals, 1 journal of the JCR category “Surgery” (*Annals of Surgery*), and 1 journal of the JCR “Health Care Sciences & Services” (*Medical Care*).

Study Selection

Two investigators (CP and AD) reviewed the titles and abstracts of all studies resulting from the database search. Nonoriginal articles (eg, letters, reviews, and editorials), studies not evaluating surgical mortality as well as studies using nonadministrative database were excluded. To ensure that all potentially relevant studies were considered for inclusion, every study with uncertain eligibility at title and abstract review was retained for further examination, especially when this initial review did not allow us to categorize paper as an original article based on surgical mortality from administrative data. Full-text documents were obtained for selected articles and 2 investigators (CP and AD) independently examined all studies to decide final inclusion in this review. Any disagreements were resolved by consensus.

Data Extraction

Data were extracted from the selected articles using a pre-designed collection form and checked independently by 2 authors (CP and AD). Specific data points collected included: study aim and design, data source, selected population, surgical mortality definition, patient and hospital confounders, and statistical/sensitivity analyses.

Papers were classified into 4 categories depending on their major study aim, as follows: time trend analysis, outcome determinants exploration, comparative effectiveness research, or health policy impact. Time trend studies described patterns of change in surgical mortality over time. Determinant studies identified factors influencing surgical mortality at the patient or hospital level. Comparative effectiveness studies compared the relationship between different surgical procedures and postoperative mortality. Health policy studies evaluated the impact of healthcare programs on surgical mortality.

Study design was described according to comparator choice to achieve comparable groups, time trend integration, and matching scheme between patients or hospitals. Three different study designs were reported based on comparator group choice. Before-and-after studies compared outcome before and after exposure within a unique group. Studies with contemporaneous comparators compared outcomes after exposure between exposed and nonexposed groups. Difference-in-differences studies compared outcome before and after exposure between exposed and nonexposed groups.

To define surgical mortality, we distinguished intra- from extra-hospital death. We also specified the period for death occurrence from start (at hospital admission or the day of surgical procedure) until the end (at hospital discharge, 30-d, 90-d, 1-yr). Descriptions of statistical analyses included adjustment on the characteristics of patients or providers and on secular trends, model type and methods to take into account the clustering of patients within hospitals. Finally, to assess the methodological quality of each study, we defined a 5-item score. These items included 1) the presentation of selection criteria at the patient and hospital level; 2) the adjustment for patient characteristics in multivariate models; 3) the consideration of hospital characteristics or clustering in these models; 4) the control of secular trends in statistical analyses; and 5) the performance of sensitivity analyses to test the robustness of main findings.

Statistics

Characteristics of the selected studies were presented using absolute and relative frequencies for categorical variables. Continuous variables were described using mean and standard deviation or median with minimum and maximum.

To explore trends in studies’ characteristics and methodological quality over time, papers were classified according to their respective aim and the year of publication. We used the Cochran–Armitage test to highlight trends between categories. Data manipulation and analyses were performed using SAS software (version 9.3; SAS Institute Inc, Cary, NC).

RESULTS

Characteristics of Studies

The search strategy retrieved 258 citations (Fig. 1). After screening titles, abstracts, and manuscripts, 89 studies were included

TABLE 1. Study Characteristics

	n = 89	%
Aim		
Time trend	9	10
Determinant exploration	54	61
Comparative effectiveness	13	15
Health policy	13	15
Data and population		
Study period in years, median (min–max)	6 (1–18)	—
No. of patients, median (min–max)	55,830 (502–9,474,879)	—
No. of hospitals, median (min–max)*	531 (39–2691)	—
Outcome†		
In-hospital mortality	44	49
In/out-hospital mortality	44	49
Not specified	6	7
Patient confounders considered		
Demographic	82	96
Comorbidities	74	87
Emergency	38	45
Socioeconomic	32	38
Year of procedure	30	35
Procedure type	18	21
Primary diagnosis	16	19
Admission day	3	4
Procedure complexity	1	1
Hospital confounders considered		
Volume	20	65
Hospital status	16	52
Bed size	13	42
Rural/urban	13	42
Geographic area	12	39

*Forty-nine studies did not report hospital number.
 †Studies could be used in-hospital mortality and in/out-hospital mortality.

for analysis (Appendix S2, <http://links.lww.com/SLA/B162>). Among these studies, 18 were published through 1999, 31 between 2000 and 2009, and 40 from 2010. Overall, 74 studies originated from the United States and used primarily Medicare data or nationwide inpatient samples, 11 originated from the United Kingdom and were mainly based on hospital episode statistics, 4 originated from Canada, and 1 from France. Regarding study aim, 61% (54/89) of studies explored surgical mortality determinants while 39% (35/89) focused on time trend analysis, procedural effectiveness, or health policy impact (Table 1, Fig. 2).

Study populations included procedures focused on the following organ systems: digestive (53 studies), cardiothoracic (42), orthopedic (15), respiratory (13), urinary (6), nervous (1), and genital (1). Four studies did not discriminate procedure type and therefore presumably included interventions spanning multiple organ systems.

There was no consensus in the definition of surgical mortality among reviewed papers. Surgical mortality was not defined in 6 studies. Twenty-four studies failed to identify whether surgical mortality was defined from the time of hospital admission or from the time of surgical procedure. Among 44 studies investigating in-hospital mortality, 77% (34/44) defined mortality as deaths occurring before hospital discharge, 25% (11/44) within 30 days, and 1% (1/44) within 14 days. It was not specified whether in-hospital mortality included all postoperative deaths or only focused on those occurring within surgical wards. Among 44 studies investigating in/out-hospital mortality, 45% (20/44) defined mortality as deaths occurring within 30 days, 9% (4/44) within 90 days, 11% (5/44) within 1 year, and 41% (18/44) of studies measured deaths before hospital discharge or within 30 days for patient discharged before that date.

Patient confounders were considered in 85 studies while hospital confounders were considered in 31 studies. Among 74 studies with adjustment for patient comorbidities, 38% (28/74) used

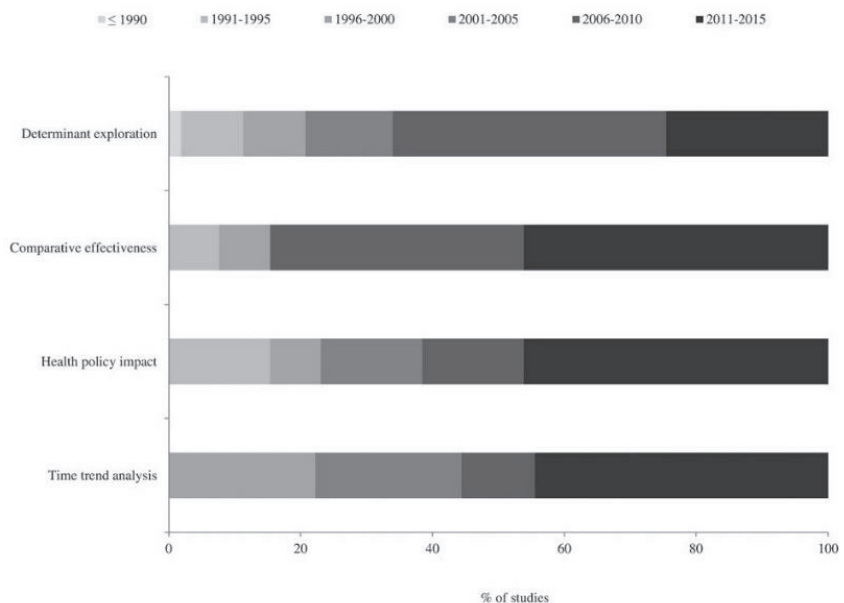


FIGURE 2. Percentage of studies according to study aim and publication year.

TABLE 2. Study Aim and Design

Study Aim	Time Trend Analysis n = 9 (%)	Determinant Exploration n = 54 (%)	Comparative Effectiveness n = 13 (%)	Health Policy Impact n = 13 (%)
Comparator				
Contemporaneous group	0 (0)	54 (100)	13 (100)	8 (61)
Before and after	0 (0)	0 (0)	0 (0)	1 (8)
Difference-in-differences	0 (0)	0 (0)	0 (0)	4 (31)
Time series	9 (100)	16 (30)	3 (23)	4 (31)
Matching	0 (0)	2 (4)	3 (34)	1 (8)

the Charlson index, 31% (23/74) the Elixhauser index, and 31% (23/74) a specific selection of comorbidities. Socioeconomic status was considered in 32 studies using insurance status (53% of studies, 17/32) or other data (56%, 18/32), based on income, education, or employment corresponding to the patient’s residence code.

Surgical mortality was modeled using logistic regression (76 studies), cox model (3), and Poisson regression (2). Clustering was considered in 38 studies using Generalized Estimating Equations models (58% of studies, 22/38) or mixed models (34%, 13/38) while 3 studies did not report the model used.

Design and Methodological Quality of Studies

A contemporaneous control group was used as comparator in all the determinant and comparative effectiveness studies (Table 2). Regarding the health policy studies, 8 used a contemporaneous control group, 4 executed a difference-in-differences study design, and 1 employed a before-and-after design. In addition to time trend studies, 23 studies used multiple measurements over time to consider secular trends. Propensity score matching was used in 6 studies. The matching was performed at the patient level in the determinant and comparative effectiveness studies, while hospitals were matched in 1 health policy study.

Figure 3 shows that population selection criteria at both the patient and hospital levels were not described in 79 studies, without improvement over time. Adjustment for patient characteristics was common, but 46 studies failed to consider the influence of hospital

characteristics on mortality or patient clustering. However, the percentage of studies that took into account hospital factors increased from 28% (5/18) until 1999 to 42% (13/31) between 2000 and 2009, and 63% (25/40) after 2010 (Fig. 3, *P* = 0.010). Secular trends during the study period were not considered in 57 studies and there was no improvement over time. Finally, 70 studies conducted sensitivity analyses, increasing from 56% (10/18) until 1999 to 77% (24/31) between 2000 and 2009, and 90% (36/40) after 2010 (Fig. 3, *P* = 0.003). These sensitivity analyses were focused on the testing of secondary outcomes (40 studies), other statistical approaches (29), other definitions of surgical mortality (18), population selection (14), and study design (1).

Regarding the 5-item score, 1 study did not meet any item, 10 met 1 item, 23 met 2 items, 37 met 3 items, 16 met 4 items, and 2 all 5 items. The percentage of studies with at least 3 items increased significantly from 44% (8/18) until 1999 to 52% (16/31) between 2000 and 2009 and 78% (31/40) after 2010 (*P* = 0.008).

DISCUSSION

Nationwide hospital data are increasingly used to investigate surgical mortality, with a special focus on analyzing time trends, comparing procedural effectiveness, or measuring the impact of health policy initiatives. We do not necessarily have tangible explanations to explain the trend toward relative increase of those topics in

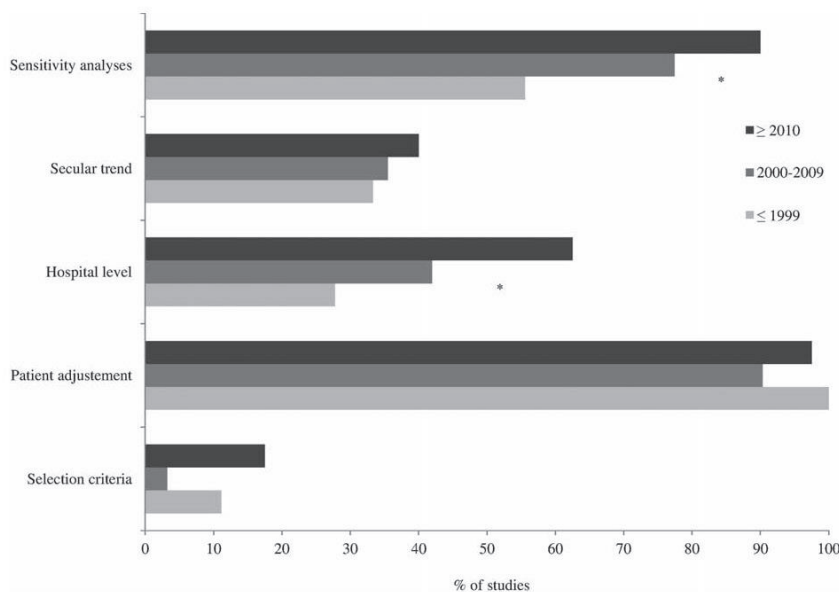


FIGURE 3. Percentage of studies according to methodological quality items and publication year. Selection criteria were at both patient and hospital levels. Hospital level corresponded to hospital adjustment or patient clustering within hospitals. **P* value <0.05, Cochran–Armitage test trend.

relation to traditional research on mortality determinants. However, recent emphasis in the fields of health services research with growing funding requires the utilization of large administrative data reflecting routine practice in large populations.¹¹ The accumulation of nationwide data over decades now allows refinements in designing longitudinal studies with contemporary control groups for investigating real changes in surgical outcomes caused by new policies or procedures. Although the methodology of studies has improved over time, this systematic review of publications from flagship journals highlights the fact that basic quality standards have not been systematically met. Nearly all publications still fail to report both patient and hospital selection criteria to homogenize study populations. A majority of studies do not take into account the influence of hospital effect, secular trend, or coding variation that may alter data interpretation in statistical analyses. Finally, sensitivity analyses are not always performed to test the robustness of study findings. However, several improvements in methodology are possible to manage potential confounders, through appropriate study design, data selection, and analysis.

No amount of sophisticated statistical analyses can compensate for poor study design. On the one hand, using before-and-after study design to measure the association between the intervention effect and outcome within a unique group is valid only if there are no underlying time-dependent trends in outcomes unrelated to the effect of intervention.¹² On the other hand, using a study with contemporaneous control comparators is valid only if baseline outcomes are the same between exposed and nonexposed groups before the start of intervention. Opting for a difference-in-differences study design can address these issues by comparing outcomes before and after exposure to the intervention between exposed and nonexposed groups.¹³ In this way, using a control group that is experiencing the same secular trends but is not exposed to the intervention allows the investigators to subtract out the background changes in outcomes if 2 assumptions are met: 1) that underlying trends in outcomes before the intervention should be the same between groups although these groups may have different outcomes at baseline; and 2) that phenomena occurring concurrent to the intervention will affect both groups equally.¹⁴ These assumptions require multiple measurements for each group to verify time trends and initial comparability between groups and assure their preservation during the entire study period.¹⁵ If these assumptions are not met, multiple measurement design could be used to take into account different trends.

One way to obtain group comparability is to choose well-matched samples of the original exposed and control groups, thereby reducing bias risk due to potential confounders.¹⁶ When the study purpose is to assess the impact of a specific healthcare policy, matching hospitals can be worthwhile for controlling hospital characteristics that may drive surgical outcomes irrespective of policy implementation.¹⁷ If a study aims at measuring the association between patient determinants and mortality or the effectiveness between surgical procedures, patient matching is of great interest;¹⁸ specifically, patient matching can be stratified within hospitals¹⁸ to know whether patients exposed to a new procedure have better outcomes than those treated with the usual procedure by the same surgical team, with the advantage of controlling for potential confounders related to variations across hospitals in perioperative care and data coding. When a large number of variables influence group choice, it can be appropriate to use propensity score analyses to balance groups according to distributions of measured covariates.^{16,19} These variables may collectively be proxies for unobserved factors; therefore, the high-dimensional propensity score method has been suggested to identify and prioritize covariables within the investigational database.^{5,19,20} Instrumental variables can also be used to control measured and unmeasured confounders. This method

employs a variable called an instrument that is predictive of treatment but has no effect on the outcome, except through its influence on treatment assignment.²¹

Given the complexity of the underlying medical and socioeconomic processes that determine surgical outcomes as well as the capacity of nationwide data for exploring healthcare, a substantial uncertainty persists about how to select study populations, to design the study and to specify statistical models. Therefore, systematic sensitivity analyses can be conducted to ensure the consistency of results. Traditionally, sensitivity analyses require changing the measurement, but they can also be performed at each step of the study for testing different control groups, population of patients, surgical procedures, and statistical models.

For example, a feature of large hospital databases is their hierarchical structure, since distinct clusters of patients are treated within different hospitals. Patients admitted to a given hospital may share common characteristics and have correlated outcomes, thus violating the independence assumptions of traditional regression models. The main consequence of not taking account of this correlation is a misinterpretation of statistical tests²² and inadequate precision of the estimates with too narrow confidence intervals.^{23,24} To avoid such a pitfall, random effects models and generalized estimating equations models²³ are commonly used and could be routinely incorporated into statistical analyses or at least into sensitivity analyses. Furthermore, logistic regression is commonly used to model in-hospital mortality, but this is unsuitable when follow-up varies from 1 patient to another or the hospital discharge represents a competing risk of death. A solution to study in-hospital mortality is to employ competing risk models in survival data analyses, integrating time to death and length of hospital stay.^{25,26} Finally, when surgical mortality is investigated over several years, it can be interesting to incorporate a time variable in the model because unmeasured confounders can be associated with background changes like coding variations.^{12,27}

The main limitation of this review was the noninclusion of all articles published in our theme for feasibility reason. Although we were not representative of studies published in all biomedical literature, we assumed that the editorial policies of selected journals were based on high-quality standards in selecting submitted works. We therefore believe that the methodological quality of papers retained in this review is above average relative to that of the entire range of studies in our theme. A further limitation was the marked heterogeneity in the aims and populations of selected studies. A broad search strategy was used intentionally to capture a full range of studies across all surgical specialties.

In 2015, the Reporting of studies Conducted using Observational Routinely-collected health Data checklist was created as an extension to the STROBE statement to address reporting items appropriate to observational studies using routinely collected health data.²⁸ This checklist encourages authors to clearly state codes, algorithms, and linkages between databases used to select study populations and define exposures, outcomes, and confounders. However, it would be worthwhile adapting this checklist for the assessment of large hospital databases that gather different data granularity levels (ie, hospital, patient, stay) across several years. This review provides a thorough evaluation of methodologies employed in studies using these data and highlights several key findings for investigating surgical outcomes. In support to the Reporting of studies Conducted using Observational Routinely-collected health Data statement, confounder control could be improved through appropriate study design, homogeneous population selection, the consideration of hospital factors, and secular trends influencing surgical mortality, and the systematic performance of sensitivity analyses (Appendix S3, <http://links.lww.com/SLA/B162>).

REFERENCES

1. Dimick J, Greenberg C. *Success in Academic Surgery: Health Services Research*. Berlin, Germany: Springer; 2014.
2. Peabody JW, Luck J, Jain S, et al. Assessing the accuracy of administrative data in health information systems. *Med Care*. 2004;42:1066–1072.
3. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med*. 1997;127(8_Part_2):666–674.
4. Mohammed MA, Deeks JJ, Girling A, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ*. 2009;338:b780.
5. Brookhart MA, Stürmer T, Glynn RJ, et al. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010;48:S114–S120.
6. Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ*. 2007;334:1044.
7. Sittig DF, Classen DC. Safe Electronic Health Record use requires a comprehensive monitoring and evaluation framework. *JAMA*. 2010;303:450–451.
8. Rice N, Leyland A. Multilevel models: applications to health data. *J Health Serv Res Policy*. 1996;1:154–164.
9. Lindenauer PK, Lagu T, Shieh M, et al. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003–2009. *JAMA*. 2012;307:1405–1413.
10. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151:264–269.
11. Moses H 3rd, Matheson DH, Cairns-Smith S, et al. The anatomy of medical research: US and international comparisons. *JAMA*. 2015;313:174–189.
12. Chen Y-F, Hemming K, Stevens AJ, et al. Secular trends and evaluation of complex interventions: the rising tide phenomenon. *BMJ Qual Saf*. 2016;25:303–310.
13. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*. 2014;312:2401–2402.
14. Ryan AM, Burgess JF, Dimick JB. Why we should not be indifferent to specification choices for difference-in-differences. *Health Serv Res*. 2015;50:1211–1235.
15. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin; 2002.
16. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci*. 2010;25:1–21.
17. Reames BN, Scally CP, Thumma JR, et al. Evaluation of the effectiveness of a surgical checklist in medicare patients. *Med care*. 2015;53:87–94.
18. Silber JH, Rosenbaum PR, Kelz RR, et al. Examining causes of racial disparities in general surgical mortality: hospital quality versus patient Risk. *Med Care*. 2015;53:619–629.
19. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. *Med Care*. 2007;45:S143–S148.
20. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20:512–522.
21. Pezzin LE, Laud P, Yen TWF, et al. Reexamining the relationship of breast cancer hospital and surgical volume to mortality: an instrumental variable analysis. *Med Care*. 2015;53:1033–1039.
22. Austi PC, Tu JV, Alte DA. Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: should we be analyzing cardiovascular outcomes data differently? *Am Heart J*. 2003;145:27–35.
23. McCulloch CE, Searle SR, Neuhaus JM. *Generalized, Linear and Mixed Models*. New York: Wiley; 2011.
24. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*. 2010;21:467–474.
25. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170:244–256.
26. Varadhan R, Weiss CO, Segal JB, et al. Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. *Med Care*. 2010;48:S96–S105.
27. Lindenauer PK, Lagu T, Shieh MS, et al. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003–2009. *JAMA*. 2012;307:1405–1413.
28. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12:e1001885.

Ce travail nous a montré que les bases de données hospitalières nationales étaient de plus en plus utilisées pour étudier la mortalité chirurgicale.

Quatre grandes finalités d'étude ont été retrouvées :

- La description de l'évolution de la mortalité au cours du temps
- L'étude des déterminants de la mortalité au niveau du patient ou de l'hôpital
- La comparaison de la mortalité entre deux interventions chirurgicales en vie réelle
- L'évaluation de politiques de santé sur la mortalité

Bien que la qualité méthodologique des études se soit améliorée au cours du temps, des pistes d'amélioration étaient envisageables à chaque étape de l'étude pour contrôler de potentiels facteurs confondants :

- Au niveau du design d'étude
- Au niveau de la stratégie de sélection de la population
- Au niveau de l'analyse statistique
- Au niveau des analyses de sensibilité éventuelles réalisées à chaque étape de l'étude

Ce travail nous a permis de proposer trois travaux visant à mieux contrôler de potentiels facteurs confondants à un ou plusieurs des niveaux définis ci-dessus, dans des thématiques de recherche actuelles : la recherche de déterminants avec l'étude du lien volume-résultat des hôpitaux, et la comparaison de l'efficacité réelle de différentes interventions.

B. Reconsidérer le lien volume-résultat en chirurgie : étude TREVORS (article 2)

Depuis les années 1970, le lien volume-résultat est la thématique de recherche la plus étudiée en recherche sur les résultats. Cette thématique est également sujette à controverse puisqu'à l'heure actuelle, le débat est toujours ouvert sur l'existence de ce lien selon la chirurgie étudiée ou si l'on étudie le volume de l'hôpital ou du chirurgien [131]. Il nous paraissait donc intéressant d'étudier ce lien volume-résultat sous un angle nouveau afin d'enrichir le débat.

Il est couramment admis que le volume d'interventions réalisées par un hôpital est un élément déterminant de la réussite d'une chirurgie. Cette intuition repose sur une notion simple selon laquelle l'habitude des équipes chirurgicales à pratiquer certaines procédures est garante d'un meilleur résultat pour le patient. Ainsi, une équipe ne réalisant que rarement une certaine chirurgie l'accomplira moins bien et en gèrera moins bien les suites opératoires qu'une équipe l'ayant intégrée à son activité de routine. Cela peut expliquer l'expertise reconnue à certaines équipes dès lors qu'elles ont déjà pris en charge par le passé de nombreux cas similaires à ceux auxquels elles seront confrontées dans le futur [132].

De nombreuses études se sont donc intéressées à la relation entre le volume de chirurgies d'un établissement et la survenue de complications [25–33]. La majorité de ces études ont montré que des patients opérés dans un établissement avec un nombre élevé d'interventions avaient de meilleurs résultats chirurgicaux que des patients opérés dans un établissement avec un nombre plus faible d'interventions, en comparant à un instant donné les résultats chirurgicaux entre des établissements à haut volume et des établissements à bas volume (approche transversale).

La principale hypothèse avancée pour expliquer cette relation repose sur l'effet d'apprentissage (« *practice makes perfect* ») [133,134], les équipes chirurgicales tendant à améliorer leurs résultats progressivement lorsqu'elles sont amenées à opérer davantage de patients. L'apprentissage reposerait sur l'expérience accumulée des individus, le transfert du savoir et de l'encadrement, mais aussi une amélioration progressive de l'organisation des soins et un investissement croissant dans des équipements plus performants [53,54,58]. Ainsi, les centres étant parvenus à développer un haut volume de procédures disposeraient

d'équipes chirurgicales travaillant de manière plus coordonnée, de systèmes pour identifier précocement, puis réagir efficacement, à la survenue de complications [60], ou bien de programmes de réhabilitation destinés à améliorer la qualité du service rendu au patient. Une seconde hypothèse suppose une orientation préférentielle des patients vers les établissements obtenant les meilleurs résultats (« *selective referral* ») [29].

Sur la base de ces hypothèses, de nombreux pays ont encouragé leurs patients à s'orienter vers des centres à haut volume en fixant des seuils minimaux d'activité pour autoriser la réalisation de certaines chirurgies [30,31,135].

Néanmoins, l'approche transversale traditionnellement utilisée pour mesurer la relation volume-outcome paraît peu adaptée pour étudier cette relation si elle repose sur un effet d'apprentissage amené à évoluer au cours du temps. L'influence de l'évolution du volume de l'hôpital au cours du temps sur les résultats chirurgicaux du patient n'est pas encore clairement définie.

Les objectifs de cette étude étaient donc de déterminer si les résultats chirurgicaux des patients étaient meilleurs lorsque les hôpitaux augmentaient significativement leur volume de chirurgies au cours du temps et si un effet d'apprentissage individuel existait au niveau de l'hôpital.

Les patients opérés dans un établissement français entre 2010 et 2014, pour l'une des dix interventions digestives, cardio-vasculaires et orthopédiques étudiées, ont été identifiés dans la base PMSI MCO. Pour chaque intervention, les hôpitaux ont été répartis en 3 groupes selon l'évolution de leur volume d'interventions (augmentation, stabilité, baisse) sur 5 ans. La mortalité, la reprise chirurgicale et la réhospitalisation potentiellement évitable dans les 30 jours suivant l'intervention ou la sortie (pour la réhospitalisation), ont été comparées entre les groupes en utilisant des modèles de survie considérant la similitude des patients au sein d'un même hôpital et divers facteurs confondants liés aux caractéristiques des patients et des hôpitaux. L'effet d'apprentissage a été étudié en considérant l'interaction entre les groupes d'hôpitaux et l'année de l'intervention. Sur 5 ans, nous avons considéré 759 928 patients et 694 établissements.

Article 2: Payet C, Polazzi S, Lifante JC, Cotte E, Grinberg D, Carty MJ, Sanchez S, Rabilloud M, Duclos A. Influence of trends in hospital volume over time on patient outcomes for high-risk surgery. BMC Health Services Research (Accepté)

Influence of trends in hospital volume over time on patient outcomes for high-risk surgery

Cécile Payet^{ab}, Stéphanie Polazzi^{ab}, Jean-Christophe Lifante^{bc}, Eddy Cotte^c, Daniel Grinberg^d, Matthew J Carty^e, Stéphane Sanchez^f, Muriel Rabilloud^{gh}, Antoine Duclos^{abc}

^aHospices Civils de Lyon, Health Data Department, Lyon, F-69003, France

^bUniversité Claude Bernard Lyon 1, Health Services and Performance Research Lab (HESPER EA7425), F-69008 Lyon, France

^cHospices Civils de Lyon, Centre Hospitalier Lyon Sud, Service de Chirurgie Digestive et Endocrinienne, F-69300 Pierre Bénite, France

^dService de Chirurgie Cardio-thoracique et Transplantation, Hôpital Cardio-thoracique Louis Pradel, Lyon-Bron, Avenue du Doyen Lépine, 69500 Bron, France

^eBrigham and Women's Hospital, Harvard Medical School, Center for Surgery and Public Health, Boston, MA, USA

^fHôpitaux Champagne Sud, Centre Hospitalier de Troyes, Pôle Information Médicale Évaluation Performance, Troyes, France

^gHospices Civils de Lyon, Pôle de Santé Publique, Service de Biostatistique, Lyon, France

^hCNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biostatistique-Santé, Villeurbanne, France

Abstract

Background: The “practice makes perfect” concept considers the more frequent a hospital performs a procedure, the better the outcome of the procedure. We aimed to study this concept by investigating whether patient outcomes improve in hospitals with a significantly increased volume of high-risk surgery over time and whether a learning effect existed at the individual hospital level.

Methods: We included all patients who underwent one of ten digestive, cardiovascular and orthopaedic procedures between 2010 and 2014 from the French nationwide hospitals database. For each procedure, we identified three groups of hospitals according to volume trend (increased, decreased, or no change). In-hospital mortality, reoperation, and unplanned hospital readmission within 30 days were compared between groups using Cox regressions, taking into account clustering of patients within hospitals and potential confounders. Individual hospital learning effect was investigated by considering the interaction between hospital groups and procedure year.

Results: Over 5 years, 759,928 patients from 694 hospitals were analysed. Patients’ mortality in hospitals with procedure volume increase or decrease over time did not clearly differ from those in hospitals with unchanged volume across the studied procedures (e.g., Hazard Ratios [95%] of 1.04 [0.93-1.17] and 1.08 [0.97-1.21] respectively for colectomy). Furthermore, patient outcomes did not improve or deteriorate in hospitals with increased or decreased volume of procedures over time (e.g., 1.01 [0.95-1.08] and 0.99 [0.92-1.05] respectively for colectomy).

Conclusions: Trend in hospital volume over time did not appear to influence patient outcomes based on real-world data.

Background

The relationship between hospital surgical procedures volume and related mortality has been extensively investigated over the past several decades [1]. Numerous studies have reported that patients who undergo operations in hospitals performing a high number of procedures achieve better outcomes [2–10]. Based on these findings, some countries have strived to consolidate specific procedures in high-volume hospitals, with varying results [7, 8, 11, 12]. Consequently, significant methodological flaws regarding the validity of the volume-outcome relationship and the definition of evidence-based volume thresholds have been pointed out [13–17]. Although high-volume

hospitals may provide, on average, safer care than low-volume hospitals, some high-volume hospitals may perform poorly while some low-volume hospitals may perform well [18–20]. Nonetheless, hospital-volume continues to be used as a proxy quality metric for high-risk surgeries.

The concept of the volume-outcome relationship is based on the assumption that hospitals that perform a complex procedure more frequently have better outcomes and could manage adverse events more effectively than those who rarely perform the procedure [21, 22]. In the same manner that surgeon or team experience may determine a procedure-specific learning curve [23, 24], this assumption may recapitulate the “practice-makes-perfect” dogma at the institutional level [25, 26]. Most prior studies on

volume-outcome relationship have compared outcomes between low- and high-volume hospitals at some point. However, they did not consider the temporal relationship that may exist between volume and outcome under the influence of a dynamic learning effect. Consequently, it remains unclear whether changes in hospital procedural volumes influence patient safety. In this nationwide study, we took another look at the volume-outcome relationship in high-risk surgery by evaluating, at the individual hospital level, the association between trends in volume and patient outcome over time.

Methods

Study design and data source

We performed a nationwide observational study to determine whether patient outcomes improve in hospitals with a significantly increased volume of high-risk surgery over time and whether a learning effect existed at the individual hospital level. We first defined three groups of hospitals according to the trend of the volume of surgical procedures over a 5-year period, that is, for a given hospital, the volume of a specific procedure increased, decreased, or did not change. Second, we compared the average patient outcomes and their evolution over time between these three defined hospital groups, taking into account potential confounding factors related to hospital and patient characteristics. To test the robustness of our results, we repeated this scheme across ten high-risk surgical and interventional procedures in various specialties and considered different patient outcomes. We assumed that if patient outcomes would be influenced by volume change of procedures over time within individual hospitals, those results would be consistent whatever the procedure or outcome studied.

Moreover, the PMSI has a system of coding with strict variable definitions and a subset of records audited on a regular basis to avoid coding errors. Inpatient stays are converted into one Diagnosis-Related Group based on standard discharge abstracts containing compulsory information about the patient, primary and secondary diagnoses using the International Classification of Diseases (10th revision - ICD-10 codes), emergency status, and

procedural codes associated with the care provided using a detailed classification.

From the PMSI database, we extracted data on patient demographics, co-morbidities according to the Elixhauser algorithm [27], the type and emergency context of the procedure, and discharge by transfer to another acute care hospital. We also characterized each hospital according to its status (i.e., teaching, public, or private for-profit), degree of specialization (i.e., proportion of admissions logged for each studied procedure in the related surgical department), and attraction rate (i.e., the proportion of patients living in another geographical area than that of the hospital location where they underwent each studied procedure). To define patients' socioeconomic status, we extracted the median income household of patient residence code provided by the National Institute of Statistics and Economic Studies.

This study was approved by Institutional Review Board IRB00009118 (Sud-Est II ethical research committee) and the French Data Protection Authority (CNIL DE-2016-028), and it was registered on clinicaltrials.gov (NCT02788331). In accordance with French ethical directives, the requirement for written informed consent was waived because the study was strictly observational and all data were blinded.

Study population and outcomes

We included all patients who underwent one of the following ten procedures from January 1, 2010 to December 31, 2014: resection of a digestive cancer (i.e., colectomy, proctectomy, esophagectomy, gastrectomy, and pancreatectomy), intervention on the cardiovascular system (i.e., percutaneous coronary intervention [PCI]), coronary-artery bypass grafting [CABG], carotid endarterectomy, and elective repair of abdominal aortic aneurysm [AAA]), and urgent hip fracture repair. The choice to focus on those procedures was guided by available evidence suggesting the existence of volume-outcome relationships based on cross-sectional studies [2–5]. Each procedure was identified from the PMSI database by combining specific diagnoses and procedural codes.

For each studied procedure, all patients from hospitals not performing at least one procedure per

Table 1 - Hospital characteristics by procedure

	Hospitals number	Volume of procedures mean (SD)	Status			Specialization degree ^a mean (SD)	Attraction rate ^b mean (SD)
			Teaching	Public or private non-for-profit	Private for profit		
Colectomy	610	141.2 (87.5)	30 (4.9%)	319 (52.3%)	261 (42.8%)	5.3 (2.4)	13.8 (14.5)
Proctectomy	522	71.0 (54.2)	33 (6.3%)	297 (56.9%)	192 (36.8%)	2.5 (2.3)	16.2 (17.3)
Esophagectomy	62	37.0 (58.4)	23 (37.1%)	27 (43.5%)	12 (19.4%)	0.8 (1.1)	33.2 (26.7)
Gastrectomy	343	28.5 (22.7)	30 (8.7%)	184 (53.6%)	129 (37.6%)	0.9 (0.9)	16.9 (18.2)
Pancreatectomy	166	37.0 (39.0)	28 (16.9%)	83 (50.0%)	55 (33.1%)	0.9 (0.9)	23.5 (22.3)
PCI	221	1,268.6 (1,086.3)	23 (10.4%)	81 (36.7%)	117 (52.9%)	19.6 (10.1)	21.4 (16.5)
CABG	45	516.6 (502.2)	22 (48.9%)	18 (40.0%)	5 (11.1%)	7.8 (7.7)	49.3 (21.7)
AAA repair	207	77.0 (78.5)	23 (11.1%)	129 (62.3%)	55 (26.6%)	2.1 (1.8)	24.0 (21.2)
Carotid endarterectomy	296	215.6 (209.1)	19 (6.4%)	193 (65.2%)	84 (28.4%)	47.7 (26.5)	20.1 (19.6)
Hip fracture repair	421	558.5 (437.9)	23 (5.5%)	133 (31.6%)	265 (62.9%)	12.9 (7.9)	11.9 (12.5)

^a Proportion of stays for each studied procedure in the surgical department (expressed as a percentage)

^b Proportion of patients living in another geographical area than the one of hospital location where they underwent each studied procedure (expressed as a percentage)
PCI percutaneous coronary intervention, CABG coronary-artery bypass grafting, AAA abdominal aortic aneurysm

Table 2 - Patient characteristics by procedure

	Patient number	Men N (%)	Age mean (SD)	No. of Elixhauser comorbidities ^a Mean (SD)	Median income, K€ Mean (SD)	Emergency admission N (%)	Discharge by transfer N (%)
Colectomy	86,102	44,994 (52.3)	71.6 (12.2)	1.8 (1.6)	20.4 (3.3)	13,201 (15.3)	2,771 (3.2)
Proctectomy	37,088	22,862 (61.6)	68.5 (11.8)	1.5 (1.5)	20.3 (3.2)	1,651 (4.5)	1,229 (3.3)
Esophagectomy	2,296	1,858 (80.9)	62.3 (9.5)	2.4 (1.7)	20.0 (3.0)	93 (4.1)	221 (9.6)
Gastrectomy	9,777	5,695 (58.2)	68.8 (12.8)	2.0 (1.7)	20.3 (3.2)	929 (9.5)	500 (5.1)
Pancreatectomy	6,148	3,255 (52.9)	65.0 (11.2)	2.1 (1.7)	20.6 (3.4)	451 (7.3)	316 (5.1)
PCI	280,369	209,058 (74.6)	66.1 (13.1)	1.5 (1.4)	20.3 (3.2)	190,154 (67.8)	33,713 (12.0)
CABG	23,247	19,019 (81.8)	67.2 (10.4)	2.6 (1.8)	20.0 (2.8)	6,586 (28.3)	4,927 (21.2)
AAA repair	15,935	14,736 (92.5)	74.2 (8.9)	1.8 (1.4)	20.2 (3.3)	0.0 (0.0)	321 (2.0)
Carotid endarterectomy	63,829	45,421 (71.2)	72.2 (9.8)	1.7 (1.4)	20.2 (3.3)	4,462 (7.0)	1,861 (2.9)
Hip fracture repair	235,137	60,169 (25.6)	81.6 (11.4)	1.4 (1.4)	20.1 (3.1)	235,137 (100.0)	3,713 (1.6)

^a Elixhauser comorbidities include congestive heart failure, cardiac arrhythmias, valvular disease, pulmonary circulation disorders, peripheral vascular disorders, hypertension uncomplicated/complicated, paralysis, other neurological disorders, chronic pulmonary disease, diabetes uncomplicated/complicated, hypothyroidism, renal failure, liver disease, peptic ulcer disease excluding bleeding, AIDS/HIV, lymphoma, metastatic cancer, solid tumor without metastasis, rheumatoid arthritis/collagen vascular diseases, coagulopathy, obesity, weight loss, fluid and electrolyte disorders, blood loss anemia, deficiency anemia, alcohol abuse, drug abuse, psychoses, depression.

PCI percutaneous coronary intervention, CABG coronary artery bypass grafting, AAA abdominal aortic aneurysm

year were removed from the dataset. Furthermore, patients <18 years old, who experienced ambulatory care, or with data inaccuracies were excluded. After the washout performed for each procedure separately since 2009, we only selected the first hospitalization of each patient identified as the index stay (except in the case of hip fracture, in which two stays were potentially included if the second stay occurred at least 30 days after the first discharge), using unique, anonymous patient numbers that linked all his/her stays in acute care.

The following patient outcomes were analysed: in-hospital mortality, reoperation, and potentially avoidable hospital readmission. In-hospital mortality and reoperation were defined as death and reoperation, respectively, within a maximum of 30 days post-procedure, whereas potentially avoidable readmission was studied within 30 days of the index stay discharge [28, 29].

Statistical analysis

To classify hospitals based on their volume change over time, we calculated individual hospital volume for each of the ten studied procedures as the total number of patients treated by each hospital within each year. Subsequently, hospitals were divided into three groups based on whether their annual procedure volumes were increasing, decreasing or remaining stable over a 5-year period. We used the random slopes of multilevel Poisson models, taking into account the annual repeated measures of hospital volume for each procedure. These slopes were categorized into three groups using the K-means method to avoid arbitrary determination of thresholds and to account for intra-group variances that could vary [30].

Categorical variables were presented using absolute and relative frequencies, and they were compared between groups using the χ^2 test. Continuous variables were presented using the means and standard deviation, and they were compared using the Mann-Whitney test. The volume change per year for each hospital was estimated from the random slope of multilevel Poisson model.

For each procedure, to determine if mortality was altered in patients admitted to hospitals with significantly increased volume changes over time and if a learning effect existed at the individual hospital level, we used cox regressions, taking into account the clustering effect of patients within hospitals with robust variance estimator (i.e., patients treated and outcomes within a particular hospital tended to be more similar than those in another hospital), the follow-up that varied from one patient to another, and the hospital discharge that represented a censure of outcome [31, 32]. Furthermore, individual hospital learning effect was investigated by examining the interaction between hospital groups and year of procedure. It was measured using a ratio of hazard-ratio (RHR) comparing the hazard-ratios (HR) between hospitals with increasing or decreasing procedure volume and hospitals with unchanged procedure volume. RHR above 1.0 implied a higher mortality over time in hospitals with increasing or decreasing volume than in hospitals with unchanged volume, while RHR below 1.0 reflected a lower mortality.

To adjust mortality for case mix variations, we considered patient (age, gender, Elixhauser list of comorbidities, type and year of procedure, transfer, emergency admission, and median income) and hospital (hospital status, volume of procedures, specialization degree, and attraction rate) characteristics. Restricted cubic splines were used for continuous variables in the adjustment scheme [33].

To test the robustness of our results, after performing mortality analysis for the ten procedures, we repeated this approach across secondary outcomes (reoperation and unplanned hospital readmission) using Fine and Gray's models to consider the, competing risk of death. Model estimates were presented as adjusted hazards ratios with corresponding 95% confidence interval (95% CI). Data manipulation and analyses were performed using SAS version 9.4 (SAS Institute Inc, Cary, NC) and R version 3.2.1 (R Foundation for Statistical Computing, Vienna, Austria) software.

Table 3 - Patient outcomes by procedure

	Mortality N (%)	Reoperation N (%)	Unplanned hospital readmission N (%)
Colectomy	3,271 (3.8)	6,349 (7.4)	6,432 (7.5)
Proctectomy	848 (2.3)	4,760 (12.8)	3,909 (10.5)
Esophagectomy	125 (5.4)	346 (15.1)	360 (15.7)
Gastrectomy	451 (4.6)	898 (9.2)	1,044 (10.7)
Pancreatectomy	283 (4.6)	713 (11.6)	760 (12.4)
PCI	6,741 (2.4)	41,362 (14.8)	34,074 (12.2)
CABG	573 (2.5)	1,762 (7.6)	1,654 (7.1)
AAA repair	204 (1.3)	411 (2.6)	923 (5.8)
Carotid endarterectomy	586 (0.9)	1,908 (3.0)	3,192 (5.0)
Hip fracture repair	9,059 (3.9)	5,827 (2.5)	12,146 (5.2)

PCI percutaneous coronary intervention, CABG coronary artery bypass grafting, AAA abdominal aortic aneurysm

Results

Characteristics of patients and hospitals

Over five years, 759,928 patients were admitted in 694 French hospitals to undergo one of ten procedures related to digestive cancer resection, cardiovascular system intervention, or hip fracture repair (Figure E1). Table 1 shows that not all procedures were performed in all hospitals. The number of hospitals performing each procedure ranged from 45 for CABG to 610 for colectomy. Furthermore, hospital volume varied from one procedure to another, averaging from 29 for gastrectomy to 1,269 for PCI.

Total number of patients ranged from 2,296 for esophagectomy to 280,369 for PCI (Table 2). Regarding adverse event rates between studied procedures (Table 3), patients who underwent esophagectomy had the highest risk of death (5.4%), reoperation (15.1%) and unplanned hospital readmission (15.7%), while those who underwent carotid endarterectomy had the lowest risk of death (0.9%) and unplanned hospital readmission (5.0%) and those who underwent hip fracture repair had the lowest risk of reoperation (2.5%).

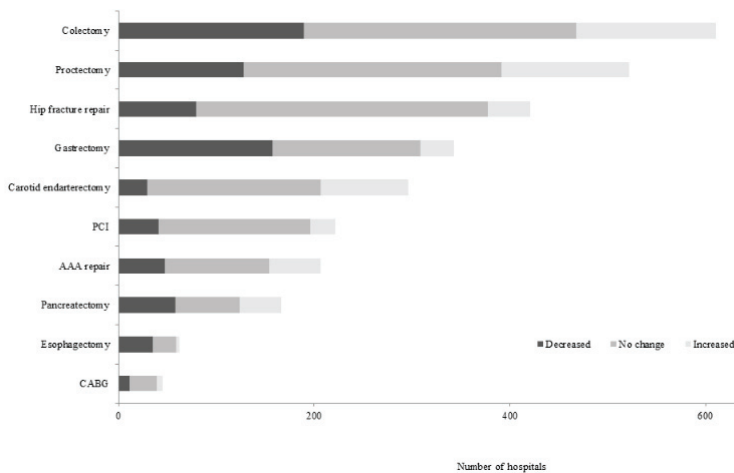


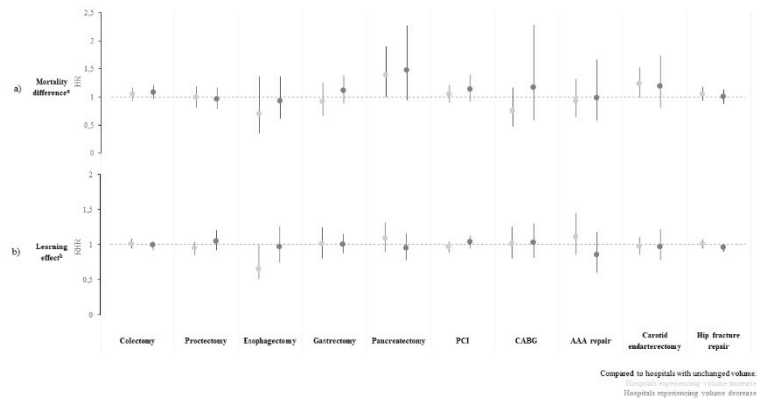
Figure - 1. Number of hospitals by trend in volume of procedures between 2010 and 2014. PCI percutaneous coronary intervention, CABG

coronary artery bypass grafting, AAA abdominal aortic aneurysm.

Relationship between trend in hospital volume and surgical outcomes

Figure 1 shows hospital distribution according to the trend in the volume of procedures. The highest proportion of hospitals with unchanged volume was for hip fracture repair (70.8%), while that of hospitals with an increasing volume and those with decreasing volume was for carotid endarterectomy (30.1%) and for esophagectomy (56.4%), respectively. Volume change per year and hospital characteristics according to procedures and trend in volume of procedures are presented in Table E1.

Figure - 2. Mortality difference and learning effect between hospital groups according to trends in procedures volume from 2010 to 2014.



CABG coronary artery bypass grafting, AAA abdominal aortic aneurysm, PCI percutaneous coronary intervention.
 a Hazard-ratios estimated from Cox model with adjustment regarding patient (age, gender, Elixhauser list of comorbidities, type and year of procedure, transfer, emergency admission, and median income) and hospital characteristics (hospital status, volume of procedures, specialization degree, and attraction rate).
 b Comparison of patient mortality in hospitals with procedure volume increase or decrease over time versus in hospitals with unchanged volume.
 c Analyse to determine if mortality improve or deteriorate over time with an increasing or decreasing volume of procedures

Figure 2a (and Table E2) shows that patient mortality were not different among the hospital groups except for pancreatectomy where the mortality rate was higher in hospitals with increasing volume than in hospitals with unchanged volume (HR 95% CI 1.39 [1.02-1.90], p=0.035). Regarding the other outcomes (Figure E2a and E3a), there was also no difference for most procedures even if some results were inconsistent. Unplanned hospital readmission rate was higher in hospitals with increasing volume for esophagectomy (1.56 [1.08-2.25], p=0.017) and carotid endarterectomy (1.13 [1.01-1.27], p=0.035). Reoperation and unplanned hospital readmission rates were higher in hospitals with decreasing volume (1.31 [1.07-1.61], p=0.010 and 1.34 [1.16-1.56], p<0.001 respectively) for carotid endarterectomy. However, reoperation rate was lower in hospitals with decreasing volume (0.43 [0.22-0.82], p=0.011) for CABG.

Regarding potential individual hospital learning effect (Figure 2b), no association was found between volume trend and mortality over time except for esophagectomy where mortality decreased over time in hospitals with increasing volume (RHR 0.65 [0.51-0.83], $p < 0.001$). Regarding the other outcomes (Figure E2b and E3b), there were also no association except for CABG where reoperation rate was lower over time in hospitals with increasing volume (0.71 [0.53-0.94], $p = 0.018$) and higher in those with decreasing volume (1.53 [1.30-1.81], $p < 0.001$), and for AAA repair where it tended to be lower over time in hospitals with decreasing volume for AAA repair (0.78 [0.64-0.96], $p = 0.018$).

Discussion

We investigated the trends in hospital volume and patient outcomes over time across ten high-risk procedures. We focused on the “practice-makes-perfect” dogma and assumed that hospitals gain expertise from repeating specific surgical care, thereby leading to improved performance.

When comparing outcomes between hospitals that increased, decreased or did not change their volumes, only reoperation and unplanned hospital readmission rates were higher in hospitals with decreasing volume for carotid endarterectomy. Others comparisons revealed either no difference in patient outcomes in hospitals with volume change over time or an inverse relationship to that expected regarding death after pancreatectomy, readmission following esophagectomy or carotid endarterectomy, as well as reoperation after CABG. In the same way, even if individual hospital learning effect may exist regarding death after esophagectomy or reoperation following CABG, patient outcomes for others metrics and procedures did not improve or diminish in hospitals with an increasing or decreasing volume of procedures, respectively. The relationship may even be inverse regarding reoperation after AAA repair. Overall, these results do not support the existence of a robust individual hospital learning effect based on real-world data.

Few prior studies examined the temporal relationship between trends in hospital volume and patient outcomes and they showed contrasting results. These studies involved a limited sample of hospitals and investigated inpatient mortality and/or readmission within 30 days [34–36]. Two studies in trauma centres showed that increasing volume was associated with improving outcomes whereas decreasing volume was associated with worsening outcomes [35, 36]. Another study focused on hip fracture and obtained consistent results with our study, showing that hospitals performing more surgeries over time did not experience outcomes improvement [34]. In our study, to corroborate the “practice-makes-perfect” dogma, we evaluated the consistency of results in a set of distinct procedures.

However, we did not find at the hospital level the learning curve that commonly exists for surgeons or surgical teams experiencing improved performance with higher case volume [23, 24]. Several factors may contribute to the relationship between team familiarity and performance, which could prevent the occurrence of intraoperative event. These factors include improved ability to anticipate the actions of other team members, a heightened willingness to relate with one another, and a greater sense of trust [23]. Surgeons who operate together need to continually interpret each other’s cues, both verbal and nonverbal, and to adjust their actions accordingly to stay coordinated. These learning effects refer to human willingness to progress both individually and collectively during the procedure to prevent intraoperative patient morbidity. Furthermore, we endeavoured revealing a potential learning effect at the hospital level through a gradual optimization of surgical care process with case repetition. Accordingly, we not only considered the operating room activities but also studied the ability of hospital systems to prevent postoperative complications or take actions necessary to mitigate untoward consequences. A hospital’s proficiency in minimizing failure to rescue could be related to a variety of factors, such as available resources and perioperative care organization [37]. The quality of recruitment, preoperative evaluation, anaesthetic management, postoperative nursing, patient follow-up, and an integrated care system with enhanced healthcare workers collaboration are also essential for safe surgery.

This study has some potential limitations. Data were extracted from large hospital databases, which have been initially implemented for billing inpatient stays. Thus, motivation of data coders was possibly influenced by financial stakes rather than epidemiological accuracy [38]. Risk adjustment could only account for factors that can be identified and measured accurately from these data [39]. Although we considered hospital and patient characteristics in the adjustment scheme, this may not be sufficient to ascertain the effect of the differential case mix on surgical outcome between groups since they were not randomly assigned. Moreover, we could not adjust patient outcomes for the volume of procedures and learning curve at the individual surgeon level as no data are available. Consequently, the individual hospital learning effect independent of the surgeon’s ability could not be estimated. We also could not identify whether the increase in volume was due to an increase in surgeon number performing the procedure in the hospital or due to an increase in the number of procedures by the same surgeons. Furthermore, specific complications for each procedure, such as major adverse cardiac and cerebrovascular events for CABG or stroke for carotid surgery, were not monitored. To analyse jointly the results across

various procedures, we opted for generic outcomes. The unclear link between those outcomes and surgical care quality may explain the difficulties in revealing the existence of a potential individual hospital learning effect. This may also be attributable to the fact that many surgical procedures have been performed for a long time and hospitals may have achieved the benefits of learning well before the study period, precluding our ability to evidence practice-makes-perfect. Hence, individual hospital learning effects would be easier to reveal with relatively new procedures, wherein hospitals initiate several changes in the facilities and staff for safe surgery. Finally, a delay in the effect of the change in hospital volume on patient outcomes is possible. Assuming that hospitals would require a sustained volume increase for several years to produce an observable effect on patient outcomes, we may have failed to identify a learning effect because of the limited duration of follow-up of hospital behaviour. Our findings have implications for the efforts aimed at surgical care improvement. We studied a wide range of procedures to investigate the dynamics of volume-outcome relationship within hospitals and showed that a significant increase in volume of procedures over time does not necessarily enhance patient safety. Accordingly, temporal variation in the volume of procedures would not seem a valid surrogate of surgical quality for guiding hospitals' pay-for-performance strategies or providing licensure of surgical practice. Patient outcomes at the individual hospital level would be influenced by factors other than potential learning effect. This may help explain the low mortality [18–20] and the good outcomes [40] in some low-volume hospitals, and possibly some other factors enabled hospitals to achieve excellent outcomes independent of the volume of procedures. For example, higher surgical quality could be attributed to better care organization or the availability of effective medical technologies. Current health policies promote centralizing specific procedures in high-volume referral centres to improve patient outcomes. However, such could result in disparities in patient access to safe surgery, unreasonable travel burdens, and potential delays in operations [41–43]. Hence, in areas where access to referral centres is limited, importing optimal practices may be essential to delivering high-quality surgery. Providing homogeneous surgical quality across the country represents an alternative to centralization, which requires the identification of best systems to achieve excellent outcome and effective strategies to implement these systems from one hospital to another.

Conclusion

The “practice makes perfect” concept considers the more frequent a hospital performs a procedure, the better the outcome of the procedure. In this study, instead of compared outcomes between low- and

high-volume hospitals at some point, we considered the temporal relationship that may exist between volume and outcome under the influence of a dynamic learning effect. Trend in hospital volume over time did not appear influencing consistently patient outcomes across ten high-risk procedures based on real-world data.

Ethics approval and consent to participate: This study was strictly observational and based on anonymous data. Therefore, in accordance with the French ethical directives, it did not require the written informed consent from the participants. In agreement with the pre-cited rules, this study was approved by Institutional Review Board IRB00009118 (Sud-Est II ethical research committee) and the French Data Protection Authority (CNIL DE-2016-028), and it was registered on clinicaltrial.gov (NCT02788331).

Availability of data and materials: The data that support the findings of this study are available from the French Medical Information System maintained by the Technical Agency for Information on Hospital Care. However, restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. To obtain this dataset for an international researcher: Email: demande_base@atih.sante.fr.

Competing interests: The authors declare that they have no competing interests

Funding: This study was supported by a grant from the Programme de Recherche sur la Performance du Système des Soins 2015 of the French Ministry of Health. The funding source had no involvement in the study design; the collection, analysis, and interpretation of data; the writing of the study; or the decision to submit the article for publication. The researchers were independent from the funder.

Authors' contributions: CP and AD had 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. Study concept: CP, SP, JCL, EC, SS, AD; study design: CP, SS, MR, AD; acquisition of data: SP, AD; analysis of data: CP, MR, AD; interpretation of data: CP, SP, JCL, EC, DG, MJC, MR, AD; drafting of the manuscript: CP, SP, AD; critical revision of the manuscript for important intellectual content: CP, SP, JCL, EC, DG, MJC, SS, MR, AD; approved current version of manuscript of publication in its current form: CP, SP, JCL, EC, DG, MJC, SS, MR, AD

References

1. Payet C, Lifante J-C, Carty MJ, Rabilloud M, Duclos A. Methodological Quality of Surgical Mortality Studies Using Large Hospital Databases:

1. A Systematic Review. *Annals of Surgery*. 2017;265:1113–8.
2. Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective Referral to High-Volume Hospitals: Estimating Potentially Avoidable Deaths. *JAMA*. 2000;283:1159.
3. Gandjour A, Bannenberg A, Lauterbach KW. Threshold Volumes Associated With Higher Survival in Health Care: A Systematic Review. *Medical Care*. 2003;41:1129–41.
4. Halm EA, Lee C, Chassin MR. Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature. *Annals of Internal Medicine*. 2002;137:511.
5. Killeen SD, O’Sullivan MJ, Coffey JC, Kirwan WO, Redmond HP. Provider volume and outcomes for oncological procedures. *British Journal of Surgery*. 2005;92:389–402.
6. Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Serv Res*. 1987;22:157–82.
7. Gordon TA, Bowman HM, Tielsch JM, Bass EB, Burleyson GP, Cameron JL. Statewide regionalization of pancreaticoduodenectomy and its effect on in-hospital mortality. *Ann Surg*. 1998;228:71–8.
8. Learn PA, Bach PB. A decade of mortality reductions in major oncologic surgery: the impact of centralization and quality improvement. *Med Care*. 2010;48:1041–9.
9. Birkmeyer JD, Dimick JB, Staiger DO. Operative mortality and procedure volume as predictors of subsequent hospital performance. *Ann Surg*. 2006;243:411–7.
10. Reames BN, Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and operative mortality in the modern era. *Ann Surg*. 2014;260:244–51.
11. Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. *N Engl J Med*. 2011;364:2128–37.
12. de Cruppé W, Ohmann C, Blum K, Geraedts M. Evaluating compulsory minimum volume standards in Germany: how many hospitals were compliant in 2004? *BMC Health Serv Res*. 2007;7:165.
13. Kim W, Wolff S, Ho V. Measuring the Volume-Outcome Relation for Complex Hospital Surgery. *Appl Health Econ Health Policy*. 2016;14:453–64.
14. LaPar DJ, Kron IL, Jones DR, Stukenborg GJ, Kozower BD. Hospital procedure volume should not be used as a measure of surgical quality. *Ann Surg*. 2012;256:606–15.
15. Livingston EH, Cao J. Procedure volume as a predictor of surgical outcomes. *JAMA*. 2010;304:95–7.
16. Duclos A, Lifante J-C. Hospital Administrative Data Should Not be Used to Study Thyroid Surgery Outcomes. *Ann Surg*. 2018;267:e78.
17. Sund R. Modeling the volume-effectiveness relationship in the case of hip fracture treatment in Finland. *BMC Health Serv Res*. 2010;10:238.
18. Horwitz LL, Lin Z, Herrin J, Bernheim S, Drye EE, Krumholz HM, et al. Association of hospital volume with readmission rates: a retrospective cross-sectional study. *BMJ*. 2015;350:h447.
19. Kanhere HA, Trochsler MI, Kanhere MH, Lord AN, Maddern GJ. Pancreaticoduodenectomy: outcomes in a low-volume, specialised Hepato Pancreato Biliary unit. *World J Surg*. 2014;38:1484–90.
20. Goldschlager T, Selvanathan S, Walker DG. Can a “novice” do aneurysm surgery? Surgical outcomes in a low-volume, non-specialised neurosurgical unit. *J Clin Neurosci*. 2007;14:1055–61.
21. Jha AK. Back to the Future: Volume as a Quality Metric. *JAMA*. 2015;314:214–5.
22. Gonzalez AA, Dimick JB, Birkmeyer JD, Ghaferi AA. Understanding the volume-outcome effect in cardiovascular surgery: the role of failure to rescue. *JAMA Surg*. 2014;149:119–23.
23. Xu R, Carty MJ, Orgill DP, Lipsitz SR, Duclos A. The teaming curve: a longitudinal study of the influence of surgical team familiarity on operative time. *Ann Surg*. 2013;258:953–7.
24. Maruthappu M, El-Harasis MA, Nagendran M, Orgill DP, McCulloch P, Duclos A, et al. Systematic review of methodological quality of individual performance measurement in surgery. *Br J Surg*. 2014;101:1491–8; discussion 1498.
25. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part I: The relation between hospital volume and outcomes for selected diagnostic categories. *Med Care*. 1984;22:98–114.
26. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part II: The relation between volume and outcomes and other hospital characteristics. *Med Care*. 1984;22:115–25.
27. 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.
28. Halfon P, Egli Y, van Melle G, Chevalier J, Wasserfallen JB, Burnand B. Measuring potentially avoidable hospital readmissions. *J Clin Epidemiol*. 2002;55:573–87.
29. Halfon P, Egli Y, Prêtre-Rohrbach I, Meylan D, Marazzi A, Burnand B. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. *Med Care*. 2006;44:972–81.
30. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*. 1979;28:100.
31. Varadhan R, Weiss CO, Segal JB, Wu AW, Scharfstein D, Boyd C. Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. *Med Care*. 2010;48 6 Suppl:S96-105.

32. Glidden DV, Vittinghoff E. Modelling clustered survival data from multicentre clinical trials. *Stat Med.* 2004;23:369–88.
33. Marrie RA, Dawson NV, Garland A. Quantile regression and restricted cubic splines are useful for exploring relationships between continuous variables. *J Clin Epidemiol.* 2009;62:511-517.e1.
34. Hamilton BH, Ho V. Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. *Med Care.* 1998;36:892–903.
35. Brown JB, Rosengart MR, Kahn JM, Mohan D, Zuckerbraun BS, Billiar TR, et al. Impact of Volume Change Over Time on Trauma Mortality in the United States. *Ann Surg.* 2017;266:173–8.
36. Marcin JP, Romano PS. Impact of between-hospital volume and within-hospital volume on mortality and readmission rates for trauma patients in California. *Crit Care Med.* 2004;32:1477–83.
37. Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and failure to rescue with high-risk surgery. *Med Care.* 2011;49:1076–81.
38. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med.* 1997;127 8 Pt 2:666–74.
39. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet.* 2004;363:1147–54.
40. Maa J, Gosnell JE, Gibbs VC, Harris HW. Exporting excellence for Whipple resection to refine the Leapfrog Initiative. *J Surg Res.* 2007;138:189–97.
41. Stitzenberg KB, Sigurdson ER, Egleston BL, Starkey RB, Meropol NJ. Centralization of cancer surgery: implications for patient access to optimal care. *J Clin Oncol.* 2009;27:4671–8.
42. Finlayson SR, Birkmeyer JD, Tosteson AN, Nease RF. Patient preferences for location of care: implications for regionalization. *Med Care.* 1999;37:204–9.
43. Birkmeyer JD, Siewers AE, Marth NJ, Goodman DC. Regionalization of high-risk surgery and implications for patient travel times. *JAMA.* 2003;290:2703–8.

Dans ce travail, nous avons étudié les évolutions de volume des hôpitaux au cours du temps en fonction des résultats des patients pour dix interventions à haut risque. Nous nous sommes concentrés sur le concept du "plus on fait, mieux on fait" et nous avons supposé que les hôpitaux acquièrent une expertise en répétant des soins chirurgicaux spécifiques, conduisant ainsi à une amélioration des résultats chirurgicaux des patients. Néanmoins, les résultats obtenus n'étaient pas en faveur d'un effet d'apprentissage individuel au niveau de l'hôpital puisque de manière générale, les résultats chirurgicaux ne dépendaient pas de l'évolution du volume des hôpitaux au cours du temps.

Dans la suite de ce travail, nous nous sommes concentrés sur une deuxième thématique de recherche mise en évidence dans notre revue de la littérature : la comparaison de l'efficacité réelle de différentes interventions. Cela a fait l'objet des deux travaux suivants.

C. Reprise chirurgicale après une chirurgie bariatrique (article 3)

L'anneau gastrique, la *sleeve* et le *bypass* sont les trois principales chirurgies bariatriques réalisées en France (Figure 14) [136]. L'anneau gastrique est un anneau réglable en silicone posé autour de la partie supérieure de l'estomac afin de ralentir le passage des aliments à ce niveau. La *sleeve gastrectomy* repose sur la résection d'une partie de l'estomac afin que celui-ci forme un tube. Enfin, le *gastric bypass* consiste à court-circuiter une grande partie de l'estomac afin de le réduire à une petite poche gastrique.

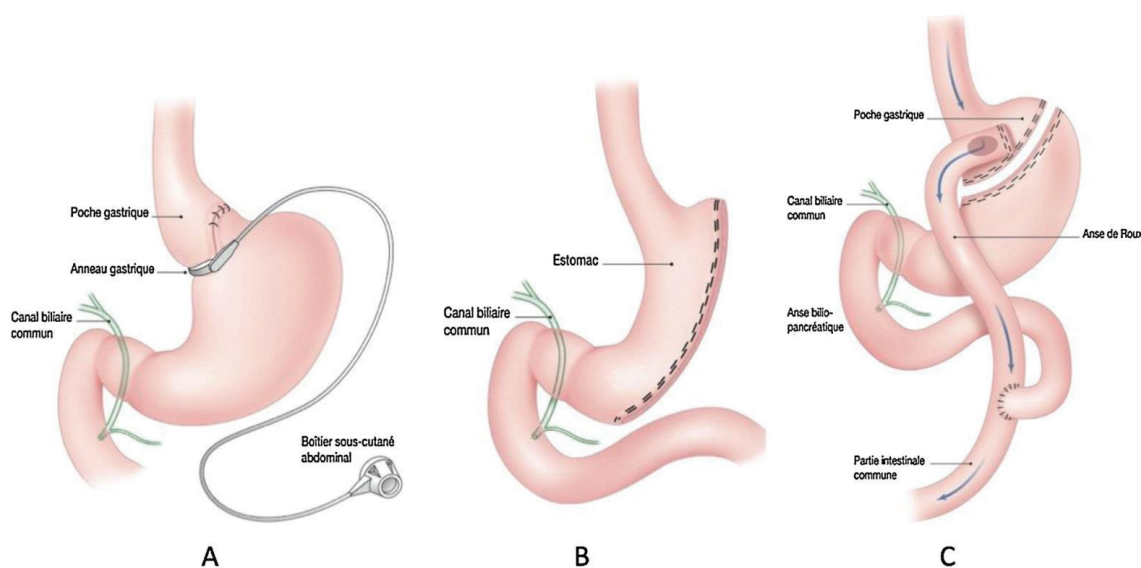


Figure 14 – Représentation schématique des trois interventions les plus réalisées en France. A : l'anneau gastrique, B : La *sleeve gastrectomy*, C : le *gastric bypass* [136].

Les complications postopératoires de ces trois chirurgies étant principalement évaluées au cours du premier mois postopératoire [137,138], l'objectif de cette étude était de comparer entre ces trois chirurgies, le risque de reprise chirurgicale et sa sévérité, dans les six mois postopératoires.

Nous avons donc sélectionné dans la base PMSI nationale MCO, les patients opérés entre 2013 et 2015 d'une sleeve, d'un bypass ou d'un anneau gastrique. Au sein de chaque hôpital, les patients opérés d'une sleeve ont été appariés aux patients opérés d'un bypass (cohorte A) ou d'un anneau (cohorte B), par des scores de propension considérant les caractéristiques des patients. Les reprises chirurgicales ont été classées selon leur sévérité (faible, modérée,

élevée), à l'aveugle, par trois chirurgiens sur la base des actes CCAM et de la classification du Dindo-Clavien. Le risque de survenue d'une reprise chirurgicale et sa sévérité ont été comparés dans les deux cohortes avec des modèles de survie tenant compte du risque compétitif de décès, dans les six mois postopératoires.

Article 3: Brunaud L, Payet C, Polazzi S, Bihain F, Quilliot D, Lifante JC, Duclos A. Reoperations incidence and severity within 6 months after bariatric surgery : a propensity-matched study from nationwide data. *Obesity Surgery* (Accepté).

Reoperations incidence and severity within 6 months after bariatric surgery: a propensity-matched study from nationwide data

Laurent BRUNAUD, MD ^{1,2}, Cecile PAYET, MD ³, Stephanie POLAZZI, MD ³, Florence BIHAIN, MD ¹, Didier QUILLIOT, MD ⁴, Jean-Christophe LIFANTE, MD ⁵, Antoine DUCLOS, MD ³

1: Department of Gastrointestinal, Metabolic, and Surgical Oncology (DCVMC). Multidisciplinary unit of obesity surgery (UMCO). University of Lorraine, CHRU Nancy, Brabois Hospital, Vandoeuvre-les-Nancy, France

2: INSERM U1256, « Nutrition, Genetics, Environmental Risks », Faculty of Medicine, University of Lorraine, Nancy, France

3: Department of Medical Information Evaluation and Research, Lyon University Hospital, Lyon, France Health Services and Performance Research Lab (EA 7425 HESPER), Lyon 1 Claude Bernard University, Lyon, France

4: Department of Endocrinology, Diabetology and Nutrition, University of Lorraine, CHRU Nancy, Brabois Hospital, Nancy, France

5: Department of General and Endocrine Surgery, Lyon University Hospital, Lyon, France

Corresponding author: Laurent BRUNAUD, Département de Chirurgie viscérale, Métabolique et Cancérologique (DCVMC), CHRU Nancy, Hopital de Brabois (7ème étage), Université de Lorraine, 11 allée du morvan, 54511 Vandoeuvre-les-Nancy, FRANCE, Tel: (33)-383153120; Fax: (33)-383153121

Abstract

Background: Data about incidence and severity of reoperations up to 6 months after bariatric surgery are currently limited. The aim of this cohort study was to evaluate the incidence and severity of reoperations after initial bariatric surgical procedures and to compare this between the 3 most frequent current surgical procedures (sleeve, gastric bypass, gastric banding).

Study design: Nationwide observational cohort study using data from French Hospital Information System (2013-2015) to evaluate incidence and severity of reoperations within 6 months after bariatric surgery. Hazard ratios (HR) of longitudinal comparison between historical propensity-matched cohorts were estimated from a Fine and Gray's model using competing risk of death.

Results: Cumulative reoperation rates increased from postoperative day-30 to day-180. Consequently, 31.1 to 90.0% of procedures would have been missed if the reoperation rate was based solely on a 30-day follow-up. Reoperation rate at 6 months was significantly higher after gastric bypass than after sleeve (HR 0.64; IC 95% [0.53-0.77]) and corresponded to moderate-risk reoperations (HR 0.65; IC 95% [0.53-0.78]). Reoperation rate at 6 months was significantly higher after gastric banding than after sleeve (HR 0.08; IC 95% [0.07-0.09]) and corresponded to moderate-risk reoperations (HR 0.08; IC 95% [0.07-0.10]).

Conclusion: Cumulative incidence of reoperations increased from 30-day to 6 months after sleeve, gastric bypass, or gastric banding and corresponded to moderate-risk surgical procedures. Consequently, 30-day reoperation rate should no longer be considered when evaluating complications and surgical performance after bariatric surgery.

Introduction

Bariatric surgery is currently the optimal management to obtain long-term weight loss with subsequent positive impacts in obese patients (1, 2). Complication rates after bariatric surgical procedures are low and bariatric surgery safety has improved over time with the optimization of surgical techniques and perioperative management (3-6). However, given the large number of procedures performed worldwide, serious complications occurrence negatively impact patient outcomes and hospital costs (3). The risk factors of postoperative surgical complications are usually categorized into factors related to the patient, to the surgical procedure itself, and to the medical center including multidisciplinary expertise in bariatric surgery and the intensive care unit availability (4, 7-12). The relationship between the volume of procedures performed per institution and patient outcomes has

also been investigated as a proxy of surgical quality (4, 10, 12). However, the identification of metrics to assess accurately medical centers performance regarding bariatric surgery is still debated (13). Postoperative mortality has been rejected because this event is too rare to allow meaningful peer-based comparisons (4, 9, 11, 13, 14). Severe complication, defined as the need for patient reoperation following initial procedure has been suggested as a superior indicator of safety in bariatric surgery (8, 13, 15). This simple surrogate marker is clinically relevant to describe the quality of the postoperative period. Although well described within the first month following surgery, data about reoperations incidence up to 6 months after initial bariatric procedures are currently very limited. In addition, the severity of those reoperations has not yet been identified at a national level. This nationwide study aimed to compare the incidence and severity of reoperations

at 6 months after initial surgery between the 3 most frequent bariatric procedures (sleeve, gastric bypass, gastric banding).

Methods

Study design and data source

We performed a three-year longitudinal comparison of historical propensity-matched cohorts from nationwide data. Two comparative cohorts of patients who underwent initial bariatric surgery were designed by matching patient pairs inside hospitals: i) sleeve versus gastric bypass and ii) sleeve versus gastric banding (16). In each of these two matched cohorts, incidence and severity of reoperations were compared at different postoperative intervals within 6 months following the index procedure.

This study used the French Medical Information System (*Programme de Médicalisation des Systèmes d'Information [PMSI], source ATIH*). The PMSI is a nationwide database routinely implemented for the purpose of care reimbursement and updated weekly with data from all hospitals in France.

The database is prospectively collected and relies on a coding system with strict variable definitions. A subset of records is audited on a regular basis to avoid coding errors. As a consequence of suitable accuracy and exhaustive data collection, no patient was assumed to be lost to follow-up during the study period. Inpatient stays were recorded as standard discharge abstracts containing compulsory information about the patient and primary/secondary diagnoses, using the International Classification of Diseases 10th revision, as well as detailed procedural codes associated with the care provided. From the PMSI database, we extracted patients' demographics, comorbidities according to the Elixhauser algorithm with acceptable validity (17), the body mass index (BMI ≥ 30 to < 40 kg/m², BMI ≥ 40 to < 50 kg/m², BMI ≥ 50 kg/m²), date and type of surgery (17, 18). Every patient had a unique, anonymous identification number that linked all his/her hospital stays in acute care over time. Because this study was strictly observational and based on anonymous data, the written informed consent from the participants or the authorization from an ethical committee for dealing with human issues was not required in accordance to the French laws. This study was declared to the National Data Protection Commission (Commission Nationale Informatique et Libertés – CNIL – MR 4312040918) for potential data issues through the "Reference Methodology 005" before performing any analysis.

Study population and outcomes

All adults over 18 years old who underwent one of the following three bariatric surgical procedures from January 1, 2013 to December 31, 2015 were included: gastric bypass, gastric banding, and sleeve gastrectomy. All patients with at least one previous bariatric surgical procedure(s) (sleeve gastrectomy,

gastric banding, gastric bypass, vertical gastroplasty, biliopancreatic diversion, intragastric balloon, or gastric electrical stimulation) during a prior five-year washout period were excluded. Consequently, this study intended to include only patients hospitalized for the first time for initial bariatric surgery to avoid potential bias due to bariatric revisional cases (19).

The primary outcome was the incidence of reoperation at different postoperative intervals (1-7 days, 8-30 days, 31-60 days, and 61-180 days) after the initial bariatric surgical procedure. A reoperation was defined as the need for any surgical procedure in relation with the initial bariatric surgery, and considered as a severe postoperative complication according to Clavien et al. classification ($> 3b$)(15). As a secondary outcome analysis, we discriminated reoperation severity in three classes: high-risk, moderate-risk, and low-risk reoperations. To determine the causality between the initial surgical procedure and reoperation, as well as the ranking of severity, exhaustive screening of procedural codes was performed in a blind-fashion by 4 distinct surgeons (Annex 1). First, all reoperations (n=2,254) recorded in the PMSI database within 180 days after an initial bariatric surgical procedure were classified by 3 different surgeons separately. For each performed reoperation, the potential link with initial bariatric surgical procedure was determined. When there was an agreement between all 3 surgeons (n=2,226), their judgment was considered as definitive, meanwhile a fourth independent surgeon decided the final attribution in case of absence of consensus among those 3 first surgeons (n=28). Then, all reoperations considered as imputable to the initial surgical procedure (n=211) were evaluated for their severity. Reoperations severity was a priori classified as high-risk, moderate-risk, and low-risk surgical procedures. As a ranking system evaluating the severity of reoperations after bariatric surgery is currently not available, evaluation of reoperation severity in this study was inspired from previous published intraoperative adverse events and postoperative surgical complications classifications (15, 20). As previously, the three surgeons agreed for most procedures (n=196), while the fourth surgeon only rarely had to decide for the final attribution (n=15). To confirm the clinical relevance of our classification procedures regarding the reoperation severity, we computed the corresponding mortality rates across each of the 3 classes grading system in Annex 2. Procedures classifications regarding causality and severity of reoperations were performed before starting statistical analysis.

Statistical analysis

Comparative safety analysis has been defined as a rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients (21). To

control for the nonrandom assignment of patients between procedures, two matched-pairs cohorts of patients were formed between sleeve and bypass and between sleeve and banding. In each of the two matched cohorts, propensity scores were estimated as the predicted probability of a patient undergoing sleeve using a logistic regression model including the following covariates: sex, age, BMI, household income, procedure year, and a selection of comorbidities (i.e. congestive heart failure, cardiac arrhythmia, valvular disease, pulmonary circulation disorder, peripheral vascular disease, hypertension, chronic pulmonary disease, diabetes, renal disease, liver disease, peptic ulcer disease, coagulopathy, fluid and electrolyte disorders)(22, 23). Restricted cubic splines with 3 knots were used to model the effect of continuous covariates (24). We then matched patients with the closest propensity score inside hospital to control for confounders at hospital level, using a greedy 1:1 algorithm without replacement and requiring that the logit of the propensity score of a patient who underwent sleeve and one who underwent the other procedure be within 0.20 standard deviations of one another (25). Standardized differences were used to assess the degree of balance between the matched groups for baseline characteristics (22). An absolute standardized difference of ≤ 0.10 was chosen to indicate a negligible difference in the mean or prevalence of a variable between groups. Balance for continuous variables was also assessed with graphical methods (side-by-side boxplots, empirical cumulative distribution functions, empirical QQ-plots) to compare the distributions across groups. In each of the two matched cohorts, reoperation and severity of reoperation were assessed as time-to-event variables. Hazard Ratios (HR) were estimated from a Fine and Gray’s model using competing risk of death. As proportionality of hazards could not be assumed, a piecewise model was used to obtain HR during four periods of time (1-7, 8-30, 31-60, and 61-180 days). Matched-pair design was taken into account with robust variance estimator. Data manipulation and analyses were performed using SAS (version 9.4; SAS Institute Inc., Cary, NC) software. Categorical variables were presented using absolute and relative frequencies and continuous variables were presented using means and standard deviations. Estimates were accompanied with the corresponding 95% CI and p-values of less than 0.05 were considered to indicate statistical significance.

Results

Comparative safety of sleeve versus gastric bypass

After applying the selection criteria and retaining only hospitals performing both procedures, this study dataset included 332 hospitals having performed 62,747 sleeves and 31,209 gastric

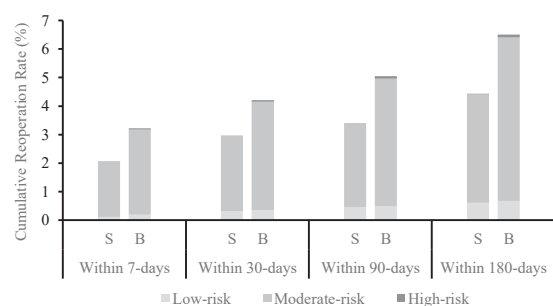
bypasses during the study period (Annex 3). Overall, 17 126 patients with sleeve were matched and compared with 17.126 patients with gastric bypass in 304 hospitals. Patients’ characteristics at baseline were similar between the two groups (Table 1).

Table 1 – Baseline characteristics of patients in matched population of patients with sleeve and bypass.

	Sleeve N=17,126 (%)	Gastric bypass N=17,126 (%)	Standardized difference
Male sex	3,343 (19.5%)	3,019 (17.6%)	0.05
Age, y mean (SD)	41.5 (11.8)	42.0 (11.5)	-0.04
BMI			0.04
30-39	4,558 (26.6%)	4,783 (27.9%)	
40-49	10,405 (60.8%)	10,472 (61.1%)	
≥ 50	2,163 (12.6%)	1,871 (10.9%)	
Income, € mean (SD)	20,203.9 (3,172.4)	20,155.5 (3,235.8)	0.02
Year			0.07
2013	5,034 (29.4%)	5,433 (31.7%)	
2014	5,717 (33.4%)	5,733 (33.5%)	
2015	6,375 (37.2%)	5,960 (34.8%)	
Congestive heart failure	99 (0.6%)	95 (0.6%)	0.00
Cardiac arrhythmias	491 (2.9%)	471 (2.8%)	0.01
Valvular disease	44 (0.3%)	45 (0.3%)	-0.00
Pulmonary circulation disorders	21 (0.1%)	29 (0.2%)	-0.01
Peripheral vascular disease	37 (0.2%)	37 (0.2%)	0.00
Hypertension	4,018 (23.5%)	4,096 (23.9%)	-0.01
Chronic pulmonary disease	722 (4.2%)	773 (4.5%)	-0.01
Diabete	2,369 (13.8%)	2,653 (15.5%)	-0.05
Renal disease	77 (0.4%)	62 (0.4%)	0.01
Mild liver disease	826 (4.8%)	954 (5.6%)	-0.03
Peptic ulcer disease	15 (0.1%)	11 (0.1%)	0.01
Coagulopathy	75 (0.4%)	67 (0.4%)	0.01
Fluid and electrolyte disorders	289 (1.7%)	246 (1.4%)	0.02

Reoperation rates increased after sleeve and gastric bypass from 2.0% at 7 days to 4.1% at 180 days and from 3.1% at 7 days to 6.1% at 180 days, respectively (Figure 1). After sleeve, 30-day reoperation rate was 2.8% and would have underestimated one third (31.1%) of the need for reoperation in comparison with 180-day reoperation rate. In a similar manner, 30-day reoperation rate after gastric bypass was 4.0% and would have missed about the same amount (33.9%) of reoperations in comparison with 180-day reoperation rate (Figure 1).

Figure 1 – Cumulative reoperations rates in patients after sleeve (S) and bypass (B) at 7, 30, 90 and 180 days, postoperatively.



When comparing the matched-pairs cohort, reoperation rates were significantly lower after sleeve compared to gastric bypass at one week (HR 0.64; IC 95% [0.56-0.73]), at 2 months (HR 0.43; IC

95% [0.28-0.66]), and at 6 months (HR 0.64; IC 95% [0.53-0.77]), postoperatively (Table 2). Those reoperations observed after sleeve gastrectomy and gastric bypass were mostly moderate-risk (93.9% and 94.4%, respectively). Details about reoperations are provided in Annex 1. Low-risk reoperation rates were similar between sleeve versus bypass group patients, while high-risk reoperations were very rare in both groups (Table 2).

Table 2 – Incidence and severity of reoperations within 180 days after sleeve versus gastric bypass in matched population of patients.

	Sleeve N=17,126 (%)	Gastric bypass N=17,126 (%)	HR [IC 95%]	P
Incidence of reoperations				
1-7 days	344 (2.0%)	536 (3.1%)	0.64 [0.56-0.73]	<0.001
8-30 days	134 (0.8%)	151 (0.9%)	0.88 [0.69-1.11]	0.266
31-60 days	29 (0.2%)	67 (0.4%)	0.43 [0.28-0.66]	<0.001
61-180 days	187 (1.1%)	286 (1.7%)	0.64 [0.53-0.77]	<0.001
High-risk reoperations				
1-7 days	1 (0.0%)	8 (0.0%)	-	
8-30 days	0 (0.0%)	2 (0.0%)	-	
31-60 days	2 (0.0%)	1 (0.0%)	-	
61-180 days	1 (0.0%)	6 (0.0%)	-	
Moderate-risk reoperations				
1-7 days	333 (1.9%)	511 (3.0%)	0.65 [0.57-0.74]	<0.001
8-30 days	119 (0.7%)	140 (0.8%)	0.84 [0.66-1.07]	0.162
31-60 days	23 (0.2%)	62 (0.4%)	0.37 [0.23-0.59]	<0.001
61-180 days	177 (1.0%)	269 (1.6%)	0.65 [0.53-0.78]	<0.001
Low-risk reoperations				
1-7 days	19 (0.1%)	33 (0.2%)	0.58 [0.33-1.01]	0.055
8-30 days	35 (0.2%)	27 (0.2%)	1.30 [0.78-2.14]	0.312
31-60 days	15 (0.1%)	17 (0.1%)	0.88 [0.44-1.77]	0.723
61-180 days	35 (0.2%)	38 (0.2%)	0.92 [0.58-1.46]	0.724

Comparative safety of sleeve versus gastric banding

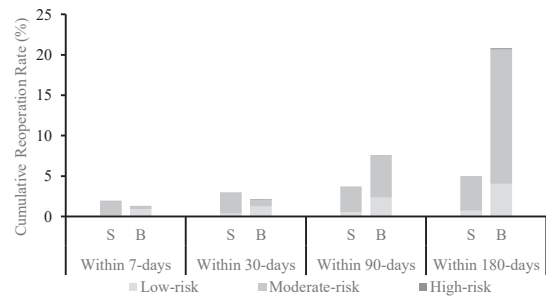
After applying the selection criteria and retaining only hospitals performing both procedures, this study dataset included 417 hospitals having performed 71,053 sleeve and 19,337 gastric banding during the study period (Annex 4).

Table 3 – Baseline characteristics of patients in matched cohorts of patients with gastric banding and sleeve.

	Sleeve N=12,603 (%)	Gastric banding N=12,603 (%)	Standardized difference
Male sex	1,887 (15.0%)	1,915 (15.2%)	-0.01
Age, y mean (SD)	39.4 (12.1)	39.9 (12.8)	-0.04
BMI			0.05
30-39	6,105 (48.4%)	6,170 (49.0%)	
40-49	5,785 (45.9%)	5,757 (45.7%)	
≥ 50	713 (5.7%)	676 (5.4%)	
Income, € mean (SD)	19,972.7 (3,329.8)	19,947.9 (3,296.0)	0.01
Year			0.07
2013	4,455 (35.3%)	4,626 (36.7%)	
2014	4,327 (34.3%)	4,353 (34.5%)	
2015	3,821 (30.3%)	3,624 (28.8%)	
Congestive heart failure	29 (0.2%)	34 (0.3%)	-0.01
Cardiac arrhythmias	107 (0.8%)	106 (0.8%)	0.00
Valvular disease	16 (0.1%)	23 (0.2%)	-0.01
Pulmonary circulation disorders	5 (0.0%)	8 (0.1%)	-0.01
Peripheral vascular disease	19 (0.2%)	15 (0.1%)	0.01
Hypertension	1,684 (13.4%)	1,732 (13.7%)	-0.01
Chronic pulmonary disease	389 (3.1%)	447 (3.5%)	-0.03
Diabete	706 (5.6%)	764 (6.1%)	-0.02
Renal disease	34 (0.3%)	29 (0.2%)	0.01
Mild liver disease	215 (1.7%)	257 (2.0%)	-0.02
Peptic ulcer disease	10 (0.1%)	6 (0.0%)	0.01
Coagulopathy	20 (0.2%)	22 (0.2%)	-0.00
Fluid and electrolyte disorders	35 (0.3%)	37 (0.3%)	-0.00

Overall, 12,603 patients with sleeve were matched and compared with 12,603 patients with gastric banding in 378 hospitals. Patients' characteristics at baseline were similar between the two group patients (Table 3). Reoperation rates increased after sleeve and gastric banding from 1.9% at 7 days to 4.5% at 180 days and from 1.2% at 7 days to 19.3% at 180 days, respectively (Figure 2). After sleeve, 30-day reoperation rate was 2.8% and would have underestimated 37.9% of the need for reoperation in comparison with 180-day reoperation rate. Most importantly, 30-day reoperation rate after gastric banding was 1.9% and would have missed 90% of reoperations in comparison with 180-day reoperation rate (Figure 2).

Figure 2 – Cumulative reoperations rates in patients after sleeve and gastric banding at 7, 30, 90 and 180 days postoperatively.



When comparing the matched-pairs cohort, reoperation rate was significantly higher after sleeve than after gastric banding at one week postoperatively (HR 1.58; IC 95% [1.29-1.94])(Table 4). These reoperations at 7 days were mostly moderate-risk reoperations after sleeve (93.3%), and lowrisk reoperations after gastric banding (76.3%). However, reoperation rate after sleeve became significantly lower than after gastric banding at 2 months (HR 0.29; IC 95% [0.21-0.39]) and at 6 months (HR 0.08; IC 95% [0.07-0.09]), postoperatively (Table 4). These reoperations after gastric banding were mostly moderate-risk reoperations (75.1% and 92.0%, respectively) (Table 4). Details about reoperations were provided in Annex 1. High-risk reoperations were very rare in both groups (Table 4).

Discussion

Principal findings

We used real-world data from a nationwide database to compare incidence and severity of reoperations between the most common bariatric surgical procedures using 2 propensity score-matched cohorts of patients. Our findings showed that reoperations occurrence continues to grow over time from 30-day to up to 6 months after all procedure types (gastric banding, gastric bypass, and sleeve). Consequently, a large amount of procedures would

have been underestimated if the reoperation rate was based solely on a 30-day follow-up. Reoperations severity was clinically relevant because most of reoperations were moderate-risk surgical procedures. Reoperation rate at 6 months was significantly lower after sleeve gastrectomy compared to gastric bypass or banding.

Table 4 – Comparison of incidence and severity of reoperations within 180 days after gastric banding versus sleeve in matched cohorts of patients.

	Sleeve N=12,603 (%)	Gastric banding N=12,603 (%)	HR [IC 95%]	P
Incidence of reoperations				
1-7 days	240 (1.9%)	152 (1.2%)	1.58 [1.29-1.94]	<0.001
8-30 days	114 (0.9%)	91 (0.7%)	1.26 [0.96-1.67]	0.094
31-60 days	48 (0.4%)	169 (1.3%)	0.29 [0.21-0.39]	<0.001
61-180 days	168 (1.3%)	2021 (16.0%)	0.08 [0.07-0.09]	<0.001
High-risk reoperations				
1-7 days	1 (0.0%)	2 (0.0%)	-	-
8-30 days	1 (0.0%)	5 (0.0%)	-	-
31-60 days	1 (0.0%)	0 (0.0%)	-	-
61-180 days	1 (0.0%)	14 (0.1%)	-	-
Moderate-risk reoperations				
1-7 days	224 (1.8%)	45 (0.4%)	5.01 [3.63-6.89]	<0.001
8-30 days	102 (0.8%)	63 (0.5%)	1.65 [1.20-2.26]	0.002
31-60 days	41 (0.3%)	127 (1.0%)	0.33 [0.23-0.47]	<0.001
61-180 days	164 (1.3%)	1860 (14.8%)	0.08 [0.07-0.10]	<0.001
Low-risk reoperations				
1-7 days	17 (0.1%)	116 (0.9%)	0.15 [0.09-0.24]	<0.001
8-30 days	27 (0.2%)	36 (0.3%)	0.74 [0.45-1.22]	0.245
31-60 days	20 (0.2%)	61 (0.5%)	0.32 [0.20-0.54]	<0.001
61-180 days	28 (0.2%)	297 (2.4%)	0.09 [0.06-0.14]	<0.001

Practical implications

The strength of the French nationwide database is the number of eligible patients and the exhaustive data available from every hospital in the country (5, 26, 27). In this study, we decided to focus on severe postoperative complications (Clavien > 3b) using a 6-month follow-up because this criterion has never been used in a national report analysis whereas it had been suggested as a superior approach for the evaluation of bariatric surgery quality (8, 13, 14). This surrogate marker is simple to define without room for ambiguity, and is captured in most medico-administrative database.

Bariatric surgery is a high-volume procedure with, although rare, high potential for serious postoperative complications (28). Consequently, a large number of patients in each group are necessary to gain enough power to evaluate the prevalence of complications between two surgical treatment arms in comparative studies (29). This explains the need for a populationbased analysis including a larger number of patients and could explain why short-term results of sleeve gastrectomy and Roux-en-Y gastric bypass regarding complications were not different at 6 months in prior randomized trials (29, 30). Numerous previous mono- or multi-center studies and one national report have also been published on this subject but postoperative follow-up in these studies were either at 30 days or shorter than 90 days (31-34). Reoperation rates after gastric bypass and sleeve for example were reported to be ranging from

0.67% to 3.41%, likely underestimating postoperative outcomes on a longer period and the actual need for reoperations (31-34).

This national report study provided evidence showing that 30-day, or even 90-day morbidity rates clearly underestimate the actual postoperative course after gastric banding, gastric bypass, and sleeve gastrectomy. Colquitt et al. concluded in the Cochrane Review that adverse event and reoperation rates have been generally poorly reported across all available studies (35).

Similarly, we believe that severe complications needing reoperation after bariatric surgery are not negligible, remaining a concern for patients and bariatric surgeons, and should be more specifically evaluated.

Data provided in this study support the extension of postoperative follow-up beyond the usual 30 days to capture more comprehensively the need for postoperative reoperations. A period of postoperative follow-up longer than 30 days has previously been proposed for other surgical procedures for similar reasons (e.g. esophagectomy or pancreatectomy)(36).

Comparison with other studies

This study supports previous reports suggesting that fewer early severe complications are associated with sleeve gastrectomy compared with gastric bypass procedure (34, 37). However, we acknowledge that this point remains controversial because some other papers concluded that early postoperative outcomes were similar between gastric bypass and sleeve gastrectomy (35). Prior major randomized controlled trials comparing gastric bypass versus sleeve showed similar reoperation rates between the two group patients although those trials were not specifically designed to evaluate major postoperative complications (29, 30, 38). We believe that current findings in this study may also correspond to more recent years data analysis and reflect the positive evolution over time of surgical techniques and perioperative management.

Gastric banding has been very popular during the last two decades because it was a simple surgical technique, easy to perform laparoscopically, and considered to be associated with low postoperative complications. Recent evaluations concluded that gastric banding was associated with worse long term weight loss and higher rate of reoperations (2, 26). This study confirmed that reoperations were more frequent after gastric banding. More precisely, it showed that reoperations within 7 days were less frequent after gastric banding than after sleeve gastrectomy. However, reoperations became significantly more frequent after gastric banding than after sleeve at 60 days and 180 days. This study provided data showing that most of those reoperations were moderate-risk surgical procedures. Taken together, our findings indicate

that gastric banding is associated with higher reoperation rates and support prior studies raising concerns about its safety, effectiveness, and value of this surgical technique (26, 27, 39).

Strengths and limitations

In this study, two propensity score matched-pairs cohorts of patients were formed to control for the nonrandom assignment of patients between procedures and for confounders at hospital level (16, 28). Among variables available within the PMSI database, we chose those with the most clinical relevance to discriminate the reoperation risk of populations but also accounting for those with a sufficient degree of validity. Furthermore, we provided information about the causality and the severity of reoperations, while prior studies generally did not identify whether reoperation is related to the initial bariatric surgical procedure or originates from a new issue that developed after inpatient discharge (40). All reoperations were blindly classified before analysis and only reoperations unanimously considered as imputable to initial bariatric surgical procedures were included. The estimation of severity of reoperations was provided prior to the statistical analysis that was confirmed posteriorly by the correspondence with postoperative mortality.

This study has several limitations. We designed a propensity score matched study using large hospital database. The risk of bias in such study is to omit some potential confounders that can alter the comparability of populations and therefore threaten the validity of conclusions. We are aware that the PMSI variables may lack of granularity to account for certain confounders accurately. For example, PMSI coding for gastric bypass during this study period corresponded to both Roux-en-Y and one anastomosis omega gastric bypass. In addition, racial and cultural differences among patients should be considered when designing strategies to optimize outcomes with bariatric surgery. However, this discrimination is forbidden by laws in a national database in France. Another weakness of medicoadministrative is the miscoding of diagnoses during hospital stays that can underestimate patients' comorbidities (41). Because this issue is influenced by a strong coding variability between health care providers and across years, we matched pairs of patients who underwent either sleeve, gastric bypass or gastric banding inside the same hospital and over the same period. Hence, we believe there is no a priori reason that miscoding would be more prominent in one cohort than another and would alter their comparability.

Conclusions

This study showed that cumulative incidence of reoperations increases continuously from 30-day to up to 6 months after sleeve, gastric bypass, or gastric

banding and corresponded to moderate-risk surgical procedures. Consequently, 30-day reoperation rate should no longer be considered when evaluating complications and surgical performance after bariatric surgery, and postoperative follow-up up to 180 days would be much more informative.

Source of funding: none

Conflicts of interest: For all authors none were declared

Acknowledgments: This study was declared to the National Data Protection Commission (Commission Nationale Informatique et Libertés – CNIL – MR4312040918) for potential data issues through the “Reference Methodology 005” before performing any analysis. Please check at: <https://www.indesante.fr/fr/repertoire-public/etude-sous-mr-4312040>

References

1. Christou NV, Sampalis JS, Liberman M, Look D, Auger S, McLean AP, et al. Surgery decreases long-term mortality, morbidity, and health care use in morbidly obese patients. *Ann Surg.* 2004;240:416-23.
2. Reges O, Greenland P, Dicker D, Leibowitz M, Hoshen M, Gofer I, et al. Association of Bariatric Surgery Using Laparoscopic Banding, Roux-en-Y Gastric Bypass, or Laparoscopic Sleeve Gastrectomy vs Usual Care Obesity Management With All-Cause Mortality. *JAMA.* 2018;319:279-90.
3. Stenberg E, Szabo E, Agren G, Na slund E, Boman L, Bylund A, et al. Early complications after laparoscopic gastric bypass surgery: results from the Scandinavian Obesity Surgery Registry. *Ann Surg.* 2014;260:1040-7.
4. Brunaud L, Polazzi S, Lifante JC, Pascal L, Nocca D, Duclos A. Health Care Institutions Volume Is Significantly Associated with Postoperative Outcomes in Bariatric Surgery. *Obes Surg.* 2018;28:923-31.
5. Caiazzo R, Baud G, Clement G, Lenne X, Torres F, Dezfoulian G, et al. Impact of centralized management of bariatric surgery complications on 90-day mortality. *Ann Surg.* 2018;268:831-7.
6. Campos GM, Khoraki J, Browning MG, Pessoa BM, Mazzini GS, Wolfe L. Changes in Utilization of Bariatric Surgery in the United States From 1993 to 2016. *Ann Surg.* 2019;Aug 13. doi: 10.1097/SLA.0000000000003554. [Epubahead of print].
7. Nguyen NT, Nguyen B, Nguyen VQ, Ziogas A, Hohmann S, Stamos MJ. Outcomes of bariatric surgery performed at accredited vs nonaccredited centers. *J Am Coll Surg.* 2012;215:467-74.
8. Dimick JB, Nicholas LH, Ryan AM, Thumma JR, Birkmeyer JD. Bariatric surgery complications before vs after implementation of a national policy

- restricting coverage to centers of excellence. *JAMA*. 2013;309:792-9.
9. Dimick JB, Osborne NH, Nicholas L, Birkmeyer JD. Identifying high-quality bariatric surgery centers: hospital volume or risk-adjusted outcomes? *J Am Coll Surg*. 2009;209:702-6.
10. Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. *N Engl J Med*. 2011;364:2128-37.
11. Fry BT, Scally CP, Thumma JR, Dimick JB. Quality Improvement in Bariatric Surgery: The Impact of Reducing Postoperative Complications on Medicare Payments. *Ann Surg*. 2018;268:22-7.
12. Markar SR, Penna M, Karthikesalingam A, Hashemi M. The impact of hospital and surgeon volume on clinical outcome following bariatric surgery. *Obes Surg*. 2012;22:1126-34.
13. Krell RW, Finks JF, English WJ, Dimick JB. Profiling hospitals on bariatric surgery quality: which outcomes are most reliable? *J Am Coll Surg*. 2014;219:725-34.
14. Brunaud L, Payet C, Lifante JC, Duclos A. Regarding Manuscript "Impact of Centralized Management of Bariatric Surgery Complications on 90-day Mortality". *Ann Surg*. 2019;270:e47-8.
15. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004;240:205-13.
16. Armoiry X, Obadia JF, Pascal L, Polazzi S, Duclos A. Comparison of transcatheter versus surgical aortic valve implantation in high-risk patients: A nationwide study in France. *J Thorac Cardiovasc Surg*. 2018;156:1017-25.
17. Haviari S, Chollet F, Polazzi S, Payet C, Beauveil A, Colin C, et al. Effect of data validation audit on hospital mortality ranking and pay for performance. *BMJ Qual Saf*. 2018;doi:10.1136/bmjqs-2018-008039.
18. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130-9.
19. Janik MR, Rogula TG, Mustafa RR, Alhaj Saleh A, Khaitan L. Safety of Revision Sleeve Gastrectomy Compared to Roux-Y Gastric Bypass After Failed Gastric Banding: Analysis of the MBSAQIP. *Ann Surg*. 2018;14:1276-82.
20. Bohnen JD, Mavros MN, Ramly EP, Chang Y, Yeh DD, Lee J, et al. Intraoperative adverse events in abdominal surgery: What happens in the operating room does not stay in the operating room. *Ann Surg*. 2017;265:1119-25.
21. Iglehart JK. Prioritizing comparative-effectiveness research—IOM recommendations. *N Engl J Med* 2009;361:325-8.
22. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41-55.
23. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res*. 2011;46:399-424.
24. Marrie RA, Dawson NV, Garland A. Quantile regression and restricted cubic splines are useful for exploring relationships between continuous variables. *J Clin Epidemiol*. 2009;62:511-7.
25. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat*. 2011;10:150-61.
26. Lazzati A, De Antonio M, Paolino L, Martini F, Azoulay D, Iannelli A, et al. Natural History of Adjustable Gastric Banding: Lifespan and Revisional Rate: A Nationwide Study on Administrative Data on 53,000 Patients. *Ann Surg*. 2017;265:439-45.
27. Debs T, Petrucciani N, Kassir R, Iannelli A, Amor IB, Gugenheim J. Trends of bariatric surgery in France during the last 10 years: analysis of 267,466 procedures from 2005-2014. *Surg Obes Relat Dis*. 2016;12:1602-9.
28. Ibrahim AM, Dimick JB. What Metrics Accurately Reflect Surgical Quality ? *Annu Rev Med*. 2018;69:481-91.
29. Peterli R, Wo Inerhanssen BK, Peters T, Vetter D, Kro ll D, Borbe ly Y, et al. Effect of Laparoscopic Sleeve Gastrectomy vs Laparoscopic Roux-en-Y Gastric Bypass on Weight Loss in Patients With Morbid Obesity: The SM-BOSS Randomized Clinical Trial. *JAMA*. 2018;319:255-355.
30. Salminen P, Helmio M, Ovaska J, Juuti A, Leivonen M, Peromaa-Haavisto P, et al. Effect of Laparoscopic Sleeve Gastrectomy vs Laparoscopic Roux-en-Y Gastric Bypass on Weight Loss at 5 Years Among Patients With Morbid Obesity: The SLEEVEPASS Randomized Clinical Trial. *JAMA*. 2018;319:241-54.
31. Ladak F, Dang JT, Switzer NJ, Mocanu V, Birch DW, Karmali S. Rates of reoperation and nonoperative intervention within 30 days of bariatric surgery. *Surg Obes Relat Dis*. 2019;15:431-40.
32. Musella M, Vitiello A. Comment on: Rates of reoperation and intervention within 30 days of bariatric surgery. *Surg Obes Relat Dis*. 2019;15:e9-10.
33. Chaar ME, Lundberg P, Stoltzfus J. Thirty-day outcomes of sleeve gastrectomy versus Roux-en-Y gastric bypass: first report based on Metabolic and Bariatric Surgery Accreditation and Quality Improvement Program database. *Surg Obes Relat Dis*. 2018;14:545-51.
34. Osland E, Yunus RM, Khan S, Alodat T, Memon B, Memon MA. Postoperative Early Major and Minor Complications in Laparoscopic Vertical Sleeve Gastrectomy (LVSG) Versus Laparoscopic Roux-en-Y Gastric Bypass (LRYGB) Procedures: A

- Meta-Analysis and Systematic Review. *Obes Surg.* 2016;26:2273-84.
35. Colquitt JL, Pickett K, Loveman E, Frampton GK. Surgery for weight loss in adults. *Cochrane Database Syst Rev.* 2014;8:<https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD003641.pub4/epdf/abstract>.
36. Talsma AK, Lingsma HF, Steyerberg EW, Wijnhoven BP, Van Lanschot JJ. The 30-Day Versus In-Hospital and 90-Day Mortality After Esophagectomy as Indicators for Quality of Care. *Ann Surg.* 2014;260:267-73.
37. Young MT, Gebhart A, Phelan MJ, Nguyen NT. Use and Outcomes of Laparoscopic Sleeve Gastrectomy vs Laparoscopic Gastric Bypass: Analysis of the American College of Surgeons NSQIP. *J Am Coll Surg.* 2015;220:880-5.
38. Fysekidis M, Catheline JM. Are Guidelines for Standardized Outcome Reporting in Bariatric Surgery Responsible for Missing the Big Picture in Bariatric Surgery Related Major Complications ? *Ann Surg.* 2018;268:e12-e3.
39. Ibrahim AM, Thumma JR, Dimick JB. Reoperation and Medicare Expenditures After Laparoscopic Gastric Band Surgery. *JAMA Surg.* 2017;152:835-42.
40. Rios-Diaz AJ, Metcalfe D, Devin CL, Berger A, Palazzo F. Six-month readmissions after bariatric surgery: Results of a nationwide analysis. *Surgery.* 2019;Aug 6. pii: S0039-6060(19)30341-1. doi: 10.1016/j.surg.2019.06.003. [Epub ahead of print].
41. Payet C, Lifante JC, Carty MJ, Rabilloud M, Duclos A. Methodological quality of surgical mortality studies using large hospital databases: a systematic review. *Ann Surg.* 2017;265:1113-8

Dans ce travail, nous avons donc comparé le risque de reprise chirurgicale et sa sévérité dans les six mois post-opératoires entre les trois principales chirurgies bariatriques : la *sleeve*, le *bypass* et l'anneau gastrique. Nous avons mis en évidence que le nombre de reprises chirurgicales continuait d'augmenter entre le 1^{er} mois post-opératoire et le 6^{ème} postopératoire quelle que soit la chirurgie bariatrique réalisée et notamment pour les reprises de sévérité modérée. Ces résultats étaient donc en faveur d'un allongement de la durée du suivi des résultats chirurgicaux post-opératoires de 1 à 6 mois. Nous avons également mis en évidence que la *sleeve* était la procédure présentant le moins de risque de reprises chirurgicales y compris pour les sévérités modérées.

Nous avons utilisé les scores de propension pour comparer les procédures chirurgicales deux à deux. Nous avons apparié les patients au sein de chaque établissement afin de contrôler les facteurs structurels et organisationnels. Pour appairer les patients, nous nous sommes limités aux caractéristiques des patients qui sont mesurées dans les bases de données hospitalières. Nous n'avons donc pas exploité toute la richesse de ces bases et notamment les nombreux codes présents dans ces bases.

Nous avons donc réalisé une nouvelle étude, afin d'exploiter l'information issue de ces codes en utilisant la méthode des scores de propension à haute dimension. Nous aurions pu tester cette méthode pour comparer ces trois procédures chirurgicales mais comme les résultats de notre étude sont concordants avec ceux de la littérature, y compris avec les résultats des essais cliniques, nous avons préféré travailler sur une nouvelle thématique pour laquelle la controverse est plus présente selon le type d'étude utilisé.

D. Performance des scores de propension à haute dimension (article 4)

En pharmaco-épidémiologie, les scores de propension à haute dimension (hdPS) ont prouvé leur utilité pour tenir compte de facteurs confondants non mesurés. Notre objectif était d'évaluer la performance des hdPS par rapport aux PS pour contrôler le biais d'indication lors de la comparaison de résultats cliniques entre des patients bénéficiant de l'implantation d'une valve aortique par voie percutanée (TAVI) et des patients bénéficiant d'un remplacement chirurgical de valve aortique (SAVR).

Nous avons choisi de comparer ces deux interventions car les résultats divergent selon le type d'étude réalisée. Les essais cliniques randomisés contrôlés montrent une efficacité et une sécurité similaire pour les patients ayant bénéficié d'un TAVI comparé aux patients opérés d'un SAVR [139–142]. Au contraire, les études réalisées sur des bases hospitalières nationales avec appariement par score de propension suggèrent de moins bons résultats pour les patients ayant bénéficié d'un TAVI comparés à ceux opérés d'un SAVR [143,144]. Une raison régulièrement évoquée pour expliquer ces différences de résultats est que dans les études par appariement sur PS, tous les facteurs confondants ne sont pas contrôlés et notamment le haut risque chirurgical des patients. En effet, les patients à haut risque chirurgical vont préférentiellement être orientés vers un TAVI. Il y aura donc plus de patients à haut risque chirurgical chez les patients TAVI que SAVR, pouvant conduire à une surmortalité par exemple, lié au fait que les patients sont plus à risque et non pas à la technique en elle-même. Ce haut risque chirurgical repose en partie sur le calcul de l'EuroSCORE [145], variable non disponible dans les bases hospitalières, pouvant ainsi conduire à un biais d'indication lors de l'étude des résultats cliniques.

En utilisant les scores de propension à haute dimension, nous souhaitons mieux contrôler ces facteurs confondants en utilisant l'apport d'information additionnelle issue des codes présents dans les bases hospitalières.

A partir de la base PMSI nationale MCO, nous avons sélectionné les patients ayant bénéficié de l'une de ces deux interventions entre 2013 et 2017. Au sein de chaque hôpital, les séjours réalisés pour TAVI ou SAVR ont été appariés par : 1) PS considérant 20 caractéristiques patients, 2) hdPS considérant 20 caractéristiques patients et 300 variables additionnelles

issues des codes d'actes et de diagnostics de l'année précédant l'hospitalisation. En utilisant chacune de ces méthodes, les risques observés de décès, reprise chirurgicale et accident vasculaire cérébral de 1 mois à 3 ans ont été comparés entre TAVI et SAVR avec des modèles de Cox ou Fine et Gray.

Article 4: Payet C, Polazzi S, Obadia JF, Armoiry X, Labarère J, Rabilloud M, Duclos A. High-dimensional propensity score to control indication bias in surgical comparative effectiveness studies : a case study on aortic valve replacement. *Journal of Clinical Epidemiology* (soumis).

High-dimensional propensity score to control indication bias in surgical comparative effectiveness studies: a case study on aortic valve replacement

Cécile Payet^{ab}, Stéphanie Polazzi^{ab}, Jean-François Obadia^c, Xavier Armoiry^{def}, José Labarère^{gh}, Muriel Rabilloud^{ij}, Antoine Duclos^{ab}

^aHospices Civils de Lyon, Pôle de Santé Publique, Service des Données de Santé, Lyon, F-69003, France

^bUniversité Claude Bernard Lyon 1, Health Services and Performance Research Lab (HESPER EA7425), F-69008 Lyon, France

^cService de Chirurgie Cardio-thoracique et Transplantation, Hôpital Cardio-thoracique Louis Pradel, Lyon-Bron, Avenue du Doyen Lépine, F-69500 Bron, France.

^dUniversity of Warwick, Warwick medical school, Division of Health Sciences, Gibbet Hill road, CV47AL Coventry, UK

^eHôpital Edouard Herriot, Département pharmacie/Place d'Arsonval, Lyon, France

^fUMR-CNRS 5510/MATEIS, F-69008, Lyon, France.

^gUniv. Grenoble Alpes, TIMC, UMR 5525 CNRS, F38706 Grenoble, France

^hQuality of care unit, CIC 1406, Grenoble Alpes University Hospital, F38043 Grenoble, France

ⁱHospices Civils de Lyon, Pôle de Santé Publique, Service de Biostatistique, Lyon, France

^jUniversité de Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biostatistique-Santé, F69100 Villeurbanne, France

Abstract

Objective: To evaluate the performance of high-dimensional propensity scores (hdPS) for controlling indication bias as compared to propensity scores (PS) in surgical comparative effectiveness studies.

Study design and Setting: Patients who underwent interventional transcatheter aortic valve implantation (TAVI) or surgical aortic valve replacement (SAVR) between 2013 and 2017 were included from the French nationwide hospitals. At each hospital level, matched pairs of TAVI and SAVR patients were formed using PS considering 20 patient baseline characteristics, and hdPS considering the same patient characteristics and 300 additional variables from procedures and diagnoses codes the year before surgery. We compared death, reoperation and stroke up to 3 years between TAVI and SAVR using Cox or Fine and Gray models.

Results: Before matching, 12/20 patient characteristics were imbalanced between the TAVI (n=29,995) and SAVR (n=29,063) patients included. No significant imbalance persisted after matching with both methods. Hazard-ratio of 1-year death, reoperation and stroke were 1.3 [1.1;1.4], 1.6[1.1;2.4] and 1.4 [1.2;1.7] for TAVI relative to SAVR with PS and 1.1 [1.0;1.3], 1.3 [0.8;2.0] and 1.3 [1.0;1.6] with hdPS.

Conclusion: hdPS is an interesting alternative for PS to control indication bias in comparative studies of surgical procedures.

What is new?

What is known on this topic?

- The high-dimensional propensity score (hdPS) method has been proven useful to control for unmeasured or residual confounding by indication
- This method is mainly used in pharmacoepidemiological studies of drug effect based on health insurance databases

What does it add?

- HdPS matching is an effective approach for controlling confounding by indication in comparative studies of surgical procedures,
- HdPS approach can be used in future comparative effectiveness studies using nationwide hospital databases

Introduction

In most industrialized countries, hospital administrative databases can reflect how health services deliver routine patient care. These data represent a valuable source for the investigation of care effectiveness and safety nationwide, but studies

may be flawed by the lack of information on important confounders [1,2]. This is of special importance when interpreting outcomes in surgical comparative effectiveness studies when certain types of patients are more likely to receive surgery owing to possible confounding by indication. For example,

when selecting a technique for a symptomatic patient with severe aortic stenosis, interventional cardiologists and cardiac surgeons may be more prone to perform an interventional transcatheter aortic valve implantation (TAVI) rather than a surgical aortic valve replacement (SAVR) if the patient is elderly or at high surgical risk. Thus, comparison of outcomes would be confounded by imbalances in baseline characteristics; that is, we would expect worse outcomes in the TAVI group, even if there was no true difference between the approaches [3]. The propensity score (PS) approach is frequently used to control confounding by indication [4–8] but is limited to investigator-specified covariates available within the examined databases.

To control for unmeasured or residual confounding by indication, the high-dimensional propensity score (hdPS) method has been proven useful in pharmacoepidemiological studies of drug effect based on health insurance databases [9–15]. Even though certain confounders may not be directly available, these databases can be understood and analyzed as a set of proxies for unmeasured factors that indirectly describe the health status of patients. Therefore, besides investigator-specified covariates, the hdPS method also uses a computerized algorithm to select a large number of potential confounders contained within the examined database [9].

Detailed codes regarding diagnoses and procedures recorded at discharge for all inpatient stays are available from nationwide hospital databases. To inform the design of future surgical comparative effectiveness studies using such databases it is therefore interesting to evaluate the potential of the

hdPS methodology for controlling confounding by indication with regard to traditional PS methodology. In the present study, we used both methodologies to compare clinical outcomes between patients undergoing TAVI versus SAVR. We opportunistically chose to compare those procedures because several randomized controlled trials (RCTs) have shown similar clinical efficacy and safety of TAVI compared with SAVR [16–19]. Conversely, evidence from PS matching studies conducted with real-life data, such as nationwide hospital databases, suggests worse outcomes for patients who undergo TAVI compared to those who had SAVR [20,21].

Methods

Study design and data source

Patients who underwent TAVI or SAVR in France over 5 years were matched within hospital to control for potential confounders related to specific care processes and dedicated surgical teams at the institutional level. Firstly, matched pairs of TAVI and SAVR patients were formed using PS. Secondly, matching was carried out using an hdPS method that employed additional information available in the dataset, especially codes for inpatient diagnoses and procedures, to improve the control for confounding factors. Finally, performance of hdPS was compared to PS using two distinct approaches, 1) by examining the balance achieved on key potential confounders between groups matched on these two scores [14,22–24], and 2) by comparing the measures of associations obtained from the matched groups to a gold standard comparator [9,10,12,14].

Table 1 – Baseline characteristics of patients before and after matching

	Before matching			PS-matching			HdPS-matching		
	SAVR N=29,063 (%)	TAVI N=29,995 (%)	Standardized difference	SAVR N=9,498 (%)	TAVI N=9,498 (%)	Standardized difference	SAVR N=7,157 (%)	TAVI N=7,157 (%)	Standardized difference
Male sex	18,818 (64.7%)	14,813 (49.4%)	-0.31	5,415 (57.0%)	5,404 (56.9%)	-0.00	4,055 (56.7%)	3,999 (55.9%)	-0.02
Age, mean (SD), y	71.7 (10.1)	83.1 (6.7)	1.32	77.7 (7.3)	78.2 (7.2)	0.08	77.8 (7.5)	78.4 (7.5)	0.07
Income, mean (SD), €	20,940.7 (3280.9)	21,226.4 (3712.2)	0.08	21,098.8 (3464.6)	21,051.2 (3499.5)	-0.01	21,145.1 (3508.1)	21,122.9 (3534.2)	-0.01
No. of days of hospitalization the previous years, mean (SD)	8.0 (11.9)	17.8 (21.0)	0.57	12.4 (16.6)	13.4 (17.5)	0.06	12.0 (15.9)	13.0 (16.8)	0.06
Emergency procedure	513 (1.8%)	508 (1.7%)	-0.01	148 (1.6%)	195 (2.1%)	0.04	157 (2.2%)	153 (2.1%)	-0.00
Congestive heart failure	7,561 (26.0%)	9,691 (32.3%)	0.14	2,901 (30.5%)	3,075 (32.4%)	0.04	2,261 (31.6%)	2,358 (32.9%)	0.03
Cardiac arrhythmias	16,823 (57.9%)	15,793 (52.7%)	-0.11	5,405 (56.9%)	5,410 (57.0%)	0.00	4,093 (57.2%)	4,069 (56.9%)	-0.01
Pulmonary circulation disorders	1,809 (6.2%)	2,935 (9.8%)	0.13	792 (8.3%)	776 (8.2%)	-0.01	606 (8.5%)	572 (8.0%)	-0.02
Peripheral vascular disease	2,528 (8.7%)	2,982 (9.9%)	0.04	970 (10.2%)	979 (10.3%)	0.00	756 (10.6%)	758 (10.6%)	0.00
Hypertension	16,326 (56.2%)	14,454 (48.2%)	-0.16	5,024 (52.9%)	4,889 (51.5%)	-0.03	3,798 (53.1%)	3,673 (51.3%)	-0.03
Chronic pulmonary disease	2,667 (9.2%)	2,781 (9.3%)	0.00	1,009 (10.6%)	965 (10.2%)	-0.02	748 (10.5%)	682 (9.5%)	-0.03
Diabetes	6,876 (23.7%)	6,090 (20.3%)	-0.08	2,319 (24.4%)	2,249 (23.7%)	-0.02	1,682 (23.5%)	1,659 (23.2%)	-0.01
Renal disease	2,299 (7.9%)	4,607 (15.4%)	0.23	1,179 (12.4%)	1230 (13.0%)	0.02	850 (11.9%)	896 (12.5%)	0.02
Liver disease	581 (2.0%)	467 (1.6%)	-0.03	255 (2.7%)	253 (2.7%)	-0.00	178 (2.5%)	182 (2.5%)	0.00
Obesity	6,567 (22.6%)	3,296 (11.0%)	-0.31	1,556 (16.4%)	1,610 (17.0%)	0.02	1,123 (15.7%)	1,189 (16.6%)	0.03
Myocardial infarction	969 (3.3%)	1,057 (3.5%)	0.01	340 (3.6%)	379 (4.0%)	0.02	269 (3.8%)	285 (4.0%)	0.01
Cerebrovascular disease	2,020 (7.0%)	1,998 (6.7%)	-0.01	707 (7.4%)	683 (7.2%)	-0.01	528 (7.4%)	541 (7.6%)	0.01
Year			0.50			0.07			0.05
2013	6,637 (22.8%)	3,061 (10.2%)		1,538 (16.2%)	1,377 (14.5%)		1,129 (15.8%)	1,051 (14.7%)	
2014	6,343 (21.8%)	4,231 (14.1%)		1,747 (18.4%)	1,645 (17.3%)		1,296 (18.1%)	1,227 (17.1%)	
2015	5,800 (20.0%)	5,878 (19.6%)		1,918 (20.2%)	1,970 (20.7%)		1,510 (21.1%)	1,477 (20.6%)	
2016	5,210 (17.9%)	7,651 (25.5%)		2,074 (21.8%)	2,180 (23.0%)		1,556 (21.7%)	1,589 (22.2%)	
2017	5,073 (17.5%)	9,174 (30.6%)		2,221 (23.4%)	2,326 (24.5%)		1,666 (23.3%)	1,813 (25.3%)	
Cardiac implant	1,327 (4.6%)	6,390 (21.3%)	0.52	920 (9.7%)	997 (10.5%)	0.03	710 (9.9%)	799 (11.2%)	0.04
Pacemaker	514 (1.8%)	1,595 (5.3%)	0.19	308 (3.2%)	325 (3.4%)	0.01	234 (3.3%)	247 (3.5%)	0.01
Mean standardized difference			0.24			0.02			0.02

PS, propensity score; hdPS, high-dimensional propensity score; TAVI, transcatheter aortic valve implantation; SAVR, surgical aortic valve replacement; SD, standard deviation

This study was based on the French Medical Information System [Programme de Médicalisation des Systèmes d'Information (PMSI), source: ATIH], which is a large hospital database with prospectively collected data from all public and private hospitals in France. The database is routinely implemented for the purpose of care reimbursement, which in turn led to strong accuracy and exhaustive collection of data. Moreover, the PMSI has a system of coding with strict variable definitions and a subset of records audited on a regular basis to avoid coding errors. Inpatient stays are converted into one Diagnosis-Related Group based on standard discharge abstracts containing compulsory information about the patient, primary and secondary diagnoses using the International Classification of Diseases (10th revision - ICD-10 codes), and procedural codes associated with the care provided using a detailed classification.

Patient demographics (sex and age), comorbidities according to Charlson and/or Elixhauser algorithms [25], the year, the type and emergency context of surgical procedure within the hospitalization stay were extracted. We subsequently used patient-unique anonymous identifier to link his/her stays in acute care hospitalizations, allowing the extraction of hospitalization-related data for the 12 months preceding TAVI or SAVR. To define patients' socioeconomic status, the median household income of the patients' residence provided by the National Institute of Statistics and Economic Studies was used.

This study was strictly observational and based on anonymous data. Therefore, in accordance with French legislation in place at the time of the it did not require the written informed consent from the participants or the authorization from an ethics committee. This study was authorized by the national data protection commission (*Commission Nationale de l'Informatique et des Libertés*, MR 0620210918).

Study population and outcomes

All adults (patients ≥ 18 years) who underwent TAVI or SAVR in French institutions between January 1, 2013, and December 31, 2017. Both interventions were identified in the PMSI database by specific procedural codes were selected. To homogenize the study population, only cases with a main diagnosis of rheumatic or nonrheumatic aortic valve disease (ICD-10, codes I06* or I35*) were included. For each procedure, the first hospitalization for each patient identified was considered as the index stay. Patients having both procedures or experienced ambulatory care, with acute infectious endocarditis, or with data inaccuracies during the study period were not retained for analysis.

The primary endpoint was all-cause death from the index hospitalization up to 3 years following TAVI or SAVR. Secondary outcomes were reoperation and stroke.

Statistical analysis

To control for the non-random assignment of patients to the 2 procedures, we formed matched pairs of TAVI and SAVR patients using either PS or hdPS method. Both PS and hdPS were estimated as the predicted probability of a patient undergoing TAVI using a logistic regression model including selected covariates. The PS model covariates were sex, age (coded as continuous variable with linear, quadratic, and cubic terms), household income (coded as continuous variable with linear, quadratic, and cubic terms), number of days spent in acute care hospitalizations the year before the index stay (coded as continuous variable with linear, quadratic, and cubic terms), emergency procedure, a selection of comorbidities (i.e. congestive heart failure, cardiac arrhythmia, pulmonary circulation disorder, peripheral vascular disease, hypertension, chronic pulmonary disease, diabetes, renal disease, liver disease, obesity, myocardial infarction, and cerebrovascular disease), cardiac implant, pacemaker, and year of the procedure.

HdPS were estimated using the variables of the PS model as well as automatically selected covariates from the database. Two dimensions from which covariates might be selected were identified: inpatient diagnoses and inpatient procedures. Candidate covariates from these dimensions were ranked by prevalence, recurrence and potential to cause bias, as described by Schneeweiss *et al.* [9]. We assessed each variable's potential for confounding based on the bias ranking method [9,22]. The first 300 of these empirically identified covariates [13] were included in the model to estimate the hdPS.

Patients with the closest PS or hdPS within hospital were then matched, using a greedy 1:1 algorithm without replacement and requiring that the logit of the propensity score of a patient who underwent TAVI and one who underwent SAVR be < 0.20 standard deviation of the propensity score logit of each other [23]. Standardized differences were used to assess the degree of balance between the matched groups for baseline characteristics [26]. An absolute standardized difference (ASD) ≤ 0.10 was chosen to indicate a negligible difference in the mean or prevalence of a variable between groups. Balance for continuous variables was also assessed using graphical methods (side-by-side boxplots, empirical cumulative distribution functions, empirical QQ-plots) to compare the distributions across the two groups.

Categorical variables were presented using absolute and relative frequencies and continuous variables were presented using means and standard deviations (SD). Estimates were provided along with 95% confidence interval (CI) and p-values <0.05 were considered to indicate statistical significance.

Outcomes were analyzed as time-to-event variables, and were evaluated at different time points (one month and then yearly up to three years after index procedure). To compare outcomes between procedure groups, hazard ratios (HR) and 95%CI for TAVI relative to SAVR patients were estimated using Cox proportional hazard models for all-cause death. Reoperation and stroke were analyzed using Fine and Gray's models to take into account death as competing risk, with robust variance estimator to account for clustering within matched pairs.

In sensitivity analyses, Cox proportional hazard models for reoperation and stroke were constructed to ensure that results were consistent with Fine and Gray's models [27]. hdPS analyses were also replicated by including 100, 200, and 500 identified covariates for the primary endpoint.

Data manipulation and analyses were performed using SAS software (version 9.4; SAS Institute Inc., Cary, NC, US).

Results

Baseline patient characteristics before matching

After applying the selection criteria, 29,995 patients who underwent TAVI and 29,063 patients underwent SAVR in 48 French hospitals between 2013 and 2017 were included (Figure E1). Among the 20 examined characteristics, 12 (60.0%) yielded an ASD >0.10 before matching, indicating imbalance between the groups (mean ASD=0.24). Compared with SAVR patients, TAVI patients were older (ASD=1.33), more likely to be women (ASD=0.31), to have renal disease (ASD=0.23) or cardiac implant (ASD=0.52), and had a greater number of days of hospitalization the previous year (ASD=0.57). The prevalence of obesity was lower among TAVI patients (ASD=0.32). The year of intervention was also unbalanced between groups (ASD=0.50) since the number of TAVI tripled in five years (3,061 in 2013 to 9,174 in 2017) while the number of SAVR decreased during the same period (6,637 to 5,073; Table 1).

PS analysis

For matching on PS, 9,498 pairs of patients were retained (Figure E1). Pre- and post-matching propensity score distributions are presented in Figure 1A and B, respectively, and show an improvement in overlap. No significant imbalance persisted after matching for all 20 examined patient characteristics;

ASD ranged from 0.00 to 0.08, and the mean ASD was 0.02 (Table 1).

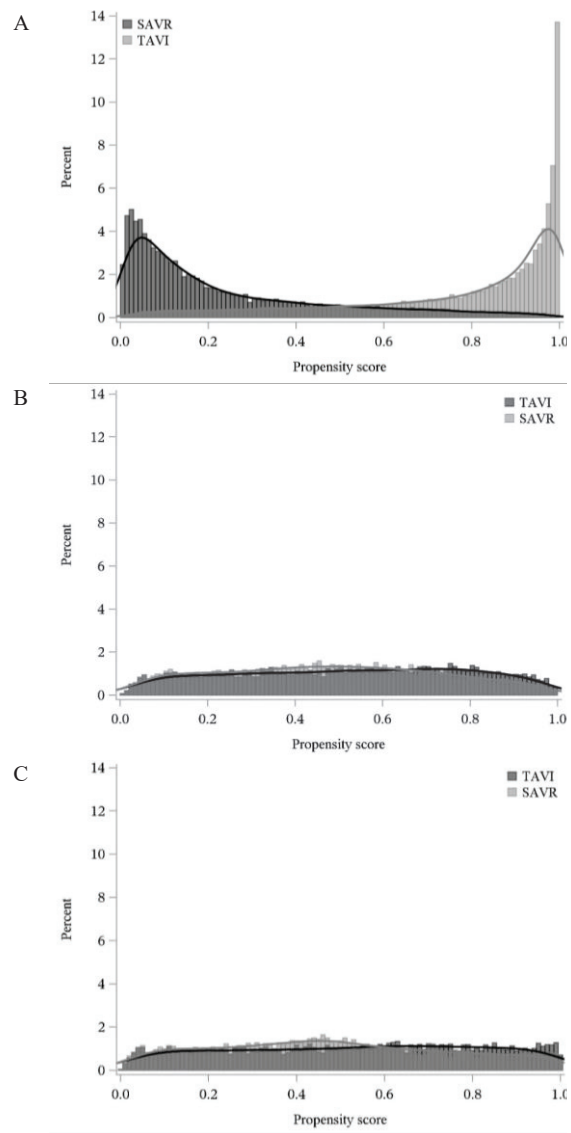


Figure 1 – Propensity score distribution before (A), after propensity score (B) and after high-dimensional propensity score matching (C). TAVI, transcatheter aortic valve implantation; SAVR, surgical aortic valve replacement.

The risk of death at 1 month was lower after TAVI than after SAVR (HR=0.80, 95%CI [0.68;0.93]); a higher risk of death was observed 1 year after TAVI (1.26 [1.14;1.40]) and this was sustained up to 3 years (1.53 [1.42;1.65]). Risk of reoperation was higher at any time point (from 1.92 [1.17;3.14] at 1 month to 1.87 [1.29;2.69] at 3 years), as was the risk of stroke (from 2.94 [1.67;5.19] at 1 month to 1.23 [1.07;1.42] at 3 years) (Table 2, Figure 2).

Table 2 – Outcomes for TAVI and SAVR after PS or hdPS matching

	PS-matching			hdPS-matching		
	SAVR N=9 498 (%)	TAVI N=9 498 (%)	HR [CI 95%]	SAVR N= 7 157 (%)	TAVI N= 7 157 (%)	HR [95%CI]
All-cause death						
At 1 month	354 (3.7%)	281 (3.0%)	0.80 [0.68-0.93]	294 (4.0%)	204 (2.8%)	0.70 [0.59-0.84]
At 1 year	686 (7.2%)	900 (9.5%)	1.26 [1.14-1.40]	560 (7.6%)	658 (9.0%)	1.14 [1.02-1.28]
At 2 year	896 (9.4%)	1376 (14.5%)	1.49 [1.37-1.62]	734 (10%)	1 010 (13.8%)	1.35 [1.22-1.48]
At 3 year	1104 (11.6%)	1717 (18.1%)	1.53 [1.42-1.65]	906 (12.4%)	1 269 (17.3%)	1.39 [1.27-1.51]
Reoperation						
At 1 month	24 (0.3%)	46 (0.5%)	1.92 [1.17-3.14]	26 (0.4%)	33 (0.4%)	1.27 [0.76-2.12]
At 1 year	40 (0.4%)	64 (0.7%)	1.60 [1.08-2.38]	37 (0.5%)	48 (0.7%)	1.30 [0.85-1.99]
At 2 year	43 (0.5%)	73 (0.8%)	1.70 [1.17-2.48]	41 (0.6%)	55 (0.7%)	1.34 [0.90-2.01]
At 3 year	44 (0.5%)	82 (0.9%)	1.87 [1.29-2.69]	43 (0.6%)	63 (0.9%)	1.47 [0.99-2.17]
Stroke						
At 1 month	16 (0.2%)	47 (0.5%)	2.94 [1.67-5.19]	16 (0.2%)	39 (0.5%)	2.44 [1.36-4.37]
At 1 year	172 (1.8%)	241 (2.5%)	1.41 [1.16-1.71]	132 (1.8%)	166 (2.3%)	1.26 [1.00-1.59]
At 2 year	280 (2.9%)	355 (3.7%)	1.28 [1.09-1.49]	221 (3.0%)	256 (3.5%)	1.17 [0.98-1.40]
At 3 year	358 (3.8%)	438 (4.6%)	1.23 [1.07-1.42]	291 (4.0%)	329 (4.5%)	1.14 [0.98-1.34]

PS, propensity score; hdPS, high-dimensional propensity score; TAVI, transcatheter aortic valve implantation; SAVR, surgical aortic valve replacement; HR, hazard ratio; CI, confidence interval

hdPS analysis

For matching on hdPS, 7,157 pairs were retained (Figure E1). Results were similar to PS analysis for the improvement in overlap of post-matching propensity score distribution (Figure 1C) and for the balance of baseline characteristics between groups. No significant imbalance persisted after matching for all 20 examined patient characteristics; ASD ranged from 0.00 to 0.07, and the mean was 0.02 (Table 1).

hdPS analysis also indicated a lower risk of death at 1 month after TAVI than after SAVR (0.70 [0.59;0.84]) and a higher risk of death after 1 year (1.14 [1.02;1.28]) that was sustained up to 3 years (1.39 [1.27;1.51]). The risk of stroke was also higher at 1 month (2.44 [1.36;4.37]) and 1 year (1.26 [1.00;1.59]). Unlike PS analysis, the risk of reoperation was not significantly different between groups at all time points studied, as was the risk of stroke at 2 years.

These findings were rather unchanged in sensitivity analyses (see Tables E1 and E2).

Discussion

In the present study, TAVI was associated with higher long-term mortality and non-significant increase in HR of stroke and reoperation, after matching on hdPS. Compared with hdPS, PS matching resulted in higher HR point estimates for mortality, reoperation, and stroke following TAVI. Overall, the results presented herein provide evidence on the feasibility of hdPS matching for controlling confounding by indication in comparative effectiveness study of surgical procedures based on a nationwide hospital database.

We examined various dimensions of hdPS relative to PS matching for comparing outcomes between TAVI and SAVR patients. Firstly, we examined residual imbalance on key potential confounders between study groups. Before matching, patient

profiles showed imbalance on many key baseline characteristics suggesting the presence of confounding by indication. After matching, comparable levels of balance were achieved on key potential confounders between study groups for both methods. The two approaches induced excellent balance on 20 examined patient characteristics and had similar mean ASD. These results were consistent with those found in other studies where hdPS improve covariate balance in similar manner [28] or induce marginally better balance compared to the conventional PS approach [14]. Secondly, we examined the extent to which HR estimates were consistent with the results of published RCTs which is traditionally done to evaluate hdPS methodology [9,10,12–14,29,30] . [16–19].

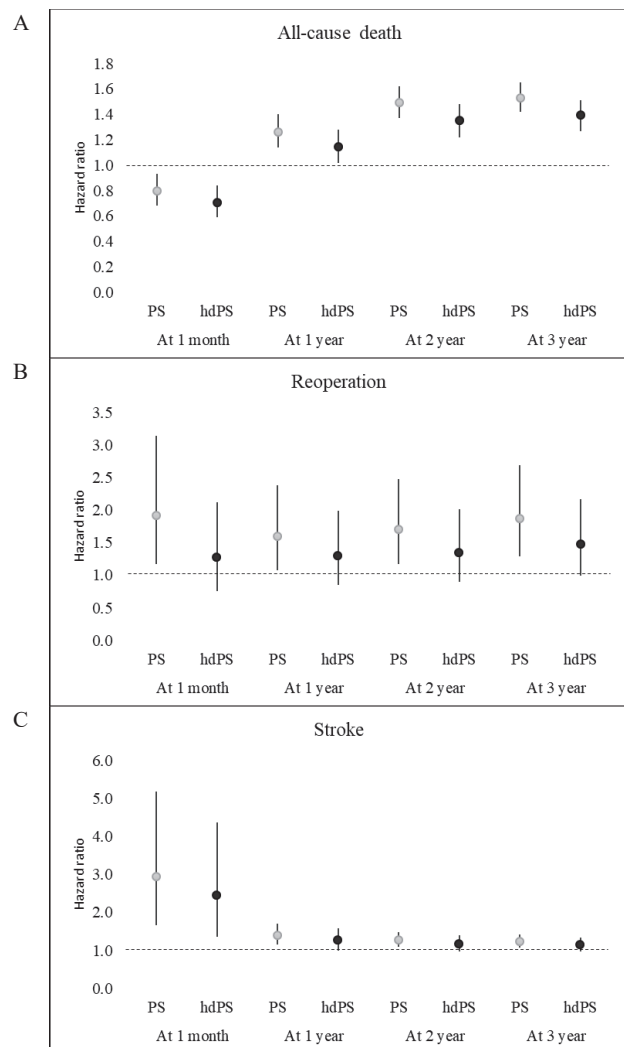


Figure 2 – Hazard Ratios and 95% confidence intervals of all-cause of death (A), reoperation (B) and stroke (C) for transcatheter aortic valve implantation (TAVI) relative to surgical aortic valve replacement (SAVR) using propensity score matching (PS) or high-dimensional propensity score (hdPS) matching to 1 month up to 3 year.

Thus, the most performant methodology had the closest results to RCTs. In our purpose, large RCTs and meta-analyses of RCTs conducted in high, intermediate or low risk patients have shown similar or improved outcomes with TAVI compared with SAVR at different endpoints

Results of the conventional PS-matched analysis herein contrasted with these studies that reported poorer outcomes with TAVI than with SAVR within 3 years except for the risk of death in the first postoperative month. This last result was new compared to other studies that used PS [20,21] and could be explained by a lower surgical risk of patients included in our study. But other results were consistent with the literature and has been emphasized in a meta-analysis [20]. Results of the hdPS-matched analysis were slightly different. Unlike PS, no significant differences in the risks of reoperation and stroke at 2 years were found between TAVI and SAVR. Moreover, when results showed poorer outcomes with TAVI, hdPS estimates moved further towards a neutral effect consistent with findings from RCTs. Thus, hdPS allowed to take into account some unobserved factors that indirectly describe the health status of patients. With this methodology, results were closer to the expected findings based on RCT even if differences persisted, in particular for death.

One possible reason for these differences was that RCTs were conducted among carefully selected populations. TAVI patients included in RCTs had generally stricter selection criteria than the real-life practice where patient eligibility for TAVI is decided by the “heart team”. Alternatively, persistent unmeasured confounding could also explain these results. The risk of bias with PS or hdPS studies is to omit some potential confounders that can alter the comparability of populations and therefore threaten the validity of outcome measures [31]. Although hdPS adjusted for hundreds of covariates and could adjust for proxies of residual confounders, the possibility that further unmeasured confounders remained after using hdPS could not be ruled out. Indeed, TAVI are preferentially channeled to patients at higher surgical risk and data granularity did not allow us to accurately describe every patients profile with respect to the surgical risk. For example, TAVI indication is based on the euro-SCORE or STS score which could not be precisely calculated in the PMSI database because the clinical variables, diagnostic or procedures codes that were available were not strictly those listed or were not as accurately defined among the factors that were accounted for in the euro-SCORE or STS score calculations. However we added the number of hospitalization days in acute care consumed the year preceding the index stay to account for unmeasured confounders. Indeed summary measures of healthcare utilization have been shown to perform as

well or even better than other comorbidity scores in predicting the risks of clinical outcomes [32]. A weakness of large hospital databases is the miscoding of diagnoses during hospital stays that can underestimate patients’ comorbidities [33]. This issue is not specific to a disease area or to certain type of procedure and is more influenced by a strong coding variability between healthcare providers and across years. Given this, we matched pairs of patients who underwent either TAVI or SAVR inside the same hospital and we took account of the procedure year.

To our knowledge, the current study is the first to use hdPS for matching in surgical comparative effectiveness study based on nationwide hospital databases. The present study highlighted the potential of hdPS for controlling confounding by indication in comparative effectiveness studies of surgical procedure using large medico-administrative databases. The use of hdPS methodology allowed to take into account some unmeasured factors that indirectly describe patient profile. Furthermore, in contrast to investigator-specified covariates, hdPS algorithm automatically selects covariates. It may be useful when there are few candidates for clinically selected variables, when they are unavailable or when investigators are uncertain about the relevant confounders [12]. Moreover, the use of hdPS methodology within future studies will become increasingly interesting with the improvement of coding of procedures and diagnoses every year. To further improve results, it would be interesting to use machine-learning extensions to optimize automated covariate selection [15,34–37]. However, we only examined the performance of hdPS in a single context and we cannot exclude that the findings would have been different in other countries or settings. Further work is needed to fully appraise the position of hdPS in such settings.

Conflicts of Interest and Source of Funding: None

References

- [1] Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding Control in Healthcare Database Research: Challenges and Potential Approaches. *Medical Care* 2010;48:S114–20. <https://doi.org/10.1097/MLR.0b013e3181d8be3>.
- [2] Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of Clinical Epidemiology* 2005;58:323–37. <https://doi.org/10.1016/j.jclinepi.2004.10.012>.
- [3] Hemmila MR. Introduction to Propensity Scores: A Case Study on the Comparative Effectiveness of Laparoscopic vs Open

- Appendectomy. *Archives of Surgery* 2010;145:939. <https://doi.org/10.1001/archsurg.2010.193>.
- [4] Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research* 2011;46:399–424. <https://doi.org/10.1080/00273171.2011.568786>.
- [5] Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 1983;70:41–55.
- [6] Groenwold RHH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. *Journal of Clinical Epidemiology* 2009;62:22–8. <https://doi.org/10.1016/j.jclinepi.2008.02.011>.
- [7] Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology* 2010;63:826–33. <https://doi.org/10.1016/j.jclinepi.2009.11.020>.
- [8] Lonjon G, Boutron I, Trinquart L, Ahmad N, Aim F, Nizard R, et al. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg* 2014;259:18–25. <https://doi.org/10.1097/SLA.0000000000000256>.
- [9] Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional Propensity Score Adjustment in Studies of Treatment Effects Using Health Care Claims Data. *Epidemiology* 2009;20:512–22. <https://doi.org/10.1097/EDE.0b013e3181a663cc>.
- [10] Garbe E, Kloss S, Suling M, Pigeot I, Schneeweiss S. High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications. *European Journal of Clinical Pharmacology* 2013;69:549–57. <https://doi.org/10.1007/s00228-012-1334-2>.
- [11] Cadarette SM, Ban JK, Consiglio GP, Black CD, Dubins D, Marin A, et al. Diffusion of Innovations model helps interpret the comparative uptake of two methodological innovations: co-authorship network analysis and recommendations for the integration of novel methods in practice. *Journal of Clinical Epidemiology* 2017;84:150–60. <https://doi.org/10.1016/j.jclinepi.2016.12.006>.
- [12] Polinski JM, Schneeweiss S, Glynn RJ, Lii J, Rassen JA. Confronting “confounding by health system use” in Medicare Part D: comparative effectiveness of propensity score approaches to confounding adjustment: COMPARATIVE EFFECTIVENESS OF PS APPROACHES. *Pharmacoepidemiology and Drug Safety* 2012;21:90–8. <https://doi.org/10.1002/pds.3250>.
- [13] Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate Selection in High-Dimensional Propensity Score Analyses of Treatment Effects in Small Samples. *American Journal of Epidemiology* 2011;173:1404–13. <https://doi.org/10.1093/aje/kwr001>.
- [14] Guertin JR, Rahme E, Dormuth CR, LeLorier J. Head to head comparison of the propensity score and the high-dimensional propensity score matching methods. *BMC Medical Research Methodology* 2016;16. <https://doi.org/10.1186/s12874-016-0119-1>.
- [15] Schneeweiss S. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clinical Epidemiology* 2018;Volume 10:771–88. <https://doi.org/10.2147/CLEP.S166545>.
- [16] Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, et al. Transcatheter Aortic-Valve Implantation for Aortic Stenosis in Patients Who Cannot Undergo Surgery. *New England Journal of Medicine* 2010;363:1597–607. <https://doi.org/10.1056/NEJMoa1008232>.
- [17] Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK, et al. Transcatheter or Surgical Aortic-Valve Replacement in Intermediate-Risk Patients. *New England Journal of Medicine* 2016;374:1609–20. <https://doi.org/10.1056/NEJMoa1514616>.
- [18] Adams DH, Popma JJ, Reardon MJ, Yakubov SJ, Coselli JS, Deeb GM, et al. Transcatheter Aortic-Valve Replacement with a Self-Expanding Prosthesis. *New England Journal of Medicine* 2014;370:1790–8. <https://doi.org/10.1056/NEJMoa1400590>.
- [19] Mack MJ, Leon MB, Thourani VH, Makkar R, Kodali SK, Russo M, et al. Transcatheter Aortic-Valve Replacement with a Balloon-Expandable Valve in Low-Risk Patients. *New England Journal of Medicine* 2019;380:1695–705. <https://doi.org/10.1056/NEJMoa1814052>.
- [20] Ando T, Takagi H, Representing ALICE (All-Literature Investigation of Cardiovascular Evidence) Group. Comparison of late mortality after transcatheter aortic valve implantation versus surgical aortic valve replacement: Insights from a meta-analysis. *Eur J Intern Med* 2017;40:43–9. <https://doi.org/10.1016/j.ejim.2017.01.023>.
- [21] Armoiry X, Obadia J-F, Pascal L, Polazzi S, Duclos A. Comparison of transcatheter versus surgical aortic valve implantation in high-risk patients: A nationwide study in France. *J Thorac Cardiovasc Surg* 2018;156:1017-1025.e4. <https://doi.org/10.1016/j.jtcvs.2018.02.092>.
- [22] Bross IDJ. Spurious effects from an extraneous variable. *Journal of Chronic Diseases* 1966;19:637–47. [https://doi.org/10.1016/0021-9681\(66\)90062-2](https://doi.org/10.1016/0021-9681(66)90062-2).
- [23] Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions

in observational studies. *Pharmaceutical Statistics* 2011;10:150–61. <https://doi.org/10.1002/pst.433>.

[24] Ali MS, Groenwold RHH, Belitser SV, Pestman WR, Hoes AW, Roes KCB, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of Clinical Epidemiology* 2015;68:122–31. <https://doi.org/10.1016/j.jclinepi.2014.08.011>.

[25] 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. <https://doi.org/10.1097/01.mlr.0000182534.19832.83>.

[26] Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine* 2009;28:3083–107. <https://doi.org/10.1002/sim.3697>.

[27] Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol* 2013;66:648–53. <https://doi.org/10.1016/j.jclinepi.2012.09.017>.

[28] Li L, Vollmer WM, Butler MG, Wu P, Kharbanda EO, Wu AC. A Comparison of Confounding Adjustment Methods for Assessment of Asthma Controller Medication Effectiveness. *American Journal of Epidemiology* 2014;179:648–59. <https://doi.org/10.1093/aje/kwt323>.

[29] Enders D, Ohlmeier C, Garbe E. The Potential of High-Dimensional Propensity Scores in Health Services Research: An Exemplary Study on the Quality of Care for Elective Percutaneous Coronary Interventions. *Health Services Research* 2018;53:197–213. <https://doi.org/10.1111/1475-6773.12653>.

[30] Hallas J, Pottgård A. Performance of the High-dimensional Propensity Score in a Nordic Healthcare Model. *Basic Clin Pharmacol Toxicol*

2017;120:312–7. <https://doi.org/10.1111/bcpt.12716>.

[31] Payet C, Lifante J-C, Carty MJ, Rabilloud M, Duclos A. Methodological Quality of Surgical Mortality Studies Using Large Hospital Databases: A Systematic Review. *Annals of Surgery* 2017;265:1113–8. <https://doi.org/10.1097/SLA.0000000000002119>.

[32] Schneeweiss S. Performance of Comorbidity Scores to Control for Confounding in Epidemiologic Studies using Claims Data. *American Journal of Epidemiology* 2001;154:854–64. <https://doi.org/10.1093/aje/154.9.854>.

[33] Armoiry X, Obadia J-F, Iung B, Polazzi S, Duclos A. Clinical outcomes and direct costs after transcatheter aortic valve implantation in French centres: a longitudinal study of 1332 patients using a national database. *Interactive Cardiovascular and Thoracic Surgery* 2016;23:883–8. <https://doi.org/10.1093/icvts/ivw278>.

[34] Franklin JM, Eddings W, Glynn RJ, Schneeweiss S. Regularized Regression Versus the High-Dimensional Propensity Score for Confounding Adjustment in Secondary Database Analyses. *American Journal of Epidemiology* 2015;182:651–9. <https://doi.org/10.1093/aje/kwv108>.

[35] Schneeweiss S, Eddings W, Glynn R, Paterno E, Rassen J, Franklin J. Variable Selection for Confounding Adjustment in High-dimensional Covariate Spaces When Analyzing Healthcare Databases. *Epidemiology* 2017;28:237–48. <https://doi.org/10.1097/EDE.0000000000000581>.

[36] Karim ME, Pang M, Platt RW. Can We Train Machine Learning Methods to Outperform the High-dimensional Propensity Score Algorithm? *Epidemiology* 2018;29:191–8. <https://doi.org/10.1097/EDE.0000000000000787>.

[37] Shortreed SM, Ertefaie A. Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics* 2017;73:1111–22. <https://doi.org/10.1111/biom.12679>.

Dans ce travail, nous avons comparé le risque de décès, de reprise chirurgicale et d'accident vasculaire cérébral dans les trois ans post-opératoires. Nous avons utilisé la méthode des PS et des hdPS pour contrôler le biais d'indication lors de la comparaison de résultats cliniques en vie réelle.

Avec la méthode PS, on retrouvait un sur-risque de décès, de reprise chirurgicale et d'accident vasculaire cérébral à long terme pour les patients TAVI. Avec la méthode hdPS, on retrouvait uniquement un sur-risque de décès à long terme. Ces estimations suggéraient des effets moins élevés qu'avec la méthode PS et se rapprochaient de celles des essais cliniques contrôlés randomisés. Nous avons donc montré que la méthode hdPS permettait de mieux contrôler le biais d'indication dans les études de comparaison de résultats cliniques menées à partir des bases hospitalières nationales.

IV. Synthèse

A. Discussion des travaux réalisés

L'objectif de la thèse était d'explorer des méthodes permettant de mieux contrôler des facteurs confondants lors de l'étude de résultats chirurgicaux dans les grandes bases de données hospitalières nationales en tenant compte de la finalité de l'étude.

Dans un premier temps, nous avons fait une revue de la littérature afin de décrire la qualité méthodologique des études analysant la mortalité chirurgicale dans ces bases de données. Cela nous a permis de mettre en évidence que ces bases étaient de plus en plus utilisées pour étudier les évolutions des résultats chirurgicaux au cours du temps, comparer l'efficacité réelle d'interventions chirurgicales, évaluer les politiques de santé et examiner les déterminants des résultats chirurgicaux. Bien que la qualité méthodologique des études se soit améliorée ces dernières années, ce travail nous a permis d'identifier des pistes d'amélioration pour contrôler de potentiels facteurs confondants à chaque étape de l'étude : au niveau du design d'étude, des stratégies de sélection de la population, de l'analyse statistique et des analyses de sensibilité éventuelles. Ce premier travail nous a donc permis de proposer différents travaux pour améliorer le contrôle de potentiels facteurs confondants, dans des thématiques de recherche actuelles. En particulier, nous nous sommes intéressés à la recherche de déterminants avec l'étude du lien volume-résultat des hôpitaux ainsi qu'à la comparaison de l'efficacité réelle de différentes interventions chirurgicales.

1. Contrôle des facteurs confondants lors de l'étude du lien volume résultat en chirurgie (Etude 2)

Nous nous sommes intéressés à la relation volume-résultat en étudiant les évolutions de volume des hôpitaux au cours du temps en fonction des résultats des patients pour dix procédures à haut risque. Nous nous sommes concentrés sur le concept du "plus on fait, mieux on fait" et nous avons supposé que les hôpitaux acquièrent une expertise en répétant des soins chirurgicaux spécifiques, conduisant ainsi à une amélioration des résultats chirurgicaux des patients. Néanmoins, les résultats obtenus n'étaient pas en faveur d'un effet

d'apprentissage individuel au niveau de l'hôpital puisque de manière générale, les résultats chirurgicaux ne dépendaient pas de l'évolution du volume des hôpitaux au cours du temps.

Pour contrôler les facteurs confondants, nous avons d'abord homogénéisé la population d'étude en utilisant des critères de sélection au niveau du patient, du séjour et de l'hôpital (Table 5). Nous avons ajusté les résultats chirurgicaux sur les caractéristiques des patients et des hôpitaux. Nous avons utilisé une variable proxy au niveau du patient qui était le revenu médian de la commune de résidence pour le patient, reflétant son niveau socio-économique [51,52]. Nous avons utilisé des variables proxies au niveau de l'hôpital, telles que le statut de l'hôpital, reflet de différences structurelles et organisationnelles pour la dotation en personnel ou la disponibilité de plateaux techniques performants par exemple [86]. Nous avons également ajusté les résultats chirurgicaux sur l'année de la chirurgie pour tenir compte des tendances séculaires dans l'évolution des résultats et des pratiques de codage. Au niveau de l'analyse statistique, nous avons tenu compte du fait que les patients d'un même hôpital présentaient des caractéristiques communes liées à la structure et l'organisation des soins en intégrant un effet centre. Enfin, nous avons multiplié les comparaisons entre procédures et résultats chirurgicaux étudiés afin de montrer la robustesse de nos résultats.

Table 5 – Prise en compte des facteurs confondants dans l'étude 2

Design	Cohorte historique avec 3 groupes d'exposition
Sélection population	Niveau du patient (wash-out d'un an pour identifier la 1 ^{ère} procédure) Niveau séjour (procédure spécifique, adulte, exclusion des soins palliatifs, séances, donneurs d'organe, prestations inter-établissement, séjours avec diagnostic principal de complication) Niveau hôpital (Métropole, établissement avec au moins 1 chirurgie par an sur la période d'étude)
Analyses statistiques	Ajustement sur les caractéristiques patients (âge, sexe, comorbidité de l'Elixhauser, année chirurgie) et une variable proxy (revenu médian) Ajustement sur les caractéristiques hôpitaux (statut de l'hôpital, volume de procédures, degré de spécialisation, taux d'attractivité) Modèle de survie tenant compte de l'effet cluster établissement et du risque compétitif selon l'évènement d'intérêt
Analyses de sensibilité	10 procédures chirurgicales 4 résultats chirurgicaux

L'utilisation de la base PMSI nationale nous a permis d'inclure tous les patients opérés d'une des chirurgies étudiées en France, sur la période de temps considérée. Cela nous a également permis de suivre les évolutions de volumes des hôpitaux ainsi que les résultats chirurgicaux sur cette même période. Les diverses hospitalisations et/ou reprises chirurgicales des patients, que ce soit dans l'établissement dans lequel ils ont eu leur intervention index ou dans tout autre établissement, ont également pu être étudiées puisqu'il est possible de relier les différents séjours des patients [82].

En revanche, une des principales limites de ce travail était liée aux données manipulées. Nous n'avons pas pu tenir compte des caractéristiques du chirurgien comme son volume de procédures ou sa courbe de performance puisque ces informations ne sont pas disponibles dans les bases. Ceci explique peut-être en partie le fait que nous n'ayons pas pu mettre en évidence la relation volume-résultat. Nous étions dans l'incapacité de savoir si les augmentations de volume de procédures étaient dues à l'augmentation du nombre de chirurgiens réalisant la procédure ou à l'augmentation du nombre de procédures réalisées par les mêmes chirurgiens. Selon que l'on soit dans un cas ou un autre, les résultats pourraient être différents.

2. Contrôle des facteurs confondants lors de la comparaison de l'efficacité réelle de deux interventions chirurgicales (Etude 3 et 4)

Dans les études 3 et 4 nous avons comparé, en vie réelle, des résultats chirurgicaux entre deux interventions, en contrôlant à chaque étape de l'étude de potentiels facteurs confondants.

Dans un premier temps, nous avons comparé le risque de reprise chirurgicale et sa sévérité, dans les six mois postopératoires pour des patients opérés en France, entre 2013 et 2015, d'une des trois principales chirurgies bariatriques : la sleeve, le bypass et l'anneau gastrique (Etude 3). Nous avons mis en évidence que le nombre de reprises chirurgicales continuait d'augmenter entre le 1^{er} mois postopératoire et le 6^{ème} postopératoire quelle que soit la chirurgie bariatrique réalisée et notamment pour les reprises de sévérité modérée. Ces résultats étaient donc en faveur d'un allongement de la durée du suivi des résultats chirurgicaux post-opératoires de 1 à 6 mois. Nous avons également mis en évidence que la

sleeve était la procédure présentant le moins de risque de reprises chirurgicales y compris pour les sévérités modérées.

Nous avons contrôlé les facteurs confondants au niveau du design d'étude en appariant les patients entre les procédures au sein des hôpitaux pour tenir compte des facteurs organisationnels ou des pratiques de codage liés à l'hôpital et qui peuvent influencer les résultats chirurgicaux (Table 6). Au niveau de la sélection de la population, nous avons utilisé des critères de sélection au niveau du patient, du séjour et de l'hôpital afin d'homogénéiser la population et de limiter ainsi les facteurs confondants. Au niveau de l'analyse statistique, nous avons utilisé un score de propension pour appairer les patients, basé sur les caractéristiques des patients. Nous avons également utilisé deux variables proxies, le revenu médian de la commune de résidence du patient et l'année de la chirurgie.

Table 6 – Prise en compte des facteurs confondants dans l'étude 3

Design	Constitution de deux cohortes historiques exposé-non exposé : sleeve vs bypass et sleeve vs anneau Appariement des patients au sein de chaque cohorte au sein d'un même hôpital
Sélection population	Niveau du patient (wash-out de 5 ans pour identifier la 1 ^{ère} procédure de chirurgie bariatrique) Niveau séjour (patient âgé de 18 à 110 ans, exclusion des séjours avec les 2 procédures étudiées) Niveau hôpital (Métropole, établissement ne réalisant pas les 2 procédures étudiées)
Analyses statistiques	PS basés sur les caractéristiques des patients (sexe, âge, IMC, revenu médian, sélection de comorbidités de l'Elixhauser, année de la chirurgie) Modèle de survie tenant compte de l'appariement et du décès comme risque compétitif

Dans un deuxième temps, nous avons souhaité améliorer le contrôle du biais d'indication lors de la comparaison de résultats cliniques entre deux interventions en ne se limitant pas aux caractéristiques patients mesurées dans la base mais en tenant compte également de l'ensemble des codes renseignés, qui indirectement nous permettaient de tenir compte de facteurs confondants non mesurés.

Au vu des résultats divergents entre les essais cliniques et les études basées sur l'appariement par PS pour le TAVI, nous avons choisi de comparer le risque de décès, de réhospitalisation et d'AVC dans les trois ans suivant l'intervention, pour des patients opérés en France, entre 2013 et 2017 d'un TAVI ou d'un SAVR (Etude 4). Nous avons donc repris la méthodologie de l'étude précédente et contrôlé le biais d'indication soit par la méthode classique des PS considérant

les caractéristiques patients soit par la méthode des hdPS considérant les mêmes caractéristiques patients ainsi que 300 variables additionnelles issues des codes d'actes et de diagnostics de l'année précédant l'hospitalisation (Table 7). Nous n'avons finalement pas étudié les codes en rapport avec les molécules onéreuses et les dispositifs médicaux car les effectifs étaient trop faibles dans la population étudiée. Nous avons également ajouté une variable proxy supplémentaire, le nombre de jours d'hospitalisation l'année précédente car les variables mesurant la consommation de soins de santé sont intéressantes pour prédire les résultats cliniques [146].

Table 7 - Prise en compte des facteurs confondants dans l'étude 4.

Design	Cohorte historique exposé- non exposé Appariement des patients entre les deux interventions au sein des hôpitaux
Sélection population	Niveau patient (Identifier la 1 ^{ère} intervention) Niveau séjour (patient ≥18 ans présentant un diagnostic principal d'insuffisance cardiaque, maladie de la valve aortique rhumatismale ou non, endocardite infectieuse) Niveau hôpital (Métropole, établissement ne réalisant pas les 2 interventions étudiées)
Analyses statistiques	Appariement par: <ul style="list-style-type: none"> - PS basé sur les caractéristiques des patients (sexe, âge, sélection de comorbidités de l'Elixhauser/Charlson, année, urgence, stimulateur et implant cardiaque) et des proxies (revenu médian, nombre de jours d'hospitalisation l'année précédente) - HDPS basé sur les caractéristiques patients + codes actes et diagnostics, dispositifs médicaux, molécules onéreuses sur l'année antérieure - Modèle de survie tenant compte de l'appariement et du décès comme risque compétitif selon le critère de jugement
Analyses de sensibilité	- Nombre de variables incluses dans le hdPS - Modèle de survie sans tenir compte du risque compétitif

Avec la méthode PS, on retrouvait un sur-risque de décès, de reprise chirurgicale et d'AVC à long terme pour les patients TAVI. Avec la méthode hdPS, on retrouvait uniquement un sur-risque de décès à long terme. Ces estimations étaient quand même plus faibles qu'avec la méthode PS et se rapprochaient de celles des essais cliniques contrôlés randomisés. Nous avons donc montré que la méthode hdPS permettait de mieux contrôler le biais d'indication dans les études de comparaison de résultats chirurgicaux menées à partir des bases hospitalières nationales. Néanmoins, nous n'avons pas pu étudier si la persistance de la

différence pour le décès était liée au fait que les populations étudiées étaient différentes entre les deux types d'étude ou bien au fait qu'il existait des facteurs confondants résiduels. En effet, les patients TAVI inclus dans les essais répondaient à des critères de sélection plus stricts que les patients TAVI de notre étude. Néanmoins, l'absence de données cliniques ne nous a pas permis de reproduire la population de l'essai dans notre étude pour s'en assurer. Nous ne pouvons donc pas écarter le fait que des facteurs confondants non mesurés persistaient.

Dans ces deux travaux, l'utilisation de la base PMSI nationale nous a permis de considérer tous les patients opérés d'une des chirurgies étudiées en France, sur les périodes de temps considérées, et d'évaluer la qualité de ces chirurgies en vie réelle. Elle nous a également permis d'avoir des effectifs suffisants pour étudier les complications de ces chirurgies, notamment pour la chirurgie bariatrique pour laquelle les complications sont relativement rares [147,148]. Les résultats chirurgicaux étudiés étaient facilement identifiables dans la base PMSI à partir des codes renseignés durant les diverses hospitalisations des patients, que ce soit dans l'établissement dans lequel ils ont eu leur intervention index ou dans tout autre établissement.

Nous avons retrouvé les limites classiques inhérentes aux bases de données hospitalières à savoir l'absence de certaines variables et le manque de granularité des données. Par exemple, lors de la comparaison des procédures bariatriques, il aurait été intéressant d'étudier la perte de poids ou des résultats métaboliques tels que le diabète ou les carences nutritionnelles, pour évaluer la balance bénéfique/risque de la procédure pour le patient. Egalement, lors de la comparaison du TAVI et du SAVR, il aurait été intéressant d'avoir l'euro-SCORE pour mesurer précisément le haut risque chirurgical des patients.

B. Perspectives

Dans ces différents travaux, nous avons proposé des stratégies pour contrôler de potentiels facteurs confondants malgré leur absence ou leur granularité insuffisantes. Cela nous a également permis de dégager trois perspectives de recherche intéressantes : la recherche du bon groupe contrôle, l'utilisation de nouvelles méthodes pour la sélection des covariables et l'apport de données additionnelles.

1. La recherche du bon groupe contrôle

Dès la conception de l'étude, le large périmètre des données hospitalières dans le temps et l'espace, permettent d'envisager divers schémas d'étude et groupes contrôles. L'utilisation de schéma d'étude de type quasi-expérimental permet de contrôler les facteurs confondants en se rapprochant de la conception de l'essai clinique randomisé. Se pose alors la question d'identifier le bon groupe contrôle. Les études avant-après permettent de mesurer l'association entre l'effet d'une intervention et d'un résultat chirurgical au sein d'un unique groupe lorsqu'il n'y a pas d'évolution sous-jacente du résultat chirurgical au cours du temps [149]. L'ajout d'un groupe comparateur peut être une solution pour remédier à ce problème si les résultats chirurgicaux sont de même niveau entre les deux groupes avant le début de l'intervention. Sinon, un design d'étude « difference-in-difference » peut être utilisé pour comparer les résultats avant et après l'intervention entre un groupe exposé et non exposé [150,151]. Ainsi, l'utilisation d'un groupe contrôle qui a les mêmes évolutions de résultats chirurgicaux au cours du temps mais qui n'est pas exposé à l'intervention, permet de tenir compte de l'évolution naturelle du résultat chirurgical si les évolutions avant l'intervention sont les mêmes dans les deux groupes et si des phénomènes concurrents affectent les deux groupes de la même manière [152]. Néanmoins, il n'est pas toujours aisé de rentrer dans la configuration d'un de ces designs car de nombreux facteurs confondants influencent le résultat chirurgical et il est difficile de trouver les deux groupes qui ne diffèrent que par l'intervention. Souvent d'autres facteurs peuvent également être différents. Comment dès lors arriver à tenir compte de ces facteurs confondants surtout lorsqu'ils ne sont pas mesurés dans les bases ? Faut-il multiplier les groupes contrôles ? Si oui, faut-il prendre plusieurs groupes contrôles différents et s'assurer que les résultats soient cohérents entre les analyses de sensibilité ? Comment alors synthétiser et conclure sur le résultat de l'étude ? Nous avons

vu avec l'étude 2 qu'il n'était pas toujours aisé de dégager un message clair quand les résultats des différents groupes divergeaient. Faut-il inclure ces différents groupes dans un même design d'étude en réalisant des études type double, triple différence [153], voir plus ? S'inspirer de la méthode des méta-analyses ? Le choix du ou des bons groupes contrôles est une piste de recherche intéressante afin de tenir compte de facteurs confondants connus ou non connus dans les bases.

2. Utiliser de nouvelles approches pour la sélection des variables

Ces dernières années, le développement des méthodes d'apprentissage automatique ou *machine learning* offre une alternative intéressante aux modèles paramétriques classiques, particulièrement lorsqu'on travaille sur des grandes bases de données [154]. Bien qu'utiles dans de nombreux contextes, les modèles paramétriques font des hypothèses fortes qui ne sont pas toujours satisfaites dans la pratique. Les méthodes d'apprentissage automatique ont été développées pour surmonter ces limites en exigeant des limites moins restrictives [155]. Plusieurs de ces méthodes ont été évaluées pour modéliser les scores de propension et se sont révélées performantes dans de nombreuses situations où les hypothèses paramétriques ne sont pas satisfaites. Cependant aucun algorithme n'est optimal, et l'algorithme le plus performant varie selon les contextes d'étude et les bases de données manipulées [156]. Dans un premier temps, il pourrait être utile de tester ces méthodes sur les données françaises afin de savoir si elles sont plus performantes que les PS ou hdPS. Dans un deuxième temps et sur le même principe que les hdPS, ces méthodes de machine learning pourraient être couplées au PS afin d'optimiser la sélection automatique des variables. Elles pourraient donc être utilisées pour identifier les codes présents dans les bases et qui seraient intéressants à inclure dans un PS. En utilisant d'autres sources de données telles que les entrepôts locaux hospitaliers, certaines méthodes de machine learning pourraient permettre d'inclure également dans un PS des informations textuelles issues des comptes rendus d'hospitalisation ou des résultats d'imagerie médicale par exemple. Enfin, il pourrait être intéressant de combiner plusieurs méthodes de machine learning, comme si l'on demandait à plusieurs experts leurs avis sur les variables pertinentes à inclure, afin de tirer parti du meilleur de chacune de ces méthodes et donc peut-être, d'améliorer le contrôle de facteurs confondants potentiels.

3. Apport de données additionnelles

Une autre perspective de recherche est l'ajout de données additionnelles. Cela permettrait d'étudier des résultats chirurgicaux spécifiques à la procédure étudiée, de mieux caractériser le risque préopératoire du patient et la complexité de l'intervention chirurgicale. Cela permettrait également d'étudier les caractéristiques du chirurgien et du bloc opératoire dans lequel il opère ainsi que la composition de l'équipe du bloc qui l'accompagne.

Premièrement, en utilisant le Système National des Données de Santé (SNDS), le PMSI s'enrichit des données de l'Assurance Maladie et du CépiciDc (causes médicales de décès) [157]. Il est ainsi possible de capter les décès qui ont lieu à l'extérieur de l'hôpital. De plus, les données de consommation de soins de ville peuvent servir à mieux tenir compte du risque patient préopératoire en étudiant les délivrances de médicaments, les consultations et examens que le patient a eus en dehors de l'hôpital. Avec ces données on capterait peut-être mieux le haut risque chirurgical des patients TAVI, par exemple. Ces données peuvent permettre également d'étudier les déterminants de la prise en charge chirurgicale en ne se limitant plus à ce qui passe à l'hôpital mais en explorant également sa prise en charge lorsque le patient est sorti de l'hôpital, ce qui peut potentiellement influencer le résultat chirurgical si l'on étudie les réhospitalisations par exemple. Etudier ce qui se passe à l'extérieur de l'hôpital devient de plus en plus important avec l'essor de la chirurgie ambulatoire.

Deuxièmement, l'appariement des données à celle d'un registre (ex France-TAVI [158], SOFFCOMM [159], EPITHOR [160]) ou à une cohorte nationale (ex Constances [161]) peut permettre d'améliorer la granularité des données et de pouvoir étudier des résultats chirurgicaux spécifiques à la chirurgie étudiée ou obtenir des informations cliniques supplémentaires pour contrôler d'éventuels facteurs confondants. Ainsi, la gravité de la maladie, la complexité de l'intervention et des soins postopératoires sont autant d'exemples de détails cliniques qui peuvent être obtenus dans les registres. La production des registres et des cohortes coûte cher en raison d'une intervention humaine indispensable pour atteindre le haut niveau de qualité requis. L'appariement aux données de routines peut permettre de diminuer ces coûts en se focalisant sur la collecte des informations qui ne sont pas disponibles dans ces bases.

Troisièmement, les données des entrepôts locaux hospitaliers sont également une autre source de données qui pourrait être utilisée. De nos jours, les hôpitaux disposent d'ensembles de données massives provenant des données collectées en routine dans leur système d'information qui sont nécessaires à la réalisation et l'analyse de l'activité de l'hôpital. Ces derniers permettent d'identifier de manière exhaustive l'ensemble des patients hospitalisés et des personnels travaillant au sein de l'hôpital. Ils contiennent diverses informations en rapport par exemple avec les données médicales du patient, le planning du personnel et notamment les informations relatives au bloc opératoire, non présentes dans le PMSI. Ainsi, il serait possible d'avoir l'Euro-score pour les patients TAVI ou bien, d'autres résultats chirurgicaux comme la perte de poids ou des résultats métaboliques pour les patients bariatriques, et de savoir quels sont les chirurgiens qui ont opéré les patients et quelles sont les équipes qui se sont occupées des patients, par exemple. Ces données permettraient donc de mieux appréhender les caractéristiques liées à l'état du patient et aux soins chirurgicaux qui lui sont prodigués de par leur granularité importante. Ces données permettraient également d'explorer d'autres éléments clés liés au chirurgien [162], à l'équipe chirurgicale et aux conditions de réalisation de l'acte chirurgical [163] qui ne peuvent pas être investigués dans les bases nationales.

A l'heure actuelle, les données de santé constituent un patrimoine riche mais extrêmement fragmenté, avec des accès parfois complexes et des délais de traitement assez long. Cela devrait être facilité avec la création de la Plateforme des Données de Santé ou *Health Data Hub* (HDS) qui vise à enrichir et à valoriser le SNDS en incluant l'ensemble des données dont la collecte est soutenue par la solidarité nationale, et à mettre à disposition des capacités de stockage, de calcul et des outils sécurisés permettant d'exploiter ces données à travers un guichet unique [164]. L'objectif du HDS est de garantir l'accès aisé et unifié aux données de santé aux acteurs publics ou privés dès lors que le projet est d'intérêt général en garantissant la sécurité de ces données dans le respect de la protection des données personnelles sur l'ensemble du territoire.

Ainsi les données du SNDS, des registres, des cohortes nationales, des entrepôts de données de certains hôpitaux, des bases de données de soins de ville pourraient être disponibles au sein d'une même plateforme. Un projet ambitieux qui ouvre des perspectives de recherche très intéressantes pour les années à venir.

V. Bibliographie

- [1] Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, Uribe-Leitz T, et al. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. *Lancet* 2015;385 Suppl 2:S11. [https://doi.org/10.1016/S0140-6736\(15\)60806-6](https://doi.org/10.1016/S0140-6736(15)60806-6).
- [2] Chiffres clés de l'hospitalisation, extraites du Programme de Médicalisation des Systèmes d'Information 2018. n.d. https://www.atih.sante.fr/sites/default/files/public/content/2554/atih_chiffres_cles_mco_2018_0.pdf (accessed December 13, 2019).
- [3] World Health Organisation ; Safe Surgery Saves Lives : The Second Global Patient Safety Challenge. n.d. <https://www.who.int/patientsafety/safesurgery/en/> (accessed December 13, 2019).
- [4] Dimick JB, Greenberg CC, editors. *Success in Academic Surgery: Health Services Research*. London: Springer London; 2014. <https://doi.org/10.1007/978-1-4471-4718-3>.
- [5] Michel P. Etude Nationale sur les Evénements Indésirables graves liés aux Soins 2009 (ENEIS2): description des résultats 2009. Rapport final à la DREES(Ministère du travail, de l'emploi et de la Santé). Paris: 2011.
- [6] Moses H, Matheson DHM, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research: US and international comparisons. *JAMA* 2015;313:174–89. <https://doi.org/10.1001/jama.2014.15939>.
- [7] Lohr KN, Steinwachs DM. Health Services Research: An Evolving Definition of the Field. *Health Serv Res* 2002;37:15–7. <https://doi.org/10.1111/1475-6773.01020>.
- [8] Agency for Healthcare Research and Quality. *Agency forHealthcare Research andQuality.What is health services research?* Rockville 2002.
- [9] Phillips CD. What do you do for a living? Toward a more succinct definition of health services research. *BMC Health Services Research* 2006. <https://doi.org/10.1186/1472-6963-6-117>.
- [10] Ban KA, Bilimoria KY. Is Health Services Research Important for Surgeons? *Advances in Surgery* 2016;50:143–55. <https://doi.org/10.1016/j.yasu.2016.03.012>.

- [11] Agency for Healthcare Research and Quality. Outcomes Research: fact Sheet 2013. <https://archive.ahrq.gov/research/findings/factsheets/outcomes/outfact/outcomes-and-research.html> (accessed December 18, 2019).
- [12] Codman E. The product of a hospital. *Surg Gynecol Obstet* 1914;18:491–6.
- [13] Roemer MI, Montoya-Aguilar C. Quality assessment and assurance in primary health care. ,. Geneva: WHO Offset Publication; 1988.
- [14] OCDE, Organisation mondiale de la Santé, World Bank Group. La qualité des services de santé : Un impératif mondial en vue de la couverture santé universelle. OECD; 2019. <https://doi.org/10.1787/62f287af-fr>.
- [15] Or Z, Com-Ruelle L. La qualité des soins en France : comment la mesurer pour l’améliorer ? 2008:20.
- [16] Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988;260:1743–8. <https://doi.org/10.1001/jama.260.12.1743>.
- [17] Veillard J, Champagne F, Klazinga N, Kazandjian V, Arah OA, Guisset A-L. A performance assessment framework for hospitals: the WHO regional office for Europe PATH project. *International Journal for Quality in Health Care* 2005;17:487–96. <https://doi.org/10.1093/intqhc/mzi072>.
- [18] Construction et utilisation des indicateurs dans le domaine de la santé : principes généraux. Agence Nationale d’Accréditation et d’Évaluation en Santé; 2002.
- [19] Donabedian A. An Introduction to Quality Assurance in Health Care. Oxford, New York: Oxford University Press; 2003.
- [20] Birkmeyer JD, Dimick JB, Birkmeyer NJO. Measuring the quality of surgical care: structure, process, or outcomes?1 1No competing interests declared. *Journal of the American College of Surgeons* 2004;198:626–32. <https://doi.org/10.1016/j.jamcollsurg.2003.11.017>.
- [21] Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA* 2002;288:1987–93. <https://doi.org/10.1001/jama.288.16.1987>.
- [22] Kane RL, Shamliyan TA, Mueller C, Duval S, Wilt TJ. The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. *Med Care* 2007;45:1195–204. <https://doi.org/10.1097/MLR.0b013e3181468ca3>.
- [23] Shekelle PG. Nurse-patient ratios as a patient safety strategy: a systematic review. *Ann Intern Med* 2013;158:404–9. <https://doi.org/10.7326/0003-4819-158-5-201303051-00007>.

- [24] Neuraz A, Guérin C, Payet C, Polazzi S, Aubrun F, Dailier F, et al. Patient Mortality Is Associated With Staff Resources and Workload in the ICU: A Multicenter Observational Study. *Crit Care Med* 2015;43:1587–94. <https://doi.org/10.1097/CCM.0000000000001015>.
- [25] Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective Referral to High-Volume Hospitals: Estimating Potentially Avoidable Deaths. *JAMA* 2000;283:1159. <https://doi.org/10.1001/jama.283.9.1159>.
- [26] Gandjour A, Bannenberg A, Lauterbach KW. Threshold Volumes Associated With Higher Survival in Health Care: A Systematic Review. *Medical Care* 2003;41:1129–41. <https://doi.org/10.1097/01.MLR.0000088301.06323.CA>.
- [27] Halm EA, Lee C, Chassin MR. Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature. *Annals of Internal Medicine* 2002;137:511. <https://doi.org/10.7326/0003-4819-137-6-200209170-00012>.
- [28] Killeen SD, O’Sullivan MJ, Coffey JC, Kirwan WO, Redmond HP. Provider volume and outcomes for oncological procedures. *British Journal of Surgery* 2005;92:389–402. <https://doi.org/10.1002/bjs.4954>.
- [29] Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Serv Res* 1987;22:157–82.
- [30] Gordon TA, Bowman HM, Tielsch JM, Bass EB, Burleyson GP, Cameron JL. Statewide regionalization of pancreaticoduodenectomy and its effect on in-hospital mortality. *Ann Surg* 1998;228:71–8. <https://doi.org/10.1097/0000658-199807000-00011>.
- [31] Learn PA, Bach PB. A decade of mortality reductions in major oncologic surgery: the impact of centralization and quality improvement. *Med Care* 2010;48:1041–9. <https://doi.org/10.1097/MLR.0b013e3181f37d5f>.
- [32] Birkmeyer JD, Dimick JB, Staiger DO. Operative mortality and procedure volume as predictors of subsequent hospital performance. *Ann Surg* 2006;243:411–7. <https://doi.org/10.1097/01.sla.0000201800.45264.51>.
- [33] Reames BN, Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and operative mortality in the modern era. *Ann Surg* 2014;260:244–51. <https://doi.org/10.1097/SLA.0000000000000375>.
- [34] Bergman S, Feldman LS, Barkun JS. Evaluating surgical outcomes. *Surg Clin North Am* 2006;86:129–49, x. <https://doi.org/10.1016/j.suc.2005.10.007>.

- [35] Chow A, Mayer EK, Darzi AW, Athanasiou T. Patient-reported outcome measures: the importance of patient satisfaction in surgery. *Surgery* 2009;146:435–43. <https://doi.org/10.1016/j.surg.2009.03.019>.
- [36] Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. *Health Technol Assess* 2001;5:1–194. <https://doi.org/10.3310/hta5220>.
- [37] Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004;240:205–13. <https://doi.org/10.1097/01.sla.0000133083.54934.ae>.
- [38] Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg* 2009;250:187–96. <https://doi.org/10.1097/SLA.0b013e3181b13ca2>.
- [39] Clavien PA, Sanabria JR, Strasberg SM. Proposed classification of complications of surgery with examples of utility in cholecystectomy. *Surgery* 1992;111:518–26.
- [40] DeOliveira ML, Winter JM, Schafer M, Cunningham SC, Cameron JL, Yeo CJ, et al. Assessment of complications after pancreatic surgery: A novel grading system applied to 633 patients undergoing pancreaticoduodenectomy. *Ann Surg* 2006;244:931–7; discussion 937–939. <https://doi.org/10.1097/01.sla.0000246856.03918.9a>.
- [41] Miller MR, Elixhauser A, Zhan C, Meyer GS. Patient Safety Indicators: using administrative data to identify potential patient safety concerns. *Health Serv Res* 2001;36:110–32.
- [42] Zhan C. Administrative data based patient safety research: a critical review. *Quality and Safety in Health Care* 2003;12:58ii–63. https://doi.org/10.1136/qhc.12.suppl_2.ii58.
- [43] Winters BD, Bharmal A, Wilson RF, Zhang A, Engineer L, Defoe D, et al. Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid Hospital-acquired Conditions: A Systematic Review and Meta-Analysis. *Med Care* 2016;54:1105–11. <https://doi.org/10.1097/MLR.0000000000000550>.
- [44] Quan H, Eastwood C, Cunningham CT, Liu M, Flemons W, De Coster C, et al. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ Open* 2013;3:e003716. <https://doi.org/10.1136/bmjopen-2013-003716>.

- [45] Haute Autorité de Santé - Indicateurs de qualité et de sécurité des soins (IQSS) n.d. https://www.has-sante.fr/jcms/r_1500957/fr/indicateurs-de-qualite-et-de-securite-des-soins-iqss (accessed February 27, 2020).
- [46] Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien P-A, Reeves BC, et al. Challenges in evaluating surgical innovation. *The Lancet* 2009;374:1097–104. [https://doi.org/10.1016/S0140-6736\(09\)61086-2](https://doi.org/10.1016/S0140-6736(09)61086-2).
- [47] Prystowsky JB, Bordage G, Feinglass JM. Patient outcomes for segmental colon resection according to surgeon's training, certification, and experience. *Surgery* 2002;132:663–72. <https://doi.org/10.1067/msy.2002.127550>.
- [48] Patoir A, Payet C, Peix J-L, Colin C, Pascal L, Kraimps J-L, et al. Determinants of operative time in thyroid surgery: A prospective multicenter study of 3454 thyroidectomies. *PLOS ONE* 2017;12:e0181424. <https://doi.org/10.1371/journal.pone.0181424>.
- [49] Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991;78:355–60. <https://doi.org/10.1002/bjs.1800780327>.
- [50] DICOM_Gabriel.DS, DICOM_Gabriel.DS. Case-mix. Ministère des Solidarités et de la Santé 2020. <https://solidarites-sante.gouv.fr/professionnels/gerer-un-etablissement-de-sante-medico-social/financement/financement-des-etablissements-de-sante-10795/financement-des-etablissements-de-sante-glossaire/article/case-mix> (accessed February 27, 2020).
- [51] Ultee KHJ, Tjeertes EKM, Gonçalves FB, Rouwet EV, Hoofwijk AGM, Stolker RJ, et al. The relation between household income and surgical outcome in the Dutch setting of equal access to and provision of healthcare. *PLOS ONE* 2018;13:e0191464. <https://doi.org/10.1371/journal.pone.0191464>.
- [52] Glance LG, Kellermann AL, Osler TM, Li Y, Li W, Dick AW. Impact of Risk Adjustment for Socioeconomic Status on Risk-adjusted Surgical Readmission Rates. *Ann Surg* 2016;263:698–704. <https://doi.org/10.1097/SLA.0000000000001363>.
- [53] Maruthappu M, Gilbert BJ, El-Harasis MA, Nagendran M, McCulloch P, Duclos A, et al. The influence of volume and experience on individual surgical performance: a systematic review. *Ann Surg* 2015;261:642–7. <https://doi.org/10.1097/SLA.0000000000000852>.
- [54] Duclos A, Peix J-L, Colin C, Kraimps J-L, Menegaux F, Pattou F, et al. Influence of experience on performance of individual surgeons in thyroid surgery: prospective cross sectional multicentre study. *BMJ* 2012;344:d8041. <https://doi.org/10.1136/bmj.d8041>.

- [55] Ericsson KA. The road to excellence: the acquisition of expert performance in the arts and sciences, sports, and games. Lawrence Erlbaum Associates; 1996.
- [56] Bogdanovic J, Perry J, Guggenheim M, Manser T. Adaptive coordination in surgical teams: an interview study. *BMC Health Serv Res* 2015;15:128. <https://doi.org/10.1186/s12913-015-0792-5>.
- [57] Maruthappu M, Duclos A, Zhou CD, Lipsitz SR, Wright J, Orgill D, et al. The impact of team familiarity and surgical experience on operative efficiency: a retrospective analysis. *J R Soc Med* 2016;109:147–53. <https://doi.org/10.1177/0141076816634317>.
- [58] Xu R, Carty MJ, Orgill DP, Lipsitz SR, Duclos A. The teaming curve: a longitudinal study of the influence of surgical team familiarity on operative time. *Ann Surg* 2013;258:953–7. <https://doi.org/10.1097/SLA.0b013e3182864ffe>.
- [59] Elbardissi AW, Duclos A, Rawn JD, Orgill DP, Carty MJ. Cumulative team experience matters more than individual surgeon experience in cardiac surgery. *J Thorac Cardiovasc Surg* 2013;145:328–33. <https://doi.org/10.1016/j.jtcvs.2012.09.022>.
- [60] Gonzalez AA, Dimick JB, Birkmeyer JD, Ghaferi AA. Understanding the volume-outcome effect in cardiovascular surgery: the role of failure to rescue. *JAMA Surg* 2014;149:119–23. <https://doi.org/10.1001/jamasurg.2013.3649>.
- [61] Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997;127:666–74. https://doi.org/10.7326/0003-4819-127-8_part_2-199710151-00048.
- [62] Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. *Med Care* 2004;42:1066–72. <https://doi.org/10.1097/00005650-200411000-00005>.
- [63] Mohammed MA, Deeks JJ, Girling A, Rudge G, Carmalt M, Stevens AJ, et al. Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals. *BMJ* 2009;338. <https://doi.org/10.1136/bmj.b780>.
- [64] Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding Control in Healthcare Database Research: Challenges and Potential Approaches. *Medical Care* 2010;48:S114–20. <https://doi.org/10.1097/MLR.0b013e3181d8bebe3>.
- [65] Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147–54. [https://doi.org/10.1016/S0140-6736\(04\)15901-1](https://doi.org/10.1016/S0140-6736(04)15901-1).

- [66] Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;365:1159–62. [https://doi.org/10.1016/S0140-6736\(05\)71879-1](https://doi.org/10.1016/S0140-6736(05)71879-1).
- [67] Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA* 2003;290:1062–70. <https://doi.org/10.1001/jama.290.8.1062>.
- [68] Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Balliol Collaboration, et al. Evaluation and stages of surgical innovations. *Lancet* 2009;374:1089–96. [https://doi.org/10.1016/S0140-6736\(09\)61083-7](https://doi.org/10.1016/S0140-6736(09)61083-7).
- [69] Yu J, Chen W, Chen S, Jia P, Su G, Li Y, et al. Design, Conduct, and Analysis of Surgical Randomized Controlled Trials: A Cross-sectional Survey. *Ann Surg* 2019;270:1065–9. <https://doi.org/10.1097/SLA.0000000000002860>.
- [70] Stulberg JJ, Haut ER. Practical Guide to Surgical Data Sets: Healthcare Cost and Utilization Project National Inpatient Sample (NIS). *JAMA Surg* 2018;153:586–7. <https://doi.org/10.1001/jamasurg.2018.0542>.
- [71] HCUP-US Home Page n.d. <https://www.hcup-us.ahrq.gov/> (accessed February 28, 2020).
- [72] Hospital Episode Statistics: GDPR information. NHS Digital n.d. <https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/gdpr/gdpr-register/hospital-episode-statistics-gdpr/hospital-episode-statistics-hes-gdpr-information> (accessed February 28, 2020).
- [73] Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. *Med Care* 1980;18:iii, 1–53.
- [74] Fetter RB, Freeman JL. Diagnosis related groups: product line management within hospitals. *Acad Manage Rev* 1986;11:41–54.
- [75] Article L6113-7. n.d.
- [76] Article L6113-8. n.d.
- [77] Statistiques MCO, Publication ATIH n.d. <https://www.atih.sante.fr/statistiques/mco> (accessed January 13, 2020).
- [78] Guide Méthodologique MCO 2020 | Publication ATIH n.d. <https://www.atih.sante.fr/guide-methodologique-mco-2020> (accessed February 28, 2020).

- [79] CIM-10 FR 2020 à usage PMSI | Publication ATIH n.d. <https://www.atih.sante.fr/cim-10-fr-2020-usage-pmsi> (accessed February 28, 2020).
- [80] WHO | International Classification of Diseases, 10th Revision (ICD-10). WHO n.d. <http://www.who.int/classifications/icd/en/> (accessed February 28, 2020).
- [81] CCAM en ligne - CCAM n.d. <https://www.ameli.fr/accueil-de-la-ccam/index.php> (accessed February 28, 2020).
- [82] ATIH : Agence technique de l'information sur l'hospitalisation n.d. <https://www.atih.sante.fr/> (accessed February 28, 2020).
- [83] Kyriacou DN, Lewis RJ. Confounding by Indication in Clinical Research. *JAMA* 2016;316:1818–9. <https://doi.org/10.1001/jama.2016.16435>.
- [84] Goldschlager T, Selvanathan S, Walker DG. Can a “novice” do aneurysm surgery? Surgical outcomes in a low-volume, non-subspecialised neurosurgical unit. *J Clin Neurosci* 2007;14:1055–61. <https://doi.org/10.1016/j.jocn.2006.12.002>.
- [85] Patorno E, Grotta A, Bellocco R, Schneeweiss S. Propensity score methodology for confounding control in health care utilization databases. *Epidemiology, Biostatistics and Public Health* 2013;10. <https://doi.org/10.2427/8940>.
- [86] Derogar M, Blomberg J, Sadr-Azodi O. Hospital teaching status and volume related to mortality after pancreatic cancer surgery in a national cohort. *BJS (British Journal of Surgery)* 2015;102:548–57. <https://doi.org/10.1002/bjs.9754>.
- [87] Bouyer J. *Épidémiologie: principes et méthodes quantitatives*. Lavoisier; 2009.
- [88] *Generalized, Linear, and Mixed Models, 2nd Edition* | Wiley. WileyCom n.d. <https://www.wiley.com/en-us/Generalized%2C+Linear%2C+and+Mixed+Models%2C+2nd+Edition-p-9780470073711> (accessed March 19, 2020).
- [89] Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 1983;70:41–55.
- [90] Joffe MM, Rosenbaum PR. Invited Commentary: Propensity Scores. *Am J Epidemiol* 1999;150:327–33. <https://doi.org/10.1093/oxfordjournals.aje.a010011>.
- [91] Westreich D, Lessler J, Funk MJ. Propensity score estimation: machine learning and classification methods as alternatives to logistic regression. *J Clin Epidemiol* 2010;63:826–33. <https://doi.org/10.1016/j.jclinepi.2009.11.020>.

- [92] Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol* 2011;174:1213–22. <https://doi.org/10.1093/aje/kwr364>.
- [93] Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127:757–63. https://doi.org/10.7326/0003-4819-127-8_part_2-199710151-00064.
- [94] Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006;163:1149–56. <https://doi.org/10.1093/aje/kwj149>.
- [95] Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. *Stat Methods Med Res* 2012;21:273–93. <https://doi.org/10.1177/0962280210394483>.
- [96] Rosenbaum PR, Rubin DB. The bias due to incomplete matching. *Biometrics* 1985;41:103–16.
- [97] Ming K, Rosenbaum PR. Substantial gains in bias reduction from matching with a variable number of controls. *Biometrics* 2000;56:118–24. <https://doi.org/10.1111/j.0006-341x.2000.00118.x>.
- [98] Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci* 2010;25:1–21. <https://doi.org/10.1214/09-STS313>.
- [99] Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics* 2011;10:150–61. <https://doi.org/10.1002/pst.433>.
- [100] Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research* 2011;46:399–424. <https://doi.org/10.1080/00273171.2011.568786>.
- [101] Cochran WG. The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. *Biometrics* 1968;24:295–313. <https://doi.org/10.2307/2528036>.
- [102] Rosenbaum PR, Rubin DB. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association* 1984;79:516–24. <https://doi.org/10.2307/2288398>.
- [103] Linden A. A comparison of approaches for stratifying on the propensity score to reduce bias. *J Eval Clin Pract* 2017;23:690–6. <https://doi.org/10.1111/jep.12701>.

- [104] Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015;34:3661–79. <https://doi.org/10.1002/sim.6607>.
- [105] Austin PC. The performance of different propensity-score methods for estimating relative risks. *Journal of Clinical Epidemiology* 2008;61:537–45. <https://doi.org/10.1016/j.jclinepi.2007.07.011>.
- [106] Ali MS, Groenwold RHH, Belitser SV, Pestman WR, Hoes AW, Roes KCB, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of Clinical Epidemiology* 2015;68:122–31. <https://doi.org/10.1016/j.jclinepi.2014.08.011>.
- [107] Flury BK, Riedwyl H. Standard Distance in Univariate and Multivariate Analysis. *The American Statistician* 1986;40:249–51. <https://doi.org/10.2307/2684560>.
- [108] Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine* 2009;28:3083–107. <https://doi.org/10.1002/sim.3697>.
- [109] Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol* 2006;59:437–47. <https://doi.org/10.1016/j.jclinepi.2005.07.004>.
- [110] Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional Propensity Score Adjustment in Studies of Treatment Effects Using Health Care Claims Data. *Epidemiology* 2009;20:512–22. <https://doi.org/10.1097/EDE.0b013e3181a663cc>.
- [111] Garbe E, Kloss S, Suling M, Pigeot I, Schneeweiss S. High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications. *European Journal of Clinical Pharmacology* 2013;69:549–57. <https://doi.org/10.1007/s00228-012-1334-2>.
- [112] Cadarette SM, Ban JK, Consiglio GP, Black CD, Dubins D, Marin A, et al. Diffusion of Innovations model helps interpret the comparative uptake of two methodological innovations: co-authorship network analysis and recommendations for the integration of novel methods in practice. *Journal of Clinical Epidemiology* 2017;84:150–60. <https://doi.org/10.1016/j.jclinepi.2016.12.006>.

- [113] Polinski JM, Schneeweiss S, Glynn RJ, Lii J, Rassen JA. Confronting “confounding by health system use” in Medicare Part D: comparative effectiveness of propensity score approaches to confounding adjustment: COMPARATIVE EFFECTIVENESS OF PS APPROACHES. *Pharmacoepidemiology and Drug Safety* 2012;21:90–8. <https://doi.org/10.1002/pds.3250>.
- [114] Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate Selection in High-Dimensional Propensity Score Analyses of Treatment Effects in Small Samples. *American Journal of Epidemiology* 2011;173:1404–13. <https://doi.org/10.1093/aje/kwr001>.
- [115] Guertin JR, Rahme E, Dormuth CR, LeLorier J. Head to head comparison of the propensity score and the high-dimensional propensity score matching methods. *BMC Medical Research Methodology* 2016;16. <https://doi.org/10.1186/s12874-016-0119-1>.
- [116] Schneeweiss S. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clinical Epidemiology* 2018;Volume 10:771–88. <https://doi.org/10.2147/CLEP.S166545>.
- [117] Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of Clinical Epidemiology* 2005;58:323–37. <https://doi.org/10.1016/j.jclinepi.2004.10.012>.
- [118] Bross IDJ. Spurious effects from an extraneous variable. *Journal of Chronic Diseases* 1966;19:637–47. [https://doi.org/10.1016/0021-9681\(66\)90062-2](https://doi.org/10.1016/0021-9681(66)90062-2).
- [119] Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010;19:537–54. <https://doi.org/10.1002/pds.1908>.
- [120] Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 1996;91:444–55. <https://doi.org/10.1080/01621459.1996.10476902>.
- [121] Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;29:722–9. <https://doi.org/10.1093/ije/29.4.722>.
- [122] Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology* 2006;17:260–7. <https://doi.org/10.1097/01.ede.0000215160.88317.cb>.
- [123] Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health* 1998;19:17–34. <https://doi.org/10.1146/annurev.publhealth.19.1.17>.

- [124] Bagiella E, Karamlou T, Chang H, Spivack J. Instrumental variable methods in clinical research. *The Journal of Thoracic and Cardiovascular Surgery* 2015;150:779–82. <https://doi.org/10.1016/j.jtcvs.2015.07.056>.
- [125] McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994;272:859–66.
- [126] Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods. *JAMA* 2007;297:278–85. <https://doi.org/10.1001/jama.297.3.278>.
- [127] Basu A. Estimating Decision-Relevant Comparative Effects Using Instrumental Variables. *Stat Biosci* 2011;3:6–27. <https://doi.org/10.1007/s12561-011-9033-6>.
- [128] Ho V, Hamilton BH, Roos LL. Multiple approaches to assessing the effects of delays for hip fracture patients in the United States and Canada. *Health Serv Res* 2000;34:1499–518.
- [129] Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007;334:1044. <https://doi.org/10.1136/bmj.39168.496366.55>.
- [130] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009;151:W65-94. <https://doi.org/10.7326/0003-4819-151-4-200908180-00136>.
- [131] Lillemoe KD. Surgical Volume/Outcome Debate. *Annals of Surgery* 2017;265:270. <https://doi.org/10.1097/SLA.0000000000002110>.
- [132] Jha AK. Back to the Future: Volume as a Quality Metric. *JAMA* 2015;314:214–5. <https://doi.org/10.1001/jama.2015.7580>.
- [133] Flood AB, Scott WR, Ewy W. Does practice make perfect? Part I: The relation between hospital volume and outcomes for selected diagnostic categories. *Med Care* 1984;22:98–114.
- [134] Flood AB, Scott WR, Ewy W. Does practice make perfect? Part II: The relation between volume and outcomes and other hospital characteristics. *Med Care* 1984;22:115–25.
- [135] de Cruppé W, Ohmann C, Blum K, Geraedts M. Evaluating compulsory minimum volume standards in Germany: how many hospitals were compliant in 2004? *BMC Health Serv Res* 2007;7:165. <https://doi.org/10.1186/1472-6963-7-165>.

- [136] Caiazzo R, Pattou F. Anneau, bypass ou sleeve : que choisir ? *Journal de Chirurgie Viscérale* 2013;150:104–15. <https://doi.org/10.1016/j.jchirv.2012.12.001>.
- [137] Dimick JB, Nicholas LH, Ryan AM, Thumma JR, Birkmeyer JD. Bariatric surgery complications before vs after implementation of a national policy restricting coverage to centers of excellence. *JAMA* 2013;309:792–9. <https://doi.org/10.1001/jama.2013.755>.
- [138] Krell RW, Finks JF, English WJ, Dimick JB. Profiling hospitals on bariatric surgery quality: which outcomes are most reliable? *J Am Coll Surg* 2014;219:725-734.e3. <https://doi.org/10.1016/j.jamcollsurg.2014.06.006>.
- [139] Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, et al. Transcatheter Aortic-Valve Implantation for Aortic Stenosis in Patients Who Cannot Undergo Surgery. *New England Journal of Medicine* 2010;363:1597–607. <https://doi.org/10.1056/NEJMoa1008232>.
- [140] Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK, et al. Transcatheter or Surgical Aortic-Valve Replacement in Intermediate-Risk Patients. *New England Journal of Medicine* 2016;374:1609–20. <https://doi.org/10.1056/NEJMoa1514616>.
- [141] Adams DH, Popma JJ, Reardon MJ, Yakubov SJ, Coselli JS, Deeb GM, et al. Transcatheter Aortic-Valve Replacement with a Self-Expanding Prosthesis. *New England Journal of Medicine* 2014;370:1790–8. <https://doi.org/10.1056/NEJMoa1400590>.
- [142] Mack MJ, Leon MB, Thourani VH, Makkar R, Kodali SK, Russo M, et al. Transcatheter Aortic-Valve Replacement with a Balloon-Expandable Valve in Low-Risk Patients. *New England Journal of Medicine* 2019;380:1695–705. <https://doi.org/10.1056/NEJMoa1814052>.
- [143] Ando T, Takagi H, Representing ALICE (All-Literature Investigation of Cardiovascular Evidence) Group. Comparison of late mortality after transcatheter aortic valve implantation versus surgical aortic valve replacement: Insights from a meta-analysis. *Eur J Intern Med* 2017;40:43–9. <https://doi.org/10.1016/j.ejim.2017.01.023>.
- [144] Armoiry X, Obadia J-F, Pascal L, Polazzi S, Duclos A. Comparison of transcatheter versus surgical aortic valve implantation in high-risk patients: A nationwide study in France. *J Thorac Cardiovasc Surg* 2018;156:1017-1025.e4. <https://doi.org/10.1016/j.jtcvs.2018.02.092>.
- [145] Roques F, Michel P, Goldstone AR, Nashef S a. M. The logistic EuroSCORE. *Eur Heart J* 2003;24:881–2. [https://doi.org/10.1016/s0195-668x\(02\)00799-6](https://doi.org/10.1016/s0195-668x(02)00799-6).
- [146] Schneeweiss S. Performance of Comorbidity Scores to Control for Confounding in Epidemiologic Studies using Claims Data. *American Journal of Epidemiology* 2001;154:854–64. <https://doi.org/10.1093/aje/154.9.854>.

- [147] Ibrahim AM, Dimick JB. What Metrics Accurately Reflect Surgical Quality? *Annu Rev Med* 2018;69:481–91. <https://doi.org/10.1146/annurev-med-060116-022805>.
- [148] Peterli R, Wölnerhanssen BK, Peters T, Vetter D, Kröll D, Borbély Y, et al. Effect of Laparoscopic Sleeve Gastrectomy vs Laparoscopic Roux-en-Y Gastric Bypass on Weight Loss in Patients With Morbid Obesity: The SM-BOSS Randomized Clinical Trial. *JAMA* 2018;319:255–65. <https://doi.org/10.1001/jama.2017.20897>.
- [149] Chen Y-F, Hemming K, Stevens AJ, Lilford RJ. Secular trends and evaluation of complex interventions: the rising tide phenomenon. *BMJ Quality & Safety* 2016;25:303–10. <https://doi.org/10.1136/bmjqs-2015-004372>.
- [150] Dimick JB, Ryan AM. Methods for Evaluating Changes in Health Care Policy: The Difference-in-Differences Approach. *JAMA* 2014;312:2401–2. <https://doi.org/10.1001/jama.2014.16153>.
- [151] Sanchez S, Payet C, Lifante J-C, Polazzi S, Chollet F, Carty MJ, et al. Surgical risks associated with winter sport tourism. *PLoS ONE* 2015;10:e0124644. <https://doi.org/10.1371/journal.pone.0124644>.
- [152] Ryan AM, Burgess JF, Dimick JB. Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences. *Health Serv Res* 2015;50:1211–35. <https://doi.org/10.1111/1475-6773.12270>.
- [153] Kristensen SR, Meacock R, Turner AJ, Boaden R, McDonald R, Roland M, et al. Long-Term Effect of Hospital Pay for Performance on Mortality in England. *New England Journal of Medicine* 2014;371:540–8. <https://doi.org/10.1056/NEJMoa1400962>.
- [154] Breiman L. Statistical modeling: The two cultures. *Statistical Science* 2001.
- [155] The Elements of Statistical Learning: Trevor Hastie: 9780387848570 n.d. <https://www.bookdepository.com/Elements-Statistical-Learning-Trevor-Hastie/9780387848570> (accessed February 28, 2020).
- [156] Wyss R, Schneeweiss S, van der Laan M, Lendle SD, Ju C, Franklin JM. Using Super Learner Prediction Modeling to Improve High-dimensional Propensity Score Estimation. *Epidemiology* 2018;29:96–106. <https://doi.org/10.1097/EDE.0000000000000762>.
- [157] Accueil | SNDS n.d. <https://www.snds.gouv.fr/SNDS/Accueil> (accessed February 28, 2020).

- [158] FRANCE TAVI - Registre des bioprothèses valvulaires aortiques implantées par cathéter | SFCardio n.d. <https://www.sfcario.fr/recherche/france-tavi-registre-des-bioprotheses-valvulaires-aortiques-implantees-par-catheter> (accessed February 28, 2020).
- [159] SOFFCOMM : Chirurgie de l'Obésité et des Maladies Métaboliques n.d. <https://soffcomm.org/> (accessed February 28, 2020).
- [160] S.F.C.T.C.V. Société Française de Chirurgie Thoracique et Cardio-Vasculaire. Société Française de Chirurgie Thoracique et Cardio-Vasculaire n.d. <https://www.sfctcv.org> (accessed March 19, 2020).
- [161] Constances | Améliorer la santé de demain n.d. <https://www.constances.fr/> (accessed February 28, 2020).
- [162] Carter D. The surgeon as a risk factor. *BMJ* 2003;326:832–3. <https://doi.org/10.1136/bmj.326.7394.832>.
- [163] Vincent C, Moorthy K, Sarker SK, Chang A, Darzi AW. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg* 2004;239:475–82. <https://doi.org/10.1097/01.sla.0000118753.22830.41>.
- [164] Health data hub | Plateforme Des Données De Santé | France. Healthdatahub n.d. <https://www.health-data-hub.fr> (accessed February 28, 2020).

VI. Annexes

A. Annexes des articles présentés dans ce travail

1. Annexes de l'article 1

S1 - Search strategy via PubMed

((surgery[Title/Abstract] OR surgical[Title/Abstract]) AND
(mortality[Title/Abstract] OR death[Title/Abstract] OR survival[Title/Abstract]) AND
("hospital episode statistics"[Title/Abstract] OR medicare[Title/Abstract] OR "nationwide
inpatient sample" [Title/Abstract] OR administrative data*[Title/Abstract] OR administrative
record*[Title/Abstract] OR discharge data*[Title/Abstract] OR discharge record*
[Title/Abstract] OR discharge abstract* [Title/Abstract] OR routine data* [Title/Abstract] OR
hospital data*[Title/Abstract] OR hospital record* [Title/Abstract]) AND
(NEJM[journal] OR Lancet[journal] OR bmj[journal] OR jama[journal] OR Ann
Surg[journal] OR Med care[journal]))
NOT letter[publication type] NOT comment[publication type] NOT review[publication type]
NOT Lancet oncol[journal]

In the final search strategy MeSH terms were not used. They exist only for mortality but their use decreased by half the number of eligible studies.

S2 – Studies selected for review

Showstack, J.A., et al., *Association of volume with outcome of coronary artery bypass graft surgery: Scheduled vs nonscheduled operations*. JAMA, 1987. **257**(6): p. 785-789.
Silber, J.H., et al., *Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue*. Medical care, 1992. **30**(7): p. 616-629.
Luft, H.S. and P.S. Romano, *Chance, continuity, and change in hospital mortality rates: Coronary artery bypass graft patients in california hospitals, 1983 to 1989*. JAMA, 1993. **270**(3): p. 331-337.
Lu-Yao, G.L., et al., *An assessment of radical prostatectomy: Time trends, geographic variation, and outcomes*. JAMA, 1993. **269**(20): p. 2633-2636.
Stremple, J.F., et al., *Comparison of postoperative mortality in VA and private hospitals*. Annals of Surgery, 1993. **217**(3): p. 277-285.
Steiner, C.A., et al., *Surgical Rates and Operative Mortality for Open and Laparoscopic Cholecystectomy in Maryland*. New England Journal of Medicine, 1994. **330**(6): p. 403-408.
Escarce, J.J., W. Chen, and J. Schwartz, *Falling cholecystectomy thresholds since the introduction of laparoscopic cholecystectomy*. JAMA, 1995. **273**(20): p. 1581-1585.

- Gordon, T.A., et al., *The effects of regionalization on cost and outcome for one general high-risk surgical procedure*. *Annals of Surgery*, 1995. **221**(1): p. 43-49.
- Grumbach, K., et al., *Regionalization of cardiac surgery in the united states and canada: Geographic access, choice, and outcomes*. *JAMA*, 1995. **274**(16): p. 1282-1288.
- Wen, S., R. Hernandez, and C. Naylor, *Pitfalls in nonrandomized outcomes studies: The case of incidental appendectomy with open cholecystectomy*. *JAMA*, 1995. **274**(21): p. 1687-1691.
- ROOS, L.L., et al., *Short-Term Mortality After Repair of Hip Fracture: Do Manitoba Elderly Do Worse?* *Medical care*, 1996. **34**(4): p. 310-326.
- Weber, A.M. and J.-C. Lee, *Use of Alternative Techniques of Hysterectomy in Ohio, 1988–1994*. *New England Journal of Medicine*, 1996. **335**(7): p. 483-489.
- Ghali, W.A., et al., *Statewide quality improvement initiatives and mortality after cardiac surgery*. *JAMA*, 1997. **277**(5): p. 379-382.
- Cebul, R.D., et al., *Indications, outcomes, and provider volumes for carotid endarterectomy*. *JAMA*, 1998. **279**(16): p. 1282-1287.
- Gordon, T.A., et al., *Statewide regionalization of pancreaticoduodenectomy and its effect on in-hospital mortality*. *Annals of Surgery*, 1998. **228**(1): p. 71-78.
- Sosa, J.A., et al., *Importance of hospital volume in the overall management of pancreatic cancer*. *Annals of Surgery*, 1998. **228**(3): p. 429-438.
- Wennberg, D.E., et al., *Variation in carotid endarterectomy mortality in the medicare population: Trial hospitals, volume, and patient characteristics*. *JAMA*, 1998. **279**(16): p. 1278-1281.
- Wilt, T.J., et al., *An Evaluation of Radical Prostatectomy at Veterans Affairs Medical Centers: Time Trends and Geographic Variation in Utilization and Outcomes*. *Medical care*, 1999. **37**(10): p. 1046-1056.
- McGrath, P.D., et al., *Relation between operator and hospital volume and outcomes following percutaneous coronary interventions in the era of the coronary stent*. *JAMA*, 2000. **284**(24): p. 3139-3144.
- Birkmeyer, J.D., et al., *Hospital Volume and Surgical Mortality in the United States*. *New England Journal of Medicine*, 2002. **346**(15): p. 1128-1137.
- Dimick, J.B., et al., *Variation in Death Rate After Abdominal Aortic Aneurysmectomy in the United States: Impact of Hospital Volume, Gender, and Age*. *Annals of Surgery*, 2002. **235**(4): p. 579-585.
- Vaughan-Sarrazin, M.S., et al., *Mortality in medicare beneficiaries following coronary artery bypass graft surgery in states with and without certificate of need regulation*. *JAMA*, 2002. **288**(15): p. 1859-1866.
- Birkmeyer, J.D., et al., *Surgeon Volume and Operative Mortality in the United States*. *New England Journal of Medicine*, 2003. **349**(22): p. 2117-2127.
- Christian, C.K., et al., *The Leapfrog Volume Criteria May Fall Short in Identifying High-Quality Surgical Centers*. *Annals of Surgery*, 2003. **238**(4): p. 447-457.
- Hannan, E.L., et al., *Provider Profiling and Quality Improvement Efforts in Coronary Artery Bypass Graft Surgery: The Effect on Short-Term Mortality Among Medicare Beneficiaries*. *Medical care*, 2003. **41**(10): p. 1164-1172.
- Aylin, P., et al., *Paediatric cardiac surgical mortality in England after Bristol: descriptive analysis of hospital episode statistics 1991-2002*. *BMJ*, 2004. **329**(7470): p. 825.

- Urbach, D.R. and N.N. Baxter, *Does it matter what a hospital is "high volume" for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data.* BMJ, 2004.
- Flum, D.R., et al., *Early mortality among medicare beneficiaries undergoing bariatric surgical procedures.* JAMA, 2005. **294**(15): p. 1903-1908.
- Fong, Y., et al., *Long-Term Survival Is Superior After Resection for Cancer in High-Volume Centers.* Annals of Surgery, 2005. **242**(4): p. 540-547.
- Santry, H.P., D.L. Gillen, and D.S. Lauderdale, *Trends in bariatric surgical procedures.* JAMA, 2005. **294**(15): p. 1909-1917.
- Birkmeyer, J.D., J.B. Dimick, and D.O. Staiger, *Operative Mortality and Procedure Volume as Predictors of Subsequent Hospital Performance.* Annals of Surgery, 2006. **243**(3): p. 411-417.
- Dimick, J.B., D.O. Staiger, and J.D. Birkmeyer, *Are mortality rates for different operations related? Implications for measuring the quality of non-cardiac surgery.* Medical care, 2006. **44**(8): p. 774-778.
- Lucas, F.L., et al., *Race and Surgical Mortality in the United States.* Annals of Surgery, 2006. **243**(2): p. 281-286.
- Hwang, C.W., et al., *Comorbidity and Outcomes of Coronary Artery Bypass Graft Surgery at Cardiac Specialty Hospitals Versus General Hospitals.* Medical care, 2007. **45**(8): p. 720-728.
- McPhee, J.T., et al., *Perioperative Mortality for Pancreatectomy: A National Perspective.* Annals of Surgery, 2007. **246**(2): p. 246-253.
- Murr, M.M., et al., *A State-wide Review of Contemporary Outcomes of Gastric Bypass in Florida: Does Provider Volume Impact Outcomes?* Annals of Surgery, 2007. **245**(5): p. 699-706.
- Neighbors, C.J., et al., *Ethnic/Racial Disparities in Hospital Procedure Volume for Lung Resection for Lung Cancer.* Medical care, 2007. **45**(7): p. 655-663.
- Sosa, J.A., et al., *Racial Disparities in Clinical and Economic Outcomes From Thyroidectomy.* Annals of Surgery, 2007. **246**(6): p. 1083-1091.
- Birkmeyer, N.J.O., et al., *Socioeconomic Status and Surgical Mortality in the Elderly.* Medical care, 2008. **46**(9): p. 893-899.
- Faiz, O., et al., *Traditional and Laparoscopic Appendectomy in Adults: Outcomes in English NHS Hospitals Between 1996 and 2006.* Annals of Surgery, 2008. **248**(5): p. 800-806.
- Malenka, D.J., et al., *Outcomes Following Coronary Stenting in the Era of Bare-Metal vs the Era of Drug-Eluting Stents.* JAMA, 2008. **299**(24): p. 2868-2876.
- Schermerhorn, M.L., et al., *Endovascular vs. Open Repair of Abdominal Aortic Aneurysms in the Medicare Population.* New England Journal of Medicine, 2008. **358**(5): p. 464-474.
- Englesbe, M.J., et al., *Mortality in Medicare Patients Undergoing Surgery in July in Teaching Hospitals.* Annals of Surgery, 2009. **249**(6): p. 871-876.

- Eppsteiner, R.W., et al., *Surgeon Volume Impacts Hospital Mortality for Pancreatic Resection*. *Annals of Surgery*, 2009. **249**(4): p. 635-640.
- Farrar, S., et al., *Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis*. *BMJ*, 2009. **339**.
- Gasper, W.J., et al., *Has Recognition of the Relationship Between Mortality Rates and Hospital Volume for Major Cancer Surgery in California Made a Difference?: A Follow-Up Analysis of Another Decade*. *Annals of Surgery*, 2009. **250**(3): p. 472-483.
- Karanicolas, P.J., et al., *The More the Better?: The Impact of Surgeon and Hospital Volume on in-Hospital Mortality Following Colorectal Resection*. *Annals of Surgery*, 2009. **249**(6): p. 954-959.
- Kurichi, J.E., et al., *Possible Incremental Benefits of Specialized Rehabilitation Bed Units Among Veterans After Lower Extremity Amputation*. *Medical care*, 2009. **47**(4): p. 457-465.
- London, J.A., et al., *Lack of Insurance is Associated With Increased Risk for Hernia Complications*. *Annals of Surgery*, 2009. **250**(2): p. 331-337.
- Allareddy, V., et al., *Effect of Meeting Leapfrog Volume Thresholds on Complication Rates Following Complex Surgical Procedures*. *Annals of Surgery*, 2010. **251**(2): p. 377-383.
- Bennett, K.M., et al., *Patient Socioeconomic Status Is an Independent Predictor of Operative Mortality*. *Annals of Surgery*, 2010. **252**(3): p. 552-558.
- Burns, E.M., et al., *Introduction of laparoscopic bariatric surgery in England: observational population cohort study*. *BMJ*, 2010. **341**.
- Conrad, M.F., et al., *Management of Diseases of the Descending Thoracic Aorta in the Endovascular Era: A Medicare Population Study*. *Annals of Surgery*, 2010. **252**(4): p. 603-610.
- Deyo, R.A., et al., *Trends, Major Medical Complications, and Charges Associated with Surgery for Lumbar Spinal Stenosis in Older Adults*. *JAMA : the journal of the American Medical Association*, 2010. **303**(13): p. 1259-1265.
- Eggl, Y., et al., *Surgical Safety and Hospital Volume Across a Wide Range of Interventions*. *Medical care*, 2010. **48**(11): p. 962-971.
- Hagen, T.P., M.S. Vaughan-Sarrazin, and P. Cram, *Relation between hospital orthopaedic specialisation and outcomes in patients aged 65 and older: retrospective analysis of US Medicare data*. *BMJ*, 2010. **340**.
- LaPar, D.J., et al., *Primary Payer Status Affects Mortality for Major Surgical Operations*. *Annals of Surgery*, 2010. **252**(3): p. 544-551.
- Lazzarino, A.I., et al., *Open Versus Minimally Invasive Esophagectomy: Trends of Utilization and Associated Outcomes in England*. *Annals of Surgery*, 2010. **252**(2): p. 292-298.
- Learn, P.A. and P.B. Bach, *A Decade of Mortality Reductions in Major Oncologic Surgery: The Impact of Centralization and Quality Improvement*. *Medical care*, 2010. **48**(12): p. 1041-1049.
- Wang, Y.R., J.E. Richter, and D.T. Dempsey, *Trends and Outcomes of Hospitalizations for Peptic Ulcer Disease in the United States, 1993 to 2006*. *Annals of Surgery*, 2010. **251**(1): p. 51-58.

- Cram, P., et al., *Clinical characteristics and outcomes of Medicare patients undergoing total hip arthroplasty, 1991–2008*. JAMA : the journal of the American Medical Association, 2011. **305**(15): p. 1560-1567.
- Finks, J.F., N.H. Osborne, and J.D. Birkmeyer, *Trends in Hospital Volume and Operative Mortality for High-Risk Surgery*. New England Journal of Medicine, 2011. **364**(22): p. 2128-2137.
- Finlayson, E., et al., *Major Abdominal Surgery in Nursing Home Residents: A National Study*. Annals of Surgery, 2011. **254**(6): p. 921-926.
- Flum, D.R., et al., *The Use, Safety and Cost of Bariatric Surgery Before and After Medicare's National Coverage Decision*. Annals of Surgery, 2011. **254**(6): p. 860-865.
- Funk, L.M., et al., *Esophagectomy Outcomes at Low-Volume Hospitals: The Association Between Systems Characteristics and Mortality*. Annals of Surgery, 2011. **253**(5): p. 912-917.
- Ghaferi, A.A., J.D. Birkmeyer, and J.B. Dimick, *Hospital Volume and Failure to Rescue With High-risk Surgery*. Medical care, 2011. **49**(12): p. 1076-1081.
- Holena, D.N., et al., *Teaching Status: The Impact on Emergency and Elective Surgical Care in the US*. Annals of Surgery, 2011. **253**(5): p. 1017-1023.
- Kozower, B.D. and G.J. Stukenborg, *The Relationship Between Hospital Lung Cancer Resection Volume and Patient Mortality Risk*. Annals of Surgery, 2011. **254**(6): p. 1032-1037.
- Livhits, M., et al., *Risk of Surgery Following Recent Myocardial Infarction*. Annals of Surgery, 2011. **253**(5): p. 857-864.
- O'Malley, A.J., et al., *Improving Observational Study Estimates of Treatment Effects using Joint Modeling of Selection Effects and Outcomes: The case of AAA Repair*. Medical care, 2011. **49**(12): p. 1126-1132.
- Tiwari, M.M., et al., *Comparison of Outcomes of Laparoscopic and Open Appendectomy in Management of Uncomplicated and Complicated Appendicitis*. Annals of Surgery, 2011. **254**(6): p. 927-932.
- Birkmeyer, J.D., et al., *Hospital quality and the cost of inpatient surgery in the United States*. Annals of Surgery, 2012. **255**(1): p. 1-5.
- Mamidanna, R., et al., *Short-Term Outcomes Following Open Versus Minimally Invasive Esophagectomy for Cancer in England: A Population-Based National Study*. Annals of Surgery, 2012. **255**(2): p. 197-203.
- Moxey, P.W., et al., *Volume–Outcome Relationships in Lower Extremity Arterial Bypass Surgery*. Annals of Surgery, 2012. **256**(6): p. 1102-1107.
- Sanders, R.D., et al., *Independent Preoperative Predictors of Outcomes in Orthopedic and Vascular Surgery: The Influence of Time Interval Between an Acute Coronary Syndrome or Stroke and the Operation*. Annals of Surgery, 2012. **255**(5): p. 901-907.
- Aylin, P., et al., *Day of week of procedure and 30 day mortality for elective surgery: retrospective analysis of hospital episode statistics*. BMJ, 2013. **346**.
- Barreto-Filho, J.A., et al., *Trends in Aortic Valve Replacement for Elderly Patients in the United States, 1999–2011*. JAMA : the journal of the American Medical Association, 2013. **310**(19): p. 2078-2085.

Duclos, A., et al., *Temporal Variation in Surgical Mortality Within French Hospitals*. Medical care, 2013. **51**(12): p. 1085-1093.

Jafari, M.D., et al., *A Decade Analysis of Trends and Outcomes of Partial Versus Total Esophagectomy in the United States*. Annals of Surgery, 2013. **258**(3): p. 450-458.

Kendal, A.R., et al., *Mortality rates at 10 years after metal-on-metal hip resurfacing compared with total hip replacement in England: retrospective cohort analysis of hospital episode statistics*. BMJ, 2013. **347**.

Halabi, W.J., et al., *Surgery for Gallstone Ileus: A Nationwide Comparison of Trends and Outcomes*. Annals of Surgery, 2014. **259**(2): p. 329-335.

McCrum, M.L., et al., *Beyond Volume: Does Hospital Complexity Matter?: An Analysis of Inpatient Surgical Mortality in the United States*. Medical care, 2014. **52**(3): p. 235-242.

Morton, J.M., T. Garg, and N. Nguyen, *Does Hospital Accreditation Impact Bariatric Surgery Safety?* Annals of Surgery, 2014. **260**(3): p. 504-509.

Patel, M.S., et al., *Association of the 2011 acgme resident duty hour reforms with mortality and readmissions among hospitalized medicare patients*. JAMA, 2014. **312**(22): p. 2364-2373.

Reames, B.N., et al., *Hospital Volume and Operative Mortality in the Modern Era*. Annals of Surgery, 2014. **260**(2): p. 244-251.

Shih, T., et al., *Does Pay-for-Performance Improve Surgical Outcomes? An Evaluation of Phase 2 of the Premier Hospital Quality Incentive Demonstration*. Annals of Surgery, 2014. **259**(4): p. 677-681.

Haider, A.H., et al., *Incremental Cost of Emergency Versus Elective Surgery*. Annals of Surgery, 2015. **262**(2): p. 260-266.

Munasinghe, A., et al., *Is It Time to Centralize High-risk Cancer Care in the United States? Comparison of Outcomes of Esophagectomy Between England and the United States*. Annals of Surgery, 2015. **262**(1): p. 79-85.

Reames, B.N., et al., *Evaluation of the Effectiveness of a Surgical Checklist in Medicare Patients*. Medical care, 2015. **53**(1): p. 87-94.

S3 – An potential adapted checklist based on the RECORD checklist, that should be reported in observational studies using large hospital databases

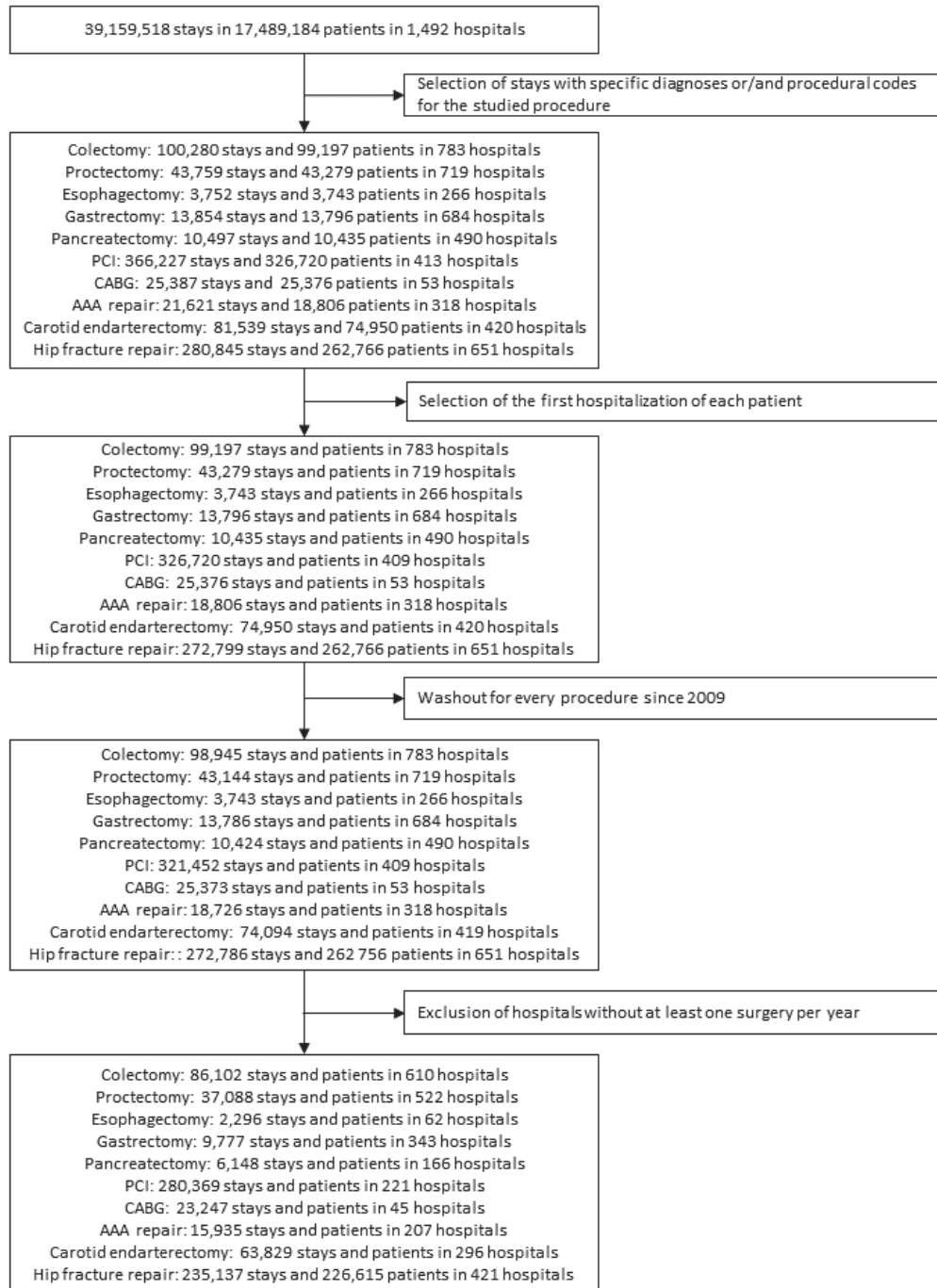
	Item No	STROBE items	RECORD items
Title and Abstract			
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.
Introduction			
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	
Objectives	3	State specific objectives, including any prespecified hypotheses	
Methods			
Study Design	4	Present key elements of study design early in the paper	How data granularity levels (e.g. stay, patient, hospital, etc.) and secular trends have been considered in study design should be described.

Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	
Participants	6	(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. <i>Case-control study</i> - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. <i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. (b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed. <i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. How study population was selected at each data granularity levels (e.g. selection criteria for stays, patients, hospitals, etc.) should be listed in detail. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided. Data granularity levels (e.g. stay, patient, hospital, etc.) for all outcomes, exposures, predictors, potential confounders and effect modifiers should be clearly defined.
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding. (b) Describe any methods used to examine subgroups and interactions. (c) Explain how missing data were addressed. (d) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed. <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed. <i>Cross-sectional study</i> - If applicable, describe analytical methods taking account of sampling strategy. (e) Describe any sensitivity analyses	How data granularity levels (e.g. stay, patient, hospital, etc.) and secular trends have been considered in statistical analysis should be described.
Data access and cleaning methods		..	RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.
Linkage		..	RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.
Results			
Participants	13	(a) Report the numbers of individuals at each stage of the study (e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed). (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	RECORD 13.1: Describe in detail the selection of the persons included in the study (i.e., study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram. The numbers of selected individuals at each data granularity levels should be reported (e.g. the number of stays, patients, hospitals, etc.).
Descriptive data	14	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders. (b) Indicate the number of participants with missing data for each variable of interest. (c) <i>Cohort study</i> - summarise follow-up time (e.g., average and total amount)	
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time. <i>Case-control study</i> - Report numbers in each exposure category, or summary measures of exposure. <i>Cross-sectional study</i> - Report numbers of outcome events or summary measures	
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included. (b) Report category boundaries when continuous variables were categorized. (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	

Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	
Generalisability	21	Discuss the generalisability (external validity) of the study results	
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	
Accessibility of protocol, raw data, and programming code		..	RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.

*Reference: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

2. Annexes de l'article 2



S

Figure E1 Study flowchart

Table E1 Hospital characteristics by procedure and trend in volume of procedures between 2010 and 2014.

	Decreased N (%)	No change N (%)	Increased N (%)	P-value
Colectomy				
Volume change per year, mean (SD)	-4.7 (1.9)	0.2 (1.4)	5.4 (2.4)	<0.001
Volume of procedures, mean (SD)	115.4 (83.9)	145.2 (86.2)	167.4 (86.3)	<0.001
Status				0.136
<i>Teaching</i>	6 (3.2%)	12 (4.3%)	12 (8.5%)	
<i>Private for profite</i>	108 (57.1%)	142 (50.9%)	69 (48.6%)	
<i>Public or private non-for-profit</i>	75 (39.7%)	125 (44.8%)	61 (43.0%)	
Specializaton degree ^a , mean (SD)	5.0 (2.2)	5.3 (2.1)	5.8 (3.0)	0.020
Attraction rate ^b , mean (SD)	12.4 (14.1)	14.6 (15.0)	14.1 (13.9)	0.180
Proctectomy				
Volume change per year, mean (SD)	-5.9 (2.4)	-0.7 (1.6)	5.0 (2.6)	<.001
Volume of procedures, mean (SD)	66.6 (50.9)	66.5 (56.1)	84.5 (51.8)	0.004
Status				0.002
<i>Teaching</i>	2 (1.6%)	19 (7.2%)	12 (9.2%)	
<i>Private for profite</i>	88 (68.8%)	133 (50.6%)	76 (58.0%)	
<i>Public or private non-for-profit</i>	38 (29.7%)	111 (42.2%)	43 (32.8%)	
Specializaton degree ^a , mean (SD)	2.5 (1.9)	2.5 (2.5)	2.6 (2.0)	0.893
Attraction rate ^b , mean (SD)	14.0 (16.2)	15.9 (17.3)	18.9 (18.2)	0.067
Esophagectomy				
Volume change per year, mean (SD)	-3.0 (2.8)	5.4 (3.1)	20.6 (6.8)	<0.001
Volume of procedures, mean (SD)	35.8 (67.6)	31.6 (31.7)	95.0 (96.3)	0.206
Status				0.759
<i>Teaching</i>	11 (31.4%)	11 (45.8%)	1 (33.3%)	
<i>Private for profite</i>	16 (45.7%)	9 (37.5%)	2 (66.7%)	
<i>Public or private non-for-profit</i>	8 (22.9%)	4 (16.7%)	0 (0.0%)	
Specializaton degree ^a , mean (SD)	0.6 (0.7)	1.0 (1.6)	1.1 (0.9)	0.479
Attraction rate ^b , mean (SD)	28.0 (23.1)	40.3 (30.1)	36.9 (34.9)	0.222
Gastrectomy				
Volume change per year, mean (SD)	2.4 (0.9)	0.4 (0.5)	-1.0 (0.6)	<0.001
Volume of procedures, mean (SD)	49.0 (27.5)	30.0 (23.0)	22.5 (18.1)	<0.001
Status				0.011
<i>Teaching</i>	8 (22.9%)	15 (9.9%)	7 (4.5%)	
<i>Private for profite</i>	15 (42.9%)	78 (51.7%)	91 (58.0%)	
<i>Public or private non-for-profit</i>	12 (34.3%)	58 (38.4%)	59 (37.6%)	
Specializaton degree ^a , mean (SD)	1.2 (1.0)	1.0 (1.1)	0.7 (0.4)	0.001
Attraction rate ^b , mean (SD)	23.0 (17.7)	17.4 (18.9)	15.0 (17.3)	0.053
Pancreatotomy				
Volume change per year, mean (SD)	-4.5 (2.0)	2.2 (1.9)	9.3 (3.4)	<0.001
Volume of procedures, mean (SD)	22.4 (29.7)	39.6 (35.0)	53.2 (48.8)	<0.001
Status				0.010
<i>Teaching</i>	3 (5.2%)	12 (18.2%)	13 (31.0%)	
<i>Private for profite</i>	35 (60.3%)	33 (50.0%)	15 (35.7%)	
<i>Public or private non-for-profit</i>	20 (34.5%)	21 (31.8%)	14 (33.3%)	
Specializaton degree ^a , mean (SD)	0.6 (0.4)	1.0 (0.8)	1.3 (1.3)	<0.001
Attraction rate ^b , mean (SD)	19.1 (22.2)	24.8 (22.0)	27.7 (22.2)	0.140
PCI				
Volume change per year, mean (SD)	-15.4 (8.5)	2.1 (4.9)	24.8 (16.6)	<0.001
Volume of procedures, mean (SD)	741.8 (860.2)	1491.8 (1117.8)	748.9 (699.5)	<0.001
Status				0.085
<i>Teaching</i>	1 (2.4%)	21 (13.5%)	1 (4.0%)	
<i>Private for profite</i>	20 (48.8%)	54 (34.8%)	7 (28.0%)	
<i>Public or private non-for-profit</i>	20 (48.8%)	80 (51.6%)	17 (68.0%)	
Specializaton degree ^a , mean (SD)	14.0 (7.8)	21.3 (10.2)	18.6 (9.4)	<0.001
Attraction rate ^b , mean (SD)	21.7 (18.5)	22.0 (16.0)	17.4 (16.2)	0.427
CABG				
Volume change per year, mean (SD)	-40.2 (11.0)	-3.3 (9.0)	29.5 (6.5)	<0.001
Volume of procedures, mean (SD)	193.6 (109.9)	683.3 (545.7)	330.8 (401.0)	0.011
Status				0.033
<i>Teaching</i>	2 (18.2%)	18 (64.3%)	2 (33.3%)	
<i>Private for profite</i>	6 (54.5%)	9 (32.1%)	3 (50.0%)	
<i>Public or private non-for-profit</i>	3 (27.3%)	1 (3.6%)	1 (16.7%)	
Specializaton degree ^a , mean (SD)	3.6 (2.6)	10.0 (8.5)	5.6 (7.1)	0.047
Attraction rate ^b , mean (SD)	50.4 (27.0)	48.7 (19.6)	50.2 (24.7)	0.972
AAA repair				
Volume change per year, mean (SD)	-10.5 (5.5)	1.9 (3.7)	15.2 (5.3)	<0.001
Volume of procedures, mean (SD)	64.3 (101.2)	72.8 (73.6)	96.6 (61.4)	0.089
Status				0.703
<i>Teaching</i>	5 (10.6%)	10 (9.3%)	8 (15.1%)	
<i>Private for profite</i>	32 (68.1%)	66 (61.7%)	31 (58.5%)	
<i>Public or private non-for-profit</i>	10 (21.3%)	31 (29.0%)	14 (26.4%)	
Specializaton degree ^a , mean (SD)	2.4 (1.7)	1.9 (1.7)	2.2 (2.1)	0.275
Attraction rate ^b , mean (SD)	22.7 (18.9)	21.1 (20.5)	31.2 (23.2)	0.016
Carotid endarterectomy				
Volume change rate per year, mean (SD)	-15.9 (6.0)	-2.6 (3.3)	8.8 (5.4)	<0.001
Volume of procedures, mean (SD)	136.2 (153.7)	243.4 (233.9)	186.0 (156.8)	0.010
Status				0.179

<i>Teaching</i>	0 (0.0%)	11 (6.2%)	8 (9.0%)	
<i>Private for profite</i>	21 (72.4%)	122 (68.5%)	50 (56.2%)	
<i>Public or private non-for-profit</i>	8 (27.6%)	45 (25.3%)	31 (34.8%)	
Specializaton degree ^a , mean (SD)	53.0 (26.1)	48.0 (28.1)	45.3 (23.1)	0.386
Attraction rate ^b , mean (SD)	20.5 (21.9)	20.9 (19.9)	18.1 (18.3)	0.536
Hip Fracture repair				
Volume change per year, mean (SD)	-12.1 (10.4)	3.6 (4.7)	30.7 (14.4)	<0.001
Volume of procedures, mean (SD)	330.3 (413.3)	640.6 (432.0)	412.3 (331.9)	<0.001
Status				<0.001
<i>Teaching</i>	5 (6.3%)	16 (5.4%)	2 (4.5%)	
<i>Private for profite</i>	43 (54.4%)	76 (25.5%)	14 (31.8%)	
<i>Public or private non-for-profit</i>	31 (39.2%)	206 (69.1%)	28 (63.6%)	
Specializaton degree ^a , mean (SD)	8.8 (8.0)	14.6 (7.5)	8.8 (6.3)	<0.001
Attraction rate ^b , mean (SD)	12.0 (13.0)	12.1 (12.8)	10.0 (9.3)	0.583

^a Proportion of stays for each studied procedure in the surgical department (expressed as a percentage)

^b Proportion of patients living in another geographical area than the one of hospital location where they underwent each studied procedure (expressed as a percentage)
 PCI percutaneous coronary intervention, CABG coronary-artery bypass grafting, AAA abdominal aortic aneurysm

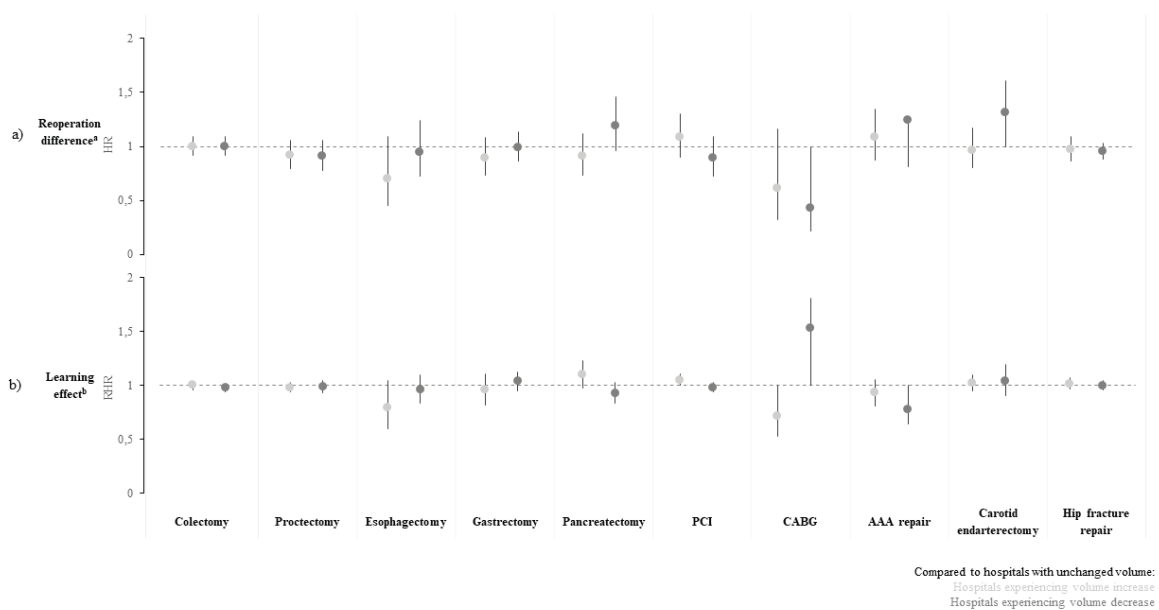


Figure E2 Reoperation difference and individual hospital learning effect between volume trend groups according to procedures, between 2010 and 2014. CABG coronary artery bypass grafting. AAA abdominal aortic aneurysm. PCI percutaneous coronary intervention. ^a Comparison of patient reoperation across hospitals by comparing hospitals with increasing or decreasing volume with hospitals with unchanged volume. Hazard-ratios estimated from Fine and Gray’s competing risk model considering patient characteristics (age, gender, Elixhauser list of comorbidities, type and year of procedure, transfer, emergency admission, and median income) and hospital characteristics (hospital status, volume of procedures, specialization degree, and attraction rate). The clustering effect of patients within hospitals was taken into account with robust variance estimator.

^b Analyse to determine if reoperation improved or deteriorated over time within hospital that increased or decreased its volume. The ratio of hazard ratio (RHR) compare the change in the reoperation rate between two groups. A RHR greater than 1 suggests that the increase of reoperation over time was greater in hospitals experiencing volume increase/decrease than in hospitals with unchanged volume.

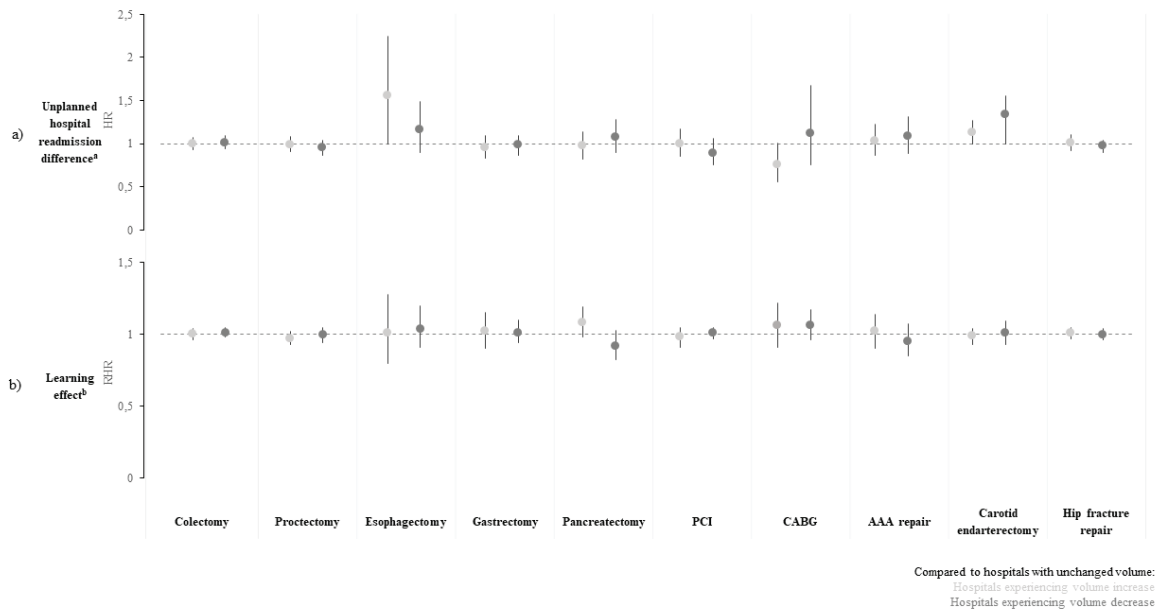


Figure E3 Unplanned hospital readmission difference and individual hospital learning effect between volume trend groups according to procedures, between 2010 and 2014.

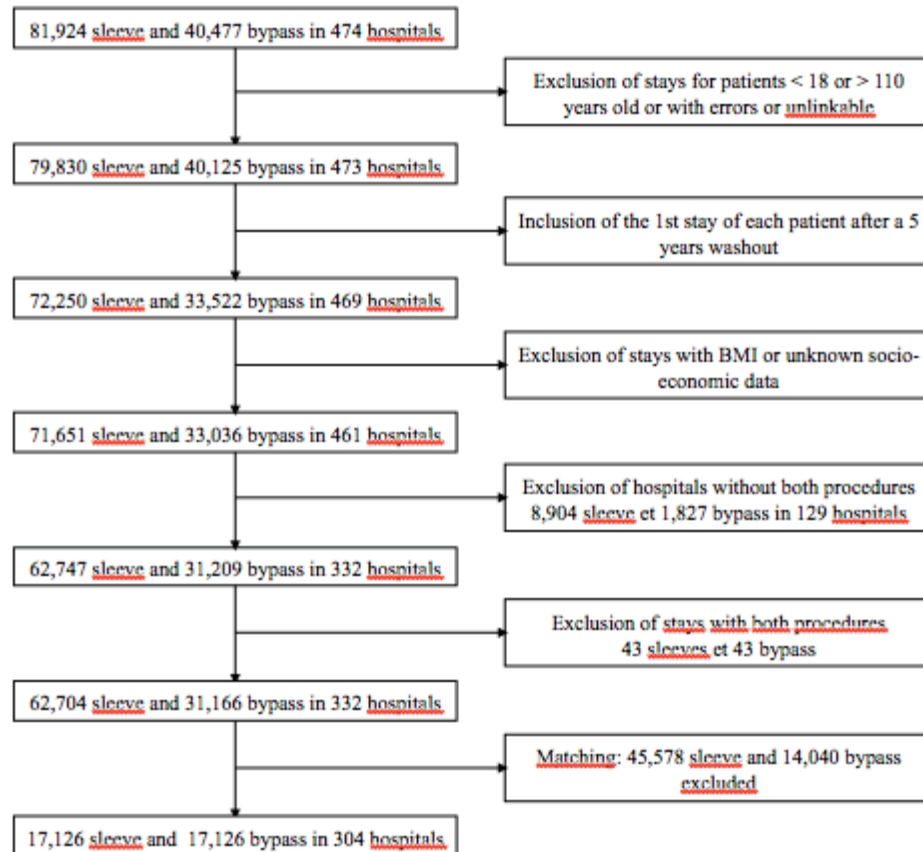
CABG coronary artery bypass grafting. AAA abdominal aortic aneurysm. PCI percutaneous coronary intervention. ^a Hazard-ratios estimated from Fine and Gray's competing risk model considering patient characteristics (age, gender, Elixhauser list of comorbidities, type and year of procedure, transfer, emergency admission, and median income) and hospital characteristics (hospital status, volume of procedures, specialization degree, and attraction rate). The clustering effect of patients within hospitals was taken into account with robust variance estimator. ^b Comparison of patient unplanned hospital readmission across hospitals by comparing hospitals with increasing or decreasing volume with hospitals with unchanged volume. ^c Analyse to determine if unplanned hospital readmission improved or deteriorated over time within hospital that increased or decreased its volume. The ratio of hazard ratio (RHR) compare the change in the unplanned readmission rate between two groups. A RHR greater than 1 suggests that the increase of readmission over time was greater in hospitals experiencing volume increase/decrease than in hospitals with unchanged volume.

Table E2 – Outcome difference and individual hospital learning effect between hospital groups according to trends in procedures volume from 2010 to 2014

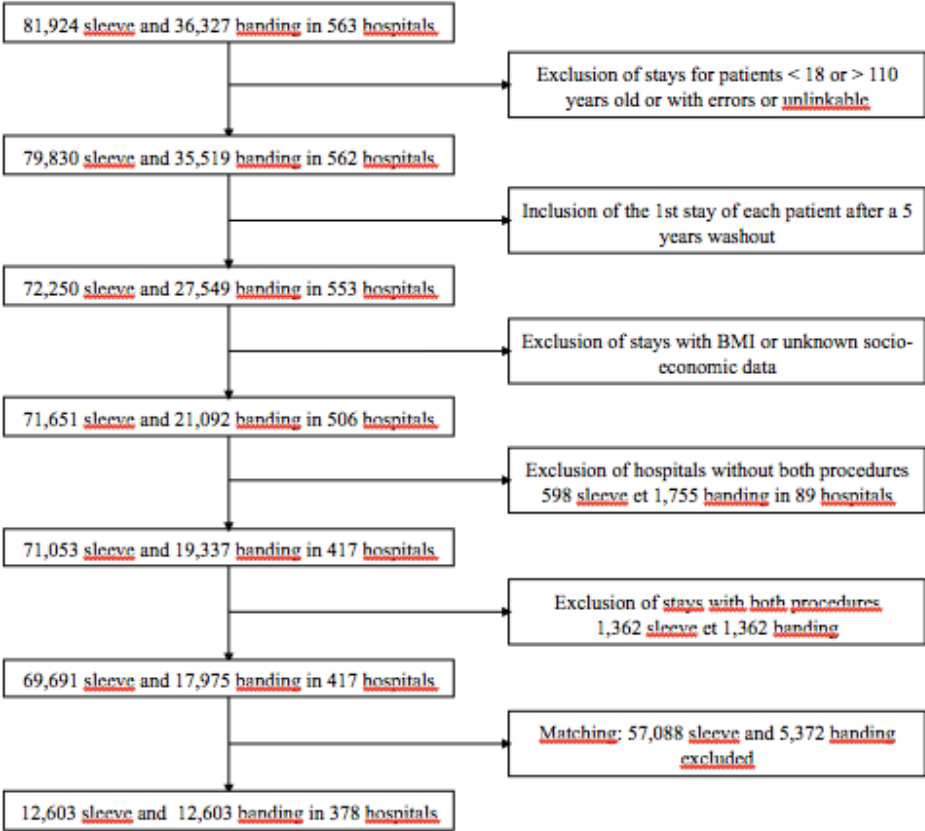
	Mortality		Reoperation		Readmission	
	Difference HR [95%CI]	Learning effect HR [95%CI]	Difference HR [95%CI]	Learning effect HR [95%CI]	Difference HR [95%CI]	Learning effect HR [95%CI]
Colectomy (ref = Unchanged volume)						
<i>Increased</i>	1.04 [0.93 1.17]	1.01 [0.95 1.08]	1.00 [0.92 1.09]	1.00 [0.96 1.03]	1.00 [0.93 1.07]	1.00 [0.96 1.04]
<i>Decreased</i>	1.08 [0.97 1.21]	0.99 [0.92 1.05]	1.00 [0.92 1.09]	0.98 [0.94 1.02]	1.01 [0.94 1.09]	1.01 [0.98 1.05]
Proctectomy (ref = Unchanged volume)						
<i>Increased</i>	0.99 [0.82 1.19]	0.95 [0.85 1.05]	0.92 [0.79 1.06]	0.98 [0.94 1.03]	0.99 [0.91 1.08]	0.97 [0.93 1.02]
<i>Decreased</i>	0.96 [0.79 1.16]	1.05 [0.92 1.20]	0.91 [0.78 1.06]	0.99 [0.93 1.05]	0.95 [0.86 1.04]	1.00 [0.94 1.05]
Esophagectomy (ref = Unchanged volume)						
<i>Increased</i>	0.69 [0.35 1.36]	0.65 [0.51 0.83]	0.70 [0.45 1.09]	0.79 [0.60 1.05]	1.56 [1.08 2.25]	1.01 [0.80 1.28]
<i>Decreased</i>	0.92 [0.62 1.36]	0.97 [0.75 1.26]	0.94 [0.72 1.24]	0.96 [0.84 1.10]	1.16 [0.90 1.49]	1.04 [0.91 1.20]
Gastrectomy (ref = Unchanged volume)						
<i>Increased</i>	0.91 [0.66 1.25]	1.01 [0.81 1.25]	0.89 [0.73 1.08]	0.96 [0.82 1.11]	0.95 [0.83 1.10]	1.02 [0.90 1.15]
<i>Decreased</i>	1.11 [0.89 1.39]	1.00 [0.88 1.15]	0.99 [0.86 1.14]	1.04 [0.95 1.13]	0.98 [0.87 1.10]	1.01 [0.94 1.10]
Pancreatotomy (ref = Unchanged volume)						
<i>Increased</i>	1.39 [1.02 1.90]	1.09 [0.90 1.32]	0.91 [0.73 1.12]	1.10 [0.98 1.23]	0.97 [0.82 1.14]	1.08 [0.98 1.19]
<i>Decreased</i>	1.47 [0.95 2.27]	0.95 [0.78 1.16]	1.19 [0.96 1.46]	0.93 [0.84 1.03]	1.07 [0.90 1.28]	0.92 [0.82 1.03]
PCI (ref = Unchanged volume)						
<i>Increased</i>	1.05 [0.90 1.21]	0.97 [0.89 1.05]	1.08 [0.90 1.30]	1.05 [1.00 1.11]	1.00 [0.85 1.17]	0.98 [0.91 1.05]
<i>Decreased</i>	1.13 [0.92 1.40]	1.04 [0.95 1.13]	0.89 [0.72 1.09]	0.98 [0.94 1.03]	0.89 [0.75 1.06]	1.01 [0.97 1.05]
CABG (ref = Unchanged volume)						
<i>Increased</i>	0.75 [0.48 1.17]	1.01 [0.81 1.26]	0.71 [0.53 0.94]	0.71 [0.53 0.94]	1.06 [0.91 1.22]	1.06 [0.91 1.22]
<i>Decreased</i>	1.16 [0.59 2.28]	1.03 [0.82 1.30]	0.43 [0.22 0.82]	1.53 [1.30 1.81]	1.53 [1.30 1.81]	1.06 [0.96 1.17]
AAA repair (ref = Unchanged volume)						
<i>Increased</i>	0.92 [0.65 1.32]	1.11 [0.86 1.45]	1.08 [0.87 1.35]	0.93 [0.81 1.06]	1.03 [0.87 1.23]	1.02 [0.90 1.14]
<i>Decreased</i>	0.98 [0.58 1.67]	0.85 [0.61 1.18]	1.24 [0.94 1.64]	0.78 [0.64 0.96]	1.08 [0.89 1.31]	0.95 [0.85 1.07]
Carotid endarterectomy (ref = Unchanged volume)						
<i>Increased</i>	1.23 [0.99 1.53]	0.98 [0.86 1.11]	0.96 [0.80 1.17]	1.02 [0.95 1.10]	1.13 [1.01 1.27]	0.99 [0.93 1.04]
<i>Decreased</i>	1.19 [0.81 1.74]	0.97 [0.78 1.22]	1.31 [1.07 1.61]	1.04 [0.91 1.20]	1.34 [1.16 1.56]	1.01 [0.93 1.09]
Hip fracture repair (ref = Unchanged volume)						
<i>Increased</i>	1.05 [0.93 1.18]	1.01 [0.95 1.07]	0.97 [0.86 1.09]	1.01 [0.97 1.07]	1.01 [0.92 1.11]	1.01 [0.97 1.05]
<i>Decreased</i>	1.00 [0.88 1.13]	0.96 [0.91 1.01]	0.95 [0.88 1.03]	1.00 [0.96 1.05]	0.97 [0.90 1.04]	1.00 [0.96 1.04]

3. Annexes de l'article 3

Annex 3 – Flowchart of patients for the comparison between sleeve and gastric bypass



Annex 4 – Flowchart of patients for the comparison between sleeve and gastric banding



4. Annexes de l'article 4

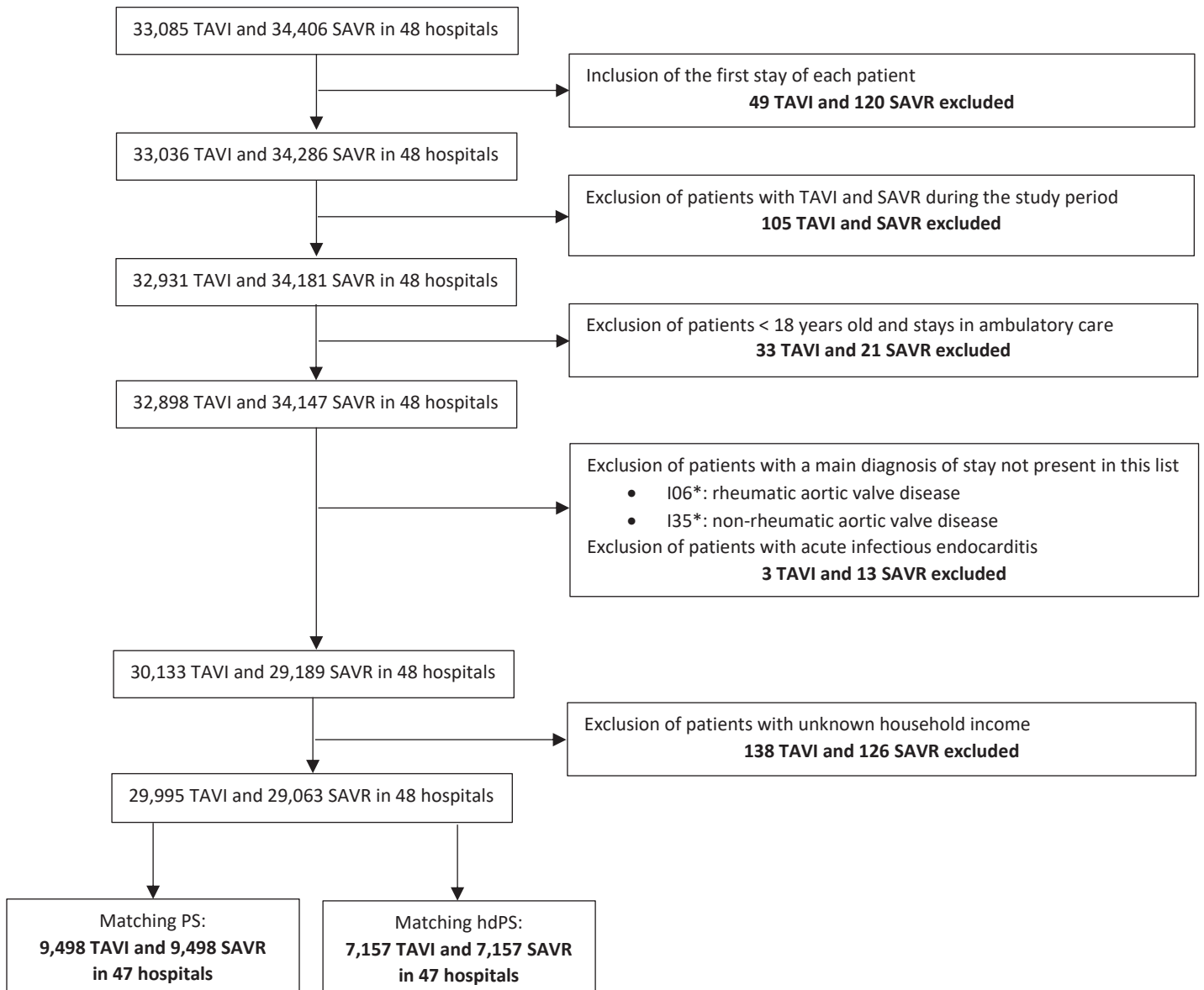


Figure E15 – Study flowchart. TAVI, transcatheter aortic valve implantation; SAVR, surgical aortic valve replacement; PS, propensity score; hdPS, high-dimensional propensity score.

Table E1 - Compared outcomes between transcatheter aortic valve implantation (TAVI) and surgical aortic valve replacement (SAVR) with Cox Model

	PS-matching HR [95%CI]	hdPS-matching HR [95%CI]
Reoperation		
<i>At 1 month</i>	1.92 [1.17-3.14]	1.27 [0.76-2.12]
<i>At 1 year</i>	1.60 [1.08-2.38]	1.30 [0.85-1.99]
<i>At 2 year</i>	1.70 [1.17-2.48]	1.34 [0.90-2.01]
<i>At 3 year</i>	1.87 [1.30-2.70]	1.47 [1.00-2.17]
Stroke		
<i>At 1 month</i>	2.94 [1.67-5.19]	2.44 [1.36-4.37]
<i>At 1 year</i>	1.41 [1.16-1.71]	1.26 [1.00-1.59]
<i>At 2 year</i>	1.28 [1.10-1.50]	1.17 [0.98-1.40]
<i>At 3 year</i>	1.25 [1.08-1.43]	1.15 [0.98-1.35]

PS, propensity score; hdPS, high-dimensional propensity score; HR, hazard ratio; CI, confidence interval

Table E2 - Compared outcomes between transcatheter aortic valve implantation (TAVI) and surgical aortic valve replacement (SAVR) using high-dimensional propensity score (hdPS) matching method with 100, 200 and 500 of empirically identified covariates

	k=100 HR [95%CI]	k=200 HR [95%CI]	k=500 HR [95%CI]
All-cause death			
<i>At 1 month</i>	0.79 [0.67-0.94]	0.82 [0.69-0.97]	0.82 [0.69-0.98]
<i>At 1 year</i>	1.21 [1.08-1.35]	1.20 [1.08-1.34]	1.19 [1.06-1.33]
<i>At 2 year</i>	1.39 [1.27-1.53]	1.39 [1.27-1.53]	1.32 [1.20-1.46]
<i>At 3 year</i>	1.44 [1.33-1.57]	1.46 [1.35-1.59]	1.36 [1.25-1.49]

HR, hazard ratio; CI, confidence interval

B. Publications et communications de ce travail

1. Publications

Payet C, Lifante JC, Carty MJ, Rabilloud M, Duclos A. Methodological Quality of Surgical Mortality Studies Using Large Hospital Databases: A Systematic Review. *Ann Surg.* 2017 Jun;265(6):1113-1118

Payet C, Polazzi S, Lifante JC, Cotte E, Grinberg D, Carty MJ, Sanchez S, Rabilloud M, Duclos A. Influence of trends in hospital volume over time on patient outcomes for high-risk surgery. *BMC Health Services Research* (Accepté)

Brunaud L, Payet C, Polazzi S, Bihain F, Quilliot D, Lifante JC, Duclos A. Reoperations incidence and severity within 6 months after bariatric surgery : a propensity-matched study from nationwide data. *Obesity Surgery* (Accepté)

Payet C, Polazzi S, Obadia JF, Armoiry X, Labarère J, Rabilloud M, Duclos A. High-dimensional propensity score to control indication bias in surgical comparative effectiveness studies : a case study on aortic valve replacement. *Journal of Clinical Epidemiology* (Soumis)

2. Communications orales

Payet C, Rabilloud M, Lifante JC, Duclos A. Prise en compte des facteurs confondants pour étudier la mortalité: chirurgicale dans les bases médico-administratives : une revue de la littérature. EMOIS, Dijon 2016.

Payet C, Lifante JC, Carty MJ Rabilloud M, Duclos A. Methodological Quality of Surgical Mortality Studies Using Large Hospital Databases: A Systematic Review. ACS, Washington 2016.

Payet C, Polazzi S, Sanchez S, et al. Lien entre l'évolution du volume de procédures chirurgicales réalisées par les hôpitaux et la survenue de complications. EMOIS, Montpellier 2018.

Payet C, Polazzi S, Lifante JC, et al. Influence of Trends in Hospital Volume Over Time on Patient Outcomes for High-Risk Surgery. ACS, Boston, 2018.

Payet C, Polazzi S, Lifante JC, et al. Comparaison du risque de reprise chirurgicale en fonction du type de chirurgie bariatrique. EMOIS, Nancy 2019.

Payet C, Polazzi S, Brunaud L, Duclos A. Comparison of reoperation incidence and severity within 6 months after bariatric surgery. ACS, San Francisco 2019.

Payet C, Polazzi S, Obadia JF, et al. Performance des scores de propension à haute dimension pour contrôler le biais d'indication lors de la comparaison de résultats chirurgicaux dans les bases hospitalières nationales. EMOIS, Paris 2020.

3. Posters

Payet C, Lifante JC, Carty MJ et al. Methodological Quality of Surgical Mortality Studies Using Large Hospital Databases: A Systematic Review. ISQUA, Tokyo 2016.

Payet C, Polazzi S, Lifante JC, et al. Trends in hospital volume over time and outcomes for high-risk surgery. IEA, Lyon 2018.

Payet C, Polazzi S, Rabilloud M, Duclos A. Trends in hospital volume over time and outcomes for high-risk surgery. ISQUA, Kuala Lumpur 2018.

Payet C, Polazzi S, Brunaud L, Duclos A. Comparison of reoperation incidence and severity within 6 months after bariatric surgery. ISQUA, Cape Town 2019.