



HAL
open science

Approches innovantes dans l'analyse des données massives en santé

Andrea Lazzati

► **To cite this version:**

Andrea Lazzati. Approches innovantes dans l'analyse des données massives en santé. Sciences du Vivant [q-bio]. UPEC, 2022. tel-03948354

HAL Id: tel-03948354

<https://inria.hal.science/tel-03948354>

Submitted on 20 Jan 2023

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

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

Université Paris-Est Créteil

**Dossier de candidature à
l'Habilitation à Diriger des Recherches**

Année 2021-2022

**Approches innovantes dans l'analyse des données
massives en santé**

Andrea Lazzati MD PhD

Service de Chirurgie Générale et Digestive
Centre Hospitalier Intercommunal de Créteil

JURY

Pr Etienne AUDUREAU	Université Paris-Est Créteil	Examineur
Pr Robert CAIAZZO	Université de Lille	Rapporteur
Pr Christos CHOUAID	Université Paris-Est Créteil	Examineur
Pr Geneviève DERUMEAUX	Université Paris-Est Créteil	Examineur
Pr Caroline GRONNIER	Université de Bordeaux	Rapporteur
Pr Sandrine KATSAHIAN	Université de Paris	Examineur
Pr Emilie LANOY	Université de Paris	Rapporteur
Pr Daniele SOMMACALE	Université Paris-Est Créteil	Examineur

AVANT-PROPOS ET REMERCIEMENTS

Je tiens à témoigner ma reconnaissance à tous ceux qui m'ont accompagné durant ces années où se sont entremêlées activités hospitalières et recherche. Mes collègues hospitaliers en particulier, qui ont supporté toutes mes absences et qui sont devenus des amis.

J'aimerais remercier également les membres du jury pour avoir accepté de lire et de juger ce travail. De près ou de loin, nos vies professionnelles se sont croisées et je ne peux que vous remercier pour votre aide et votre confiance.

Un grand merci à Caroline, Emilie et Robert qui ont accepté d'être les rapporteurs de cet opus.

Un merci tout particulier à Sandrine Katsahian, qui m'a accompagné depuis de longues années, avec un master, une thèse, une mobilité et maintenant cette habilitation. Merci pour ton aide et tous tes conseils dans les moments de difficulté.

Ces remerciements sont plus particulièrement adressés à mes étudiants, qui ont accepté de participer avec moi à cette aventure scientifique, et à qui j'espère avoir transmis en retour l'envie de poursuivre dans la voie de la recherche. Vous m'avez aidé à apprendre probablement plus que ce que j'ai réussi à vous transmettre.

Pour finir, je souhaiterais plus particulièrement remercier Virginie, mon épouse, qui m'accompagne depuis 10 ans, déjà, et qui est la seule qui n'a pas eu le choix de lire tous mes travaux. Je t'aime.

SOMMAIRE

Jury.....	2
Avant-propos et remerciements	3
Sommaire.....	5
Résumé	5
Curriculum vitae.....	6
Activité professionnelle	6
Formation	6
Fonctions collectives.....	7
Congres	7
Publications	8
Capacité à concevoir, diriger, animer et coordonner des activités de recherche et encadrements d'étudiants.....	12
Encadrement	12
Coordination de protocoles de recherche clinique	14
Bases de données médico-administratives.....	16
Structure et contenu du PMSI	16
Volumes du PMSI.....	17
Axe I : Trajectoires de soins	19
Contexte général.....	19
Recherche & encadrement	25
Etude n°1.....	26
Etude n°2	37
Etude n°3.....	51
Perspectives de recherche.....	64
Axe II : Inférence causale et donnée observationnelles	66
Contexte général.....	66
Recherche & encadrement	73
Etude n°1.....	74
Etude n°2	89
Etude n°3.....	90
Perspectives de recherche.....	96
Références bibliographiques.....	97

RESUME

L'accès à des données de santé de grande dimension a ouvert des nouvelles possibilités de recherche jusque-là difficiles à concevoir. Dans plusieurs pays, dont la France, des cohortes, des registres ou des données collectées de manière routinière à des fins administratives (les bases des données médico-administratives), s'ouvrent aux chercheurs.

La base nationale des hospitalisations (le programme de Médicalisation des Systèmes d'Information, PMSI) a été au centre de mes recherches depuis plusieurs années. Le Master 2 de Statistiques en Recherche Biomédicale et le Doctorat en Epidémiologie ont utilisé cette source de données comme terrain d'investigation.

Depuis, cette source a été l'objet d'autres travaux réalisés dans l'encadrement des étudiants en Master 2, ingénieurs et assistants. Deux thèses de science sont co-encadrées sur des travaux épidémiologiques.

Actuellement deux axes principaux définissent mes champs de recherche : l'analyse des trajectoires de soins et l'inférence causale à partir de données observationnelles.

Axe I. Le parcours de soins est défini comme une succession de recours au système de soins, en particulier dans le cas des maladies chroniques et des maladies nécessitant une prise en charge multidisciplinaire. L'analyse de ces parcours de soins permet de comprendre le déploiement de la prise en charge sur le territoire, d'en identifier les forces et les faiblesses et, potentiellement d'anticiper des potentielles complications, en particulier après un geste chirurgical.

Pour analyser ces données longitudinales, une méthodologie adaptée prenant en compte la dimension temporelle doit être employée. Parmi les méthodes actuellement existant pour analyser les parcours de soins, nous en étudions deux : l'analyse de séquences, et l'analyse formelle de concept.

Le traitement du cancer et la prise en charge de l'obésité présentent des parcours de soins complexes, articulés, faisant appel à de nombreux intervenants au cours du temps. Ces deux pathologies, cancer et obésité, semblent être des cas d'étude appropriés pour l'analyse des trajectoires de soins.

Axe II. Pour répondre à des questions causales sur l'efficacité ou la sécurité comparatives entre groupes différents, les essais randomisés contrôlés sont le *gold standard* en recherche biomédicale. Lorsqu'on ne peut pas mener une expérience randomisée, pour des raisons éthiques, logistiques ou économiques, on peut alors se tourner vers les données observationnelles.

L'inférence causale à partir de grandes bases de données médicales peut être considérée comme une tentative d'émuler un essai randomisé (l'essai cible) qui répondrait une question d'intérêt. Lorsque l'objectif est de guider les décisions entre plusieurs stratégies, les analyses causales des données observationnelles peuvent être évaluées en fonction de leur efficacité à émuler un essai cible.

Une étape incontournable de l'analyse des données observationnelles est le développement et la validation d'indicateurs de qualité, de pertinence et de sécurité des soins. Ces indicateurs sont d'ailleurs considérés par les instances de santé comme une priorité pour les financements des établissements de santé basés sur la qualité. Actuellement très peu d'indicateurs sont développés à partir des données médico-administratives, alors qu'en pratique clinique on dispose davantage d'outils pour évaluer la qualité et la sécurité des soins. Une partie de notre travail consiste à identifier et valider des indicateurs de résultats en chirurgie.

CURRICULUM VITAE

Né le 20 mars 1976 à Milan (Italie)
Chirurgien Digestif
RPPS 10004394242 - n° d'inscription à l'Ordre : 94/17528
Service de Chirurgie Digestive
Centre Hospitalier Intercommunal de Créteil
40, avenue de Verdun 94000 Créteil
01.57.02.23.31 - 06.67.47.66.03
andrea.lazzati@chicreteil.fr
andrea.lazzati@gmail.com

ACTIVITE PROFESSIONNELLE

- 2018 – présent Chef de Service, Service de Chirurgie Digestive, CHI de Créteil
- 2013 - présent *Praticien Hospitalier - Service de Chirurgie Digestive, CHI de Créteil (Dr J. Poincenot).*
- 2012 - 2013 *Praticien Hospitalier Contractuel - Service de Chirurgie Digestive, Hôpital Henri Mondor (Pr. D. Azoulay)*
- 2009 - 2012 *Chef de Clinique Assistant – puis Praticien Attaché, Service de Chirurgie Digestive, Hôpital Pitié-Salpêtrière (Pr. L. Hannoun)*
- 2007 - 2009 *Chef de Clinique Assistant - Service de Chirurgie Générale, Hôpital Jean Verdier, Bondy (Pr. G. Champault)*

FORMATION

- 2015 - 2019 **Thèse de doctorat** Ecole Doctorale de Santé Publique, ED 393, Epidémiologie et Sciences de l'Information Biomédicale. Titre : « Analyse des Big Data dans l'évaluation de la chirurgie bariatrique »
- 2016 - 2017 **DIU de Pédagogie médicale**
- 2014 - 2015 **MASTER 2** « Méthodologie et Statistique en Recherche Biomédicale », Université Paris Sud
- 2012 - 2013 **MASTER 1** « Sante Publique », Université Paris Sud
- 2010 - 2011 **DU Cancérologie**, Université Paris VI
- 2008 - 2009 **DIU Proctologie**
- 2001 - 2007 **Diplôme en Chirurgie Générale**, Université des Etudes de Milan, Italie
- 1995 - 2001 Diplôme d'Etat de **Docteur en Médecine**, Université des Etudes de Milan, Italie
- 1990 - 1995 Baccalauréat en **Lettres Classiques**

UNITE DE RECHERCHE

2015-2020 INSERM U1138, équipe 22, « Information Sciences To Support Personalized Medicine », Université Paris Descartes

2020- INSERM IMRB U955, équipe « Sénescence, métabolisme et maladies cardiovasculaires », Université Paris-Est Créteil

LANGUES

Italien: maternelle ; Français: courant ; Anglais: courant ; Espagnol: scolaire (Erasmus à Lleida, Espagne)

FONCTIONS COLLECTIVES

- ✓ CME : membre élu de la CME depuis 2019
- ✓ SFCD : membre élu du Conseil d'administration de la Société Française de Chirurgie Digestive depuis 2020
- ✓ SOFFCO-MM: membre de la Société Française et Francophone de Chirurgie de l'obésité et des Maladie Métaboliques
- ✓ OBECHIC : fondateur et Président de l'Association OBECHIC

EXPERTISE DE DOSSIERS

Expertise de dossiers soumis à des demandes de financement : **PHRC** (Programme hospitalier de recherche clinique, 2016) ; et **PHRIP** (Programme hospitalier de recherche infirmière et paramédical, 2016)

Commission d'Expertise Scientifique **PHRC-I** 2019-2022

REVIEWING

Reviewer pour les revues: Obesity Surgery, Surgery for Obesity and Related Diseases, Langenbeck's Archives of Surgery, Peer J.

CONGRES

COMMUNICATIONS ORALES AUX CONGRES NATIONAUX:

- ✓ Association Française de Chirurgie (AFC) 2014, 2016, 2018
- ✓ Société Française et Francophone de Chirurgie de l'obésité et des maladies métaboliques (Soffco-MM) 2010, 2013, 2014, 2015, 2016, 2017, 2018, 2020, 2021
- ✓ Club Soffco Jeune 2015, 2018, 2021
- ✓ Société Française de Chirurgie Digestive (SFCD) 2013, 2014, 2019 (DPC), 2020
- ✓ Association Evaluation, management, organisations, santé (EMOIS) 2019, 2020
- ✓ Assises pour l'innovation en Médecine Esthétique (AIME) 2020 (DPC)
- ✓ Société française de chirurgie plastique, reconstructive et esthétique (SOF.CPRE) 2017

COMMUNICATIONS ORALES AUX CONGRES INTERNATIONAUX:

- ✓ IFSO 2009, 2013, 2014, 2015, 2017-2020

PUBLICATIONS

Nombre d'article originaux	40
Nombre d'article en premier auteur	11
Nombre d'article en dernier auteur	11
Points SIGAPS individuel	794 (au 22/10/21)
H-Index (source ResearchGate)	13
Activité de reviewing	>50 articles

PUBLICATIONS PAR AN ET PAR CATEGORIE SIGAPS

Période : 2011 - 2021									
Année	Total	A+	A	B	C	D	E	NC	Score
2011	0	0	0	0	0	0	0	0	0
2012	0	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0	0
2014	1	0	1	0	0	0	0	0	32
2015	3	0	0	2	1	0	0	0	52
2016	5	0	4	1	0	0	0	0	112
2017	3	0	3	0	0	0	0	0	96
2018	4	0	0	4	0	0	0	0	84
2019	3	0	0	3	0	0	0	0	72
2020	5	0	1	4	0	0	0	0	88
2021	16	0	2	12	1	1	0	0	258
Total	40	0	11	26	2	1	0	0	794

ORDRE PARMIS LES AUTEURS

Période : 2011 - 2021								
Position	Total	A+	A	B	C	D	E	NC
1	11	0	2	8	1	0	0	0
2	7	0	2	5	0	0	0	0
3	1	0	1	0	0	0	0	0
k	8	0	2	5	1	0	0	0
ADA	2	0	1	1	0	0	0	0
DA	11	0	3	7	0	1	0	0
Total	40	0	11	26	2	1	0	0

LISTE DE PUBLICATIONS

Moszkowicz, D ; Jacota, M ; Nkam, L ; Giovinazzo, D ; Grimaldi, L ; **Lazzati, A** Ventral Hernia Repair and Obesity: Results from a Nationwide Register Study in France According to the Timeframes of Hernia Repair and Bariatric Surgery. Obes Surg 2021, IF:4.13, Score SIGAPS: 24

Iannelli, A ; Bulsei, J ; Debs, T ; Tran, A ; **Lazzati, A** ; Gugenheim, J & al Clinical and Economic Impact of Previous Bariatric Surgery on Liver Transplantation: a Nationwide, Population-Based Retrospective Study. Obes Surg 2021, IF:4.13, Score SIGAPS: 6

Singhal, R ; Ludwig, C ; Rudge, G ; Gkoutos, GV ; Tahrani, A ; Mahawar, K & al 30-Day Morbidity and Mortality of Bariatric Surgery During the COVID-19 Pandemic: a Multinational Cohort Study of 7704 Patients from 42

Countries. *Obes Surg* 2021, IF:4.13, Score SIGAPS: 6

Spota, A ; Cereatti, F ; Granieri, S ; Antonelli, G ; Dumont, JL ; Dagher, I & al Endoscopic Management of Bariatric Surgery Complications According to a Standardized Algorithm. *Obes Surg* 2021, IF:4.13, Score SIGAPS: 6

Lazzati, A ; Salaün-Penquer, N ; Ortala, M ; Vignot, M ; De Filippo, G ; Jung, C Trends in metabolic bariatric surgery in adolescents in France: a nationwide analysis on an 11- year period. *Surg Obes Relat Dis* 2021, IF:4.73, Score SIGAPS: 24

Challine, A ; Voron, T ; Dousset, B ; Creavin, B ; Katsahian, S ; Parc, Y & al Postoperative outcomes after laparoscopic or open gastrectomy. A national cohort study of 10,343 patients. *Eur J Surg Oncol* 2021, IF:4.42, Score SIGAPS: 18

Lazzati, A ; Raphael Rousseau, M ; Bartier, S ; Dabi, Y ; Challine, A ; Haddad, B & al Impact of COVID-19 on surgical emergencies: nationwide analysis. *BJS Open* 2021, IF:3.4, Score SIGAPS: 16

Donatelli, G ; Spota, A ; Cereatti, F ; Granieri, S ; Dagher, I ; Chiche, R & al Endoscopic internal drainage for the management of leak, fistula, and collection after sleeve gastrectomy: our experience in 617 consecutive patients. *Surg Obes Relat Dis* 2021, IF:4.73, Score SIGAPS: 6

Challine, A ; **Lazzati, A** ; Katsahian, S ; Parc, Y ; Lefevre, JH Colorectal screening: We have not caught up. A surge of colorectal cancer after the coronavirus disease 2019 (COVID-19) pandemic? *Surgery* 2021, IF:3.98, Score SIGAPS: 18

Lazzati, A ; Chatellier, G ; Paolino, L ; Batahei, S ; Katsahian, S Postoperative care fragmentation in bariatric surgery and risk of mortality: a nationwide study. *Surg Obes Relat Dis* 2021, IF:4.73, Score SIGAPS: 24

Martellotto, S ; Challine, A ; Peveri, V ; Paolino, L ; **Lazzati, A** Trends in emergent diverticular disease management: a nationwide cohort study from 2009 to 2018. *Tech Coloproctol* 2021, IF:3.78, Score SIGAPS: 24

Tombini, V ; Di Capua, M ; Capsoni, N ; **Lazzati, A** ; Bergamaschi, M ; Gheda, S & al Risk Stratification in COVID-19 Pneumonia - Determining the Role of Lung Ultrasound. *Ultraschall Med* 2021, IF:6.55, Score SIGAPS: 8

Challine, A ; Dousset, B ; de'Angelis, N ; Lefèvre, JH ; Parc, Y ; Katsahian, S & al , **Lazzati, A** Impact of coronavirus disease 2019 (COVID-19) lockdown on in-hospital mortality and surgical activity in elective digestive resections: A nationwide cohort analysis. *Surgery* 2021, IF:3.98, Score SIGAPS: 24

Challine, A ; **Lazzati, A** ; Dousset, B ; Voron, T ; Parc, Y ; Lefevre, JH Colorectal screening: We have not caught up. A surge of colorectal cancer after the coronavirus disease 2019 (COVID-19) pandemic? *Surgery* 2021, IF:3.98, Score SIGAPS: 18

Chatzistergiou, TK ; Zervaki, DS ; Derouich, M ; **Lazzati, A** Laparoscopic sleeve gastrectomy and pregnancy outcomes: A systematic review. *Eur J Obstet Gynecol Reprod Biol* 2021, IF:2.44, Score SIGAPS: 12

Challine, A ; Rives-Lange, C ; **Lazzati, A** ; Czernichow, S Response to the Comment on "Impact of Oral Immunonutrition on Postoperative Morbidity in Digestive Oncologic Surgery: A Nation-wide Cohort Study". *Ann Surg* 2020, IF:12.97, Score SIGAPS: 16

Lazzati, A ; Bechet, S ; Jouma, S ; Paolino, L ; Jung, C Revision surgery after sleeve gastrectomy: a nationwide study with 10 years of follow-up. *Surg Obes Relat Dis* 2020, IF:4.73, Score SIGAPS: 24

Liagre, A ; Debs, T ; Kassir, R ; Ledit, A ; Juglard, G ; Chalret du Rieu, M & al One Anastomosis Gastric Bypass

with a Biliopancreatic Limb of 150 cm: Weight Loss, Nutritional Outcomes, Endoscopic Results, and Quality of Life at 8-Year Follow-Up. *Obes Surg* 2020, IF:4.13, Score SIGAPS: 6

Paolino, L ; Pravettoni, R ; Epaud, S ; Ortala, M ; **Lazzati, A** Comparison of Surgical Activity and Scientific Publications in Bariatric Surgery: an Epidemiological and Bibliometric Analysis. *Obes Surg* 2020, IF:4.13, Score SIGAPS: 24

Charles-Nelson, A ; **Lazzati, A** ; Katsahian, S Analysis of Trajectories of Care After Bariatric Surgery Using Data Mining Method and Health Administrative Information Systems. *Obes Surg* 2020, IF:4.13, Score SIGAPS: 18

Lazzati, A ; Chatellier, G ; Katsahian, S Readmissions After Bariatric Surgery in France, 2013-2016: a Nationwide Study on Administrative Data. *Obes Surg* 2019, IF:3.41, Score SIGAPS: 24

Challine, A ; Rives-Lange, C ; Danoussou, D ; Katsahian, S ; Ait Boudaoud, A ; Gaujoux, S & al Impact of Oral Immunonutrition on Postoperative Morbidity in Digestive Oncologic Surgery: A Nation-wide Cohort Study. *Ann Surg* 2021, IF:12.97, Score SIGAPS: 24

Bou Nassif, G ; Salloum, C ; Paolino, L ; Azoulay, D ; **Lazzati, A** Laparoscopic Sleeve Gastrectomy After Orthotopic Liver Transplantation, Video Reported. *Obes Surg* 2019, IF:3.41, Score SIGAPS: 24

Bou Nassif, G ; Paolino, L ; **Lazzati, A** Total Gastrectomy with Roux-En-Y esophagojejunostomy for Chronic Complicated Post-Sleeve Gastric Fistula-Video Report. *Obes Surg* 2019, IF:3.41, Score SIGAPS: 24

Lazzati, A ; Nassif, GB ; Paolino, L Concomitant Ventral Hernia Repair and Bariatric Surgery: a Systematic Review. *Obes Surg* 2018, IF:3.6, Score SIGAPS: 24

Lazzati, A ; Katsahian, S ; Maladry, D ; Gerard, E ; Gaucher, S Plastic surgery in bariatric patients: a nationwide study of 17,000 patients on the national administrative database. *Surg Obes Relat Dis* 2018, IF:3.76, Score SIGAPS: 24

Martini, F ; **Lazzati, A** ; Fritsch, S ; Liagre, A ; Iannelli, A ; Paolino, L General Practitioners and Bariatric Surgery in France: Are They Ready to Face the Challenge? *Obes Surg* 2018, IF:3.6, Score SIGAPS: 18

Osseis, M ; **Lazzati, A** ; Salloum, C ; Gavara, CG ; Compagnon, P ; Feray, C & al Sleeve Gastrectomy After Liver Transplantation: Feasibility and Outcomes. *Obes Surg* 2018, IF:3.6, Score SIGAPS: 18

Bou Nassif, G ; Scetbun, E ; Lecurieux-Lafayette, C ; **Lazzati, A** \"Hand-Over-Hand Grasping Technique\": A Fast and Safe Procedure for Specimen Extraction in Laparoscopic Sleeve Gastrectomy. *Obes Surg* 2017, IF:3.9, Score SIGAPS: 32

Carandina, S ; Tabbara, M ; Galiay, L ; Polliand, C ; Azoulay, D ; Barrat, C & **Lazzati, A** Long-Term Outcomes of the Laparoscopic Adjustable Gastric Banding: Weight Loss and Removal Rate. A Single Center Experience on 301 Patients with a Minimum Follow-Up of 10 years. *Obes Surg* 2017, IF:3.9, Score SIGAPS: 32

Martini, F ; Paolino, L ; Marzano, E ; D'Agostino, J ; **Lazzati, A** ; Schneck, AS & al Single-Anastomosis Pylorus-Preserving Bariatric Procedures: Review of the Literature. *Obes Surg* 2016, IF:3.95, Score SIGAPS: 8

Lazzati, A ; De Antonio, M ; Paolino, L ; Martini, F ; Azoulay, D ; Iannelli, A & al Natural History of Adjustable Gastric Banding: Lifespan and Revisional Rate: A Nationwide Study on Administrative Data on 53,000 Patients. *Ann Surg* 2017, IF:9.2, Score SIGAPS: 32

Poghosyan, T ; **Lazzati, A** ; Moszkowicz, D ; Danoussou, D ; Vychnevskaia, K ; Azoulay, D & al Conversion of sleeve gastrectomy to Roux-en-Y gastric bypass: an audit of 34 patients. *Surg Obes Relat Dis* 2016, IF:4.5, Score

SIGAPS: 24

Dabi, Y ; Darrigues, L ; Katsahian, S ; Azoulay, D ; De Antonio, M ; **Lazzati, A** Publication Trends in Bariatric Surgery: a Bibliometric Study. *Obes Surg* 2016, IF:3.95, Score SIGAPS: 32

Schneck, AS ; **Lazzati, A** ; Audureau, E ; Hemery, F ; Gugenheim, J ; Azoulay, D & al One or two steps for laparoscopic conversion of failed adjustable gastric banding to sleeve gastrectomy: a nationwide French study on 3357 morbidly obese patients. *Surg Obes Relat Dis* 2016, IF:4.5, Score SIGAPS: 24

Cornet, M ; Lim, C ; Salloum, C ; **Lazzati, A** ; Compagnon, P ; Pascal, G & al Prognostic value of sarcopenia in liver surgery. *J Visc Surg* 2015, IF:1.57, Score SIGAPS: 4

Lazzati, A ; Audureau, E ; Hemery, F ; Schneck, AS ; Gugenheim, J ; Azoulay, D & al Reduction in early mortality outcomes after bariatric surgery in France between 2007 and 2012: A nationwide study of 133,000 obese patients. *Surgery* 2016, IF:3.9, Score SIGAPS: 24

Paolino, L ; Genser, L ; Fritsch, S ; De' Angelis, N ; Azoulay, D ; **Lazzati, A** The web-surfing bariatric patient: the role of the internet in the decision-making process. *Obes Surg* 2015, IF:3.35, Score SIGAPS: 24

Lazzati, A ; Iannelli, A ; Schneck, AS ; Nelson, AC ; Katsahian, S ; Gugenheim, J & al Bariatric surgery and liver transplantation: a systematic review a new frontier for bariatric surgery. *Obes Surg* 2015, IF:3.35, Score SIGAPS: 24

Lazzati, A ; Guy-Lachuer, R ; Delaunay, V ; Szwarzensztejn, K ; Azoulay, D Bariatric surgery trends in France: 2005-2011. *Surg Obes Relat Dis* 2014, IF:4.07, Score SIGAPS: 32

CAPACITE A CONCEVOIR, DIRIGER, ANIMER ET COORDONNER DES ACTIVITES DE RECHERCHE ET ENCADREMENTS D'ETUDIANTS

ENCADREMENT

ENCADREMENT DE MASTER 2

Années	Etudiant	Projet	Université	% d'encadrement
2017-2018	Divya Danoussou	Unplanned readmission rate 30 days after major digestive surgery for cancer: a nationwide study based on 285,000 patients. (under review)	M2 Méthodologie et Statistiques en Recherche biomédicale. Université Paris 11	100%
2017-2018	Alexandre Challine	Challine A, Rives-Langes C, Danoussou D, Katsahian S, Ait Boudaoud A, Gaujoux S, Lazzati A , Czernichow S. Impact of Oral Immunonutrition on Postoperative Morbidity in Digestive Oncologic Surgery. Ann Surg. 2019; Publish Ah:1-7.	M2 Méthodologie et Statistiques en Recherche biomédicale. Université Paris 11	80%
2019-2020	Seydou Goro	Impact Of Competition On Mortality Of Patients Operated On For Colorectal Cancer: Cross-Sectionnal Study. (under review)	M2 Méthodologie et Statistiques en Recherche biomédicale. Université Paris 11	100%
2020-2021	Stylianios Tedakis	Analyse des trajectoires des patients atteints d'un cholangiocarcinome en France	Master 2 Santé publique - Parcours : Données massives en santé. Université de Paris	100%
2020-2021	Seydou Goro	Analyse de la trajectoire des patients ayant subi une chirurgie bariatrique.	Master 2 Santé publique - Parcours : Données massives en santé. Université de Paris	100%

CO-ENCADREMENT DE DOCTORANT

Années	Etudiant	Projet	Ecole doctorale	% d'encadrement
2019-	Alexandre Challine	Un score de Morbidité post-opératoire « Clavien-Dindo » ¹ adapté à la recherche sur le PMSI en chirurgie digestive oncologique	ED 393 : Pierre Louis de santé publique. Sorbonne université	90%
2021	Stylianios Tedakis	Analyse des trajectoires des patients atteints d'un cholangiocarcinome en France	ED 393 : Pierre Louis de santé publique. Sorbonne université	50%

ENCADREMENT D'INGENIEURS

Années	Etudiant	Projet	Université	% d'encadrement
2019-2020	Marie Anne Le Ray	Analyse et visualisation des flux des patients - Programmation et analyses statistiques sous r	Diplôme d'ingénieur spécialité Génie biomédical et santé – Université Paris Est	50%
2019-2020	Dounia Zamiati	Analyse et visualisation des flux des patients - Programmation et analyses statistiques sous r	Diplôme d'ingénieur spécialité Génie biomédical et santé – Université Paris Est	50%
2019-2020	Marion Bas Dit Nuges	Analyse et visualisation des flux des patients - Programmation et analyses statistiques sous r	Diplôme d'ingénieur spécialité Génie biomédical et santé – Université Paris Est	50%
2020-2021	Melwin Chhu	Analyse et visualisation des trajectoires de soins en chirurgie digestive	Diplôme d'ingénieur spécialité Génie biomédical et santé – Université Paris Est	100%
2020-2021	Asmaa Benthouami	Prise en charge du cancer en milieu carcéral, analyse de 10 ans d'expérience au niveau national	Diplôme d'ingénieur spécialité Génie biomédical et santé – Université Paris Est - Créteil	100%

ENCADREMENT DE FELLOWS

Années	Etudiant	Projet	% d'encadrement
2016-2018	Georges Bou Nassif	<ol style="list-style-type: none"> 1. Lazzati A, Nassif GB, Paolino L. Concomitant Ventral Hernia Repair and Bariatric Surgery: a Systematic Review. <i>Obes Surg. Obesity Surgery</i>; 2018;28:2949–55. 2. Bou Nassif G, Scetbun E, Lecurieux-Lafayette C, Lazzati A. “Hand-Over-Hand Grasping Technique”: A Fast and Safe Procedure for Specimen Extraction in Laparoscopic Sleeve Gastrectomy. <i>Obes Surg.</i> 2017;27:1391. 3. Bou Nassif G, Salloum C, Paolino L, Azoulay D, Lazzati A. Laparoscopic Sleeve Gastrectomy After Orthotopic Liver Transplantation, Video Reported. <i>Obes Surg [Internet]</i>. Springer New York LLC; 2019 [cited 2021 Jul 8];29:1436–8. 	100%
2018-2019	Sophie Martello	Martello S, Challine A, Peveri V, Paolino L, Lazzati A . Trends in emergent diverticular disease management: a nationwide cohort study from 2009 to 2018. <i>Tech Coloproctol [Internet]</i> . <i>Tech Coloproctol</i> ; 2021	100%
2020-2021	Saad Joumaa	Comparaison de l'efficacité du bypass gastrique de première intention versus de seconde intention. Etude de cohorte. En cours	100%

COORDINATION DE PROTOCOLES DE RECHERCHE CLINIQUE

INVESTIGATEUR PRINCIPAL :

- **STIMUL** : Impacts d'un accompagnement numérique à distance sur l'activité physique adaptée auprès de patient(e)s en parcours de soin pré-chirurgie bariatrique. Essai randomisé contrôlé.
- **PATEX** : Expérience patient en post-chirurgicale : définition de nouveaux indicateurs. L'objectif de ce travail sera de rapporter des récits basés sur des expériences des patients et de construire un questionnaire de qualité du parcours de chirurgie axé sur les besoins des patients.

CO-INVESTIGATEUR

COHORTE BARISEN (RHU en santé CARMMA): Rôle de la senescence du tissu adipeux dans la survenue d'une cardiomyopathie métabolique et d'une sarcopénie associée à l'obésité : une étude de cohorte chez des patients obèses adressés pour chirurgie bariatrique. En collaboration avec le Service d'Explorations fonctionnelles (Pr Derumeaux).

COHORTE OBANEZ : étude de cohorte, prospective sur l'analyse des résistances nasales chez les patients atteints d'obésité et candidats à une chirurgie bariatrique. En collaboration avec le Service d'ORL du CHI de Créteil (Dr Bequignon).

PHRC MINIJOB : Evaluation d'une technique de chirurgie bariatrique minimale invasive utilisant un trocart unique coelioscopique (une seule incision de 2,5 à 3 cm) au lieu de 4 à 7 trocarts dans la gastrectomie longitudinale ou sleeve gastrectomy : essai randomisé contrôlé. Inclusion terminée en juin 2020. Fin de suivi prévu en juin 2022.

PHRC SADISLEEVE : Etude prospective, multicentrique, randomisée, contrôlée de l'efficacité et de la sécurité du Bypass duodéno iléal avec sleeve gastrectomy (Single Anastomosis Duodéno Iléal bypass with sleeve gastrectomy (SADI-S)) versus Bypass gastrique Roux-en-Y (RYGB). Fin des inclusions en octobre 2021. Fin de suivi prévue en octobre 2023.

PHRC BOUST: *"Multicentric randomized prospective study assessing the impact of the bougie calibration size during laparoscopic sleeve gastrectomy on the rate of postoperative staple-line leak rate"*. Début des inclusions: septembre 2021.

PHRC DIABOUT: *"Extending time without diabetes after bariatric surgery: a randomized controlled trial with metformin versus standard care"*. Début des inclusions: septembre 2021.

PHRC BIPASS: *"Prospective multicentric randomized trial comparing the efficacy and safety of Sleeve Gastrectomy with Transit Bipartition (SG+TB) versus Roux-en-Y Gastric Bypass (RYGB)"*. Début des inclusions: à venir.

Nous allons collaborer à deux PHRC retenus en 2021 : **PHRC YOMEGA 2** et **PHRC BeSure**

FINANCEMENT

Obtenus	DGOS Approfondissement des projets médico-soignants partagé des groupements hospitaliers de territoire 30K €	Investigateur Principal : Andrea LAZZATI Année: 2019
Sous évaluation	ATIH: Appel à manifestation d'intérêt. Coopération ATIH - Universités - Organismes de recherche 300K €	Investigateur Principal : Andrea LAZZATI Année: 2021

BASES DE DONNEES MEDICO-ADMINISTRATIVES

Les travaux et les axes de recherche présentés ici se sont basés sur les données extraites de bases de données médico-administratives (BDMA) de l'Assurance Maladie et en particulier du Programme de médicalisation des systèmes d'information (PMSI),

L'Assurance Maladie couvre les coûts liés à tout type de soins : hospitalisations, médecine générale, médicaments, vaccinations, maternité, tests de laboratoire, rééducation, invalidité, etc. Cet organisme fonctionne sur la base de tarifs fixés par des conventions ou par le Ministère de la Santé. En bref, l'Assurance Maladie est chargée de gérer l'ensemble des dépenses de soins du territoire.

Aujourd'hui, les paiements des soins et les processus de remboursement sont entièrement informatisés. Chaque fois qu'une personne rend visite à son médecin ou achète un médicament en pharmacie, une nouvelle entrée est ajoutée à un système d'information dédié: le Système national d'information sur l'assurance maladie « Système National d'Information Inter-régimes de l'Assurance Maladie » (SNIIRAM). Le SNIIRAM est alimenté par les informations générées lors de la prise en charge de l'ensemble des consommations de soins et des hospitalisations en France. La masse principale est fournie par 1.2 milliard de demandes de soins chaque année, le volume estimé étant de 450 téraoctets [1]. C'est sans doute la plus grande base de données sur les soins de santé au monde. Chaque personne reçoit un code d'identité anonyme lorsqu'elle a besoin de soins pour la toute première fois. Ensuite, le même identifiant est utilisé toute sa vie. Ce système permet le suivi des soins reçus par chaque patient.

L'entrepôt de données SNIIRAM est structuré sur trois *datamart* principaux (ou sous-ensemble) : le PMSI qui regroupe les données de remboursement des soins hospitaliers, le DCIR (« Datamart Consommation Inter Régimes ») qui regroupe les données des soins extrahospitaliers, et les données sur les causes de mortalité (géré par le Centre d'épidémiologie sur les causes médicales de décès - CépiDc).

Le SNIIRAM, le PMSI et les données du CépiDc constituent les bases des données médico-administratives françaises.

STRUCTURE ET CONTENU DU PMSI

La structure du PMSI repose sur le concept des Groupes Homogènes de Malade (GHM). Ces groupes sont construits à partir d'algorithmes qui prennent en compte trois types d'information : le ou les diagnostics expliquant la prise en charge, les actes éventuels effectués lors de celle-ci, et un marqueur caractérisant le patient, principalement son âge. A partir de ces informations les algorithmes classent les patients dans un groupe qui présente une homogénéité médico-économique.

« L'homogénéité médicale repose essentiellement sur la détermination du diagnostic principal de la prise en charge permettant d'orienter le séjour dans l'une des vingt-sept catégories majeure de diagnostic (CMD) chapeautant les GHM, tandis que l'homogénéité économique est, quant à elle, établie par la mesure du pouvoir explicatif des trois types d'informations médicales sur une variable dépendante représentative du coût du séjour, originellement et encore en grande partie approché par la durée du séjour. » [2]

Les diagnostics sont codés selon la 10^{ème} version de la Classification Internationale des Maladies (CIM-10) qui est utilisée depuis l'année 1996. Les actes diagnostiques ou thérapeutiques sont classés selon la Classification Commune des Actes Médicaux (CCAM) qui est utilisée depuis 2002, et généralisée depuis 2005.

Quatre types d'hospitalisations sont concernés par le PMSI : l'activité de court séjour en médecine, chirurgie et obstétrique (MCO), les soins de suite et réadaptation (SSR), l'hospitalisation à domicile (HAD) et la psychiatrie. Dans le champ MCO, chaque séjour dans une unité médicale fait l'objet d'un résumé

d'unité médicale (RUM)¹ qui contient des informations de type administratif, des informations médicales et les informations relatives au groupage GHM. En cas de prise en charge d'un patient dans plusieurs unités médicales pendant le même séjour, plusieurs RUMs sont générées et chaînées ensemble pour constituer un résumé standard de sortie (RSS) dont le diagnostic principal (DP) est choisi par un algorithme de groupage parmi les DP des RUM.

Les différentes informations recueillies au niveau du PMSI peuvent être distinguées en trois types :

- ✓ Les **informations administratives** concernent à la fois l'établissement (numéro FINESS, nombre d'UM intéressées par un séjour donné), le patient (âge, sexe, code postal de résidence, modalité d'entrée et de sortie, durée du séjour, mois et année du séjour, l'indice de gravité simplifié (IGS II) lors d'un passage en réanimation).
- ✓ Les **informations médicales** sont constituées par les différents diagnostics (principal, relié et associé selon la CIM-10), et les actes diagnostiques ou thérapeutiques réalisés (selon la CCAM).
- ✓ Les **informations de groupage** mentionnent principalement le GHM et la CMD dans lesquels sont classés les séjours ainsi que le tarif qui lui est associé dans le cadre de la T2A.

Depuis 2001², à chaque patient est attribué un numéro d'identification anonyme, basé sur le numéro d'assuré social de l'ouvrant droit, du sexe et de la date de naissance du bénéficiaire. Cet identifiant anonyme accompagne chaque individu tout le long de sa vie et permet donc de l'identifier lors des différents séjours. Le chaînage entre les différentes hospitalisations d'un même patient devient donc possible. A partir de 2011 un chaînage est aussi possible entre la mère et son enfant³.

D'autres mesures sont appliquées afin de protéger l'identité des patients : la date de naissance est remplacée par l'âge, les dates d'entrée et de sortie par la durée du séjour et les mois et années de sortie, le code postal par le code géographique (regroupant les plus petites communes), le numéro administratif et les données sur l'UM sont supprimés. De cette façon, le résumé standardisé de sortie (RSS) devient un résumé de sortie anonyme (RSA). Et c'est sous cette forme que les données sont transmises d'abord à l'Agence Régionale de Santé (ARS) puis à l'Agence Technique de l'Information Hospitalière (ATIH).

La durée de conservation des données du PMSI est de 10 ans.

VOLUMES DU PMSI

Dans la base PMSI on a retrouvé en 2019, 32.3 millions de séjours concernant 14.5 millions de patients. La figure 1 montre l'évolution du volume de données collectées par le PMSI au cours de la période 2006-2020. On observe une augmentation régulière d'environ 1.4% par an, passant de 21.2 millions de séjours à l'hôpital en 2007 à 32.3 en 2019. La forte augmentation de 12% de plus de séjours en 2008 par rapport à 2007 s'explique par un changement des pratiques de codage [1]. Le nouveau règlement de financement a

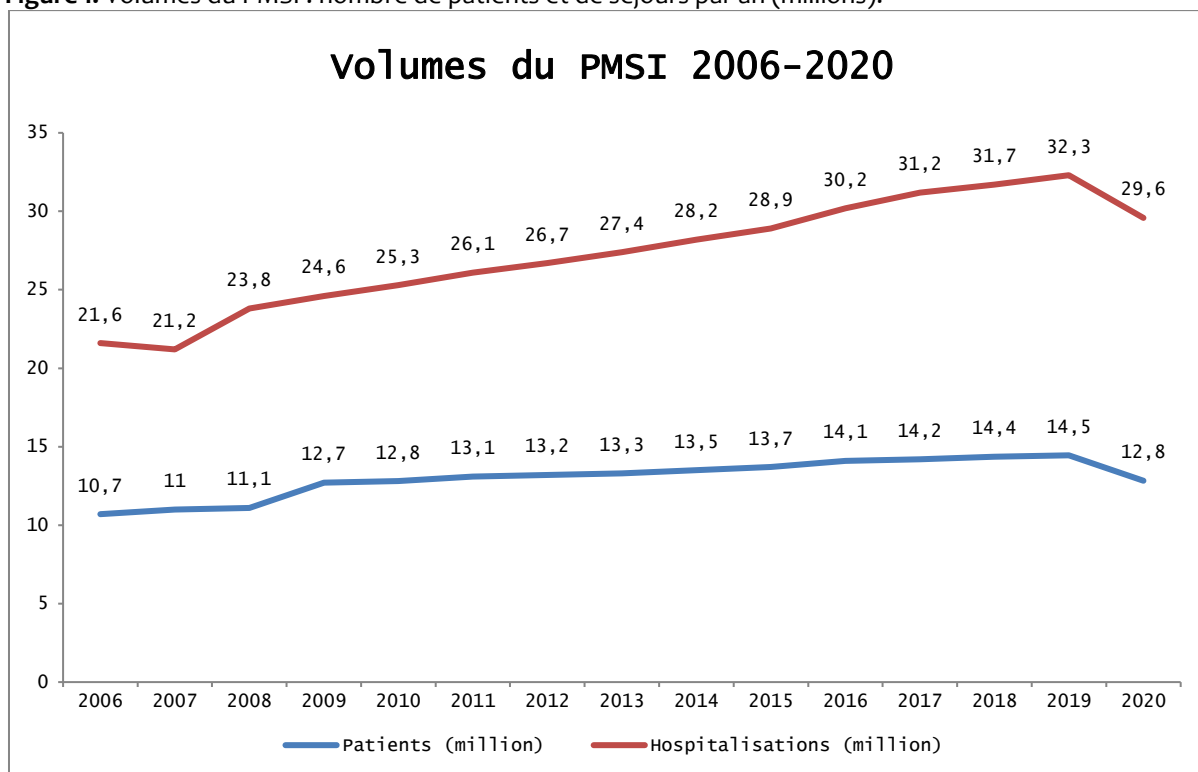
¹ Arrêté du 30 novembre 2009 modifiant l'arrêté du 22 février 2008 modifié relatif au recueil et au traitement des données d'activité médicale et des données de facturation correspondantes, produites par les établissements de santé publics ou privés ayant une activité en médecine, chirurgie, obstétrique et odontologie, et à la transmission d'informations issues de ce traitement dans les conditions définies à l'article L. 6113-8 du code de la santé publique

² Circulaire DHOS/PMSI n° 2001-106 du 22 février 2001 relative au chaînage des séjours en établissements de santé dans le cadre du programme de médicalisation des systèmes d'information (PMSI) Bulletin officiel du ministère chargé de la santé n° 2001/13 p. 197- 210.

³ Arrêté du 7 février 2011 modifiant l'arrêté du 22 février 2008 modifié relatif au recueil et au traitement des données d'activité médicale et des données de facturation correspondantes, produites par les établissements de santé publics ou privés ayant une activité en médecine, chirurgie, obstétrique et odontologie, et à la transmission d'informations issues de ce traitement dans les conditions définies à l'article L. 6113-8 du code de la santé publique.

fortement incité à rendre compte de manière exhaustive des activités hospitalières. Bien évidemment l'activité de l'année 2020 a été fortement impactée par la pandémie de Covid-19.

Figure 1. Volumes du PMSI : nombre de patients et de séjours par an (millions).



AXE I : TRAJECTOIRES DE SOINS

CONTEXTE GENERAL

L'analyse des parcours de soins est le premier axe de recherche. Le parcours de soins est défini comme une succession de recours au système de soins, en particulier dans le cas des maladies chroniques et des maladies nécessitant une prise en charge multidisciplinaire. L'analyse de ces parcours permet de comprendre le déploiement de la prise en charge sur le territoire, d'identifier les forces et faiblesses et anticiper des potentielles complications, en particulier après un geste chirurgical.

Nous avons initialement approché cette thématique pendant la thèse de doctorat avec une première étude sur les réhospitalisations après chirurgie bariatrique [3]. Notre travail avait analysé le recours à des soins hospitaliers non programmés dans la période postopératoire précoce, le taux de réhospitalisations étant considéré comme un indicateur de qualité des soins. Nous avons ensuite réalisé une deuxième étude sur le thème des réhospitalisations, en analysant l'association entre mortalité postopératoire et destination de réadmission : dans la période post-chirurgie précoce en cas de réhospitalisations, les patients peuvent être réadmis dans l'hôpital où ils viennent de se faire opérer ou être réadmis dans un autre établissement. Cette différence de destination n'est pas anodine en cas de complications, car la fragmentation des soins avec une hospitalisation dans un nouveau centre, a un impact négatif sur la mortalité postopératoire [4].

Ces premières approches nous ont introduits à l'analyse plus large des parcours de soins, domaine qui s'est considérablement développé ces dernières années avec l'importation de méthodes utilisées initialement dans d'autres domaines de recherche, tels que les sciences sociales.

En effet, afin d'analyser des données longitudinales sur un enchaînement de soins, une méthodologie adaptée prenant en compte la dimension temporelle doit être employée. Parmi les méthodes actuellement existantes pour analyser les parcours de soins, nous en étudions deux : l'analyse de séquences [5], [6], et l'analyse formelle des concepts [7].

Le traitement du cancer et la prise en charge de l'obésité présentent des parcours de soins complexes, articulés, faisant appel à de nombreux intervenants au cours du temps. Ces deux pathologies, cancer et obésité, semblent être des cas d'étude appropriés pour l'analyse des trajectoires de soins.

ANALYSE FORMELLE DE CONCEPTS

Nous avons utilisé cette première méthode en collaboration avec l'équipe du Pr. Katsahian, dans le cadre du doctorat de Mme Charles-Nelson. Nous avons appliqué cette méthode à l'analyse des événements non programmés après chirurgie bariatrique.

L'analyse formelle de concept (AFC) a été utilisée pour regrouper hiérarchiquement les objets en fonction de leurs attributs communs [8]. Dans cet exemple, tiré du travail de A. Napoli [9] on décrit les caractéristiques de différentes planètes, qui constituent les objets, et les variables descriptives (*size*, *distance from the sun*, *moon*) constituent les attributs.

Table 1. Attributs des différents objets.

The initial context of planets:

Planet	Size	Distance to Sun	Moon(s)
Jupiter	large	far	yes
Mars	small	near	yes
Mercury	small	near	no
Neptune	medium	far	yes
Pluto	small	far	yes
Saturn	large	far	yes
Earth	small	near	yes
Uranus	medium	far	yes
Venus	small	near	no

Un groupe d'objets partageant un ensemble commun d'attributs est appelé un concept. Ainsi, la méthode AFC trouve et représente tous les concepts et dépendances dans les données d'entrée tabulaires [7]. Les données d'entrée sont sous une forme matricielle booléenne, appelée contexte formel, où chaque ligne représente un objet et chaque colonne représente l'un des attributs définis [7]. C'est une matrice remplie de 0 et 1, où 1 signifie qu'un objet présente un attribut.

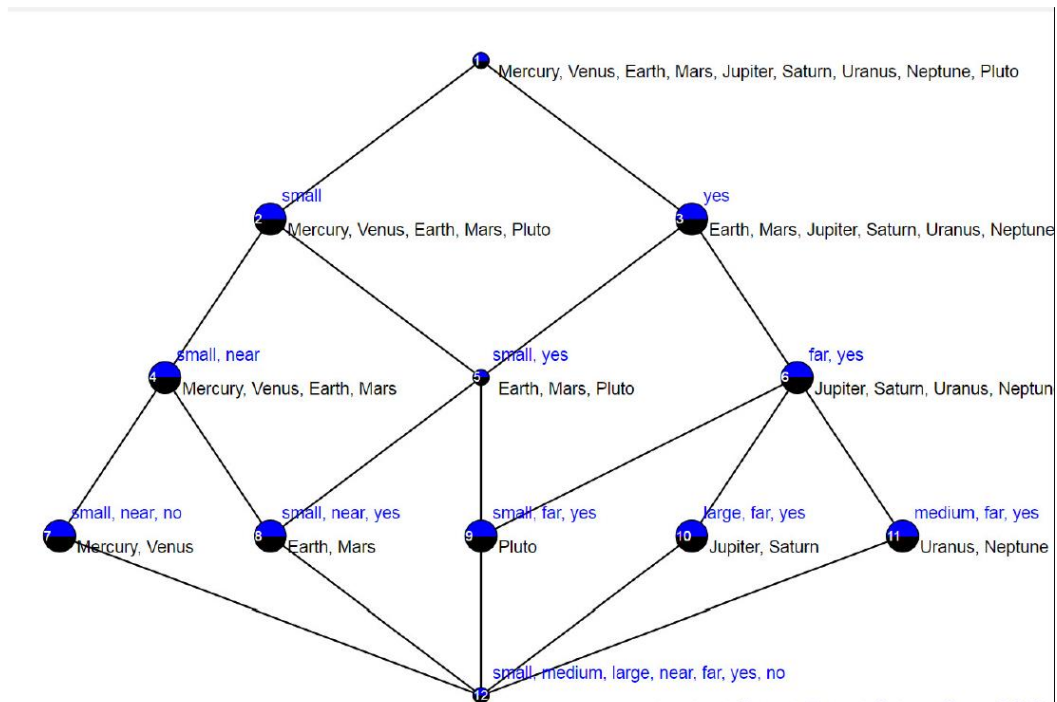
Table 2. Matrice booléenne des attributs.

The context of planets after plain scaling:

Planets	Size			Distance to Sun		Moon(s)	
	small	medium	large	near	far	yes	no
Jupiter			x		x	x	
Mars	x			x		x	
Mercury	x			x			x
Neptune		x			x	x	
Pluto	x				x	x	
Saturn			x		x	x	
Earth	x			x		x	
Uranus		x			x	x	
Venus	x			x			x

Ensuite, chaque contexte formel est transformé en une structure mathématique appelée treillis de concepts.

Figure 2. Treillis final avec la représentation des différents concepts.



L'AFC donc est une méthode d'analyse de données et de représentation des connaissances qui traite les informations sous forme de hiérarchies de concepts construites à partir de tables binaires [10]. C'est une méthode de classification non supervisée capable de découvrir et d'organiser les groupements naturels d'un ensemble d'objets liés à un ensemble d'attributs.

On utilise donc l'AFC comme méthode de classification des trajectoires de soins de santé. Cette expérience offre deux avantages. Premièrement, cela aide à regrouper les trajectoires des patients partageant certains attributs et à produire leurs visualisations graphiques. Le deuxième avantage de l'utilisation de l'AFC est qu'elle est capable de caractériser des groupes de patients par plusieurs trajectoires, et enfin, elle aide à sélectionner les treillis conceptuels pour filtrer uniquement les trajectoires des soins intéressantes.

ANALYSE DE SEQUENCES

La seconde méthode explorée est fondée sur la notion de « séquences ». Pratiquement, une séquence se définit comme une liste d'éléments ordonnés, ces éléments pouvant être de n'importe quelle nature (événements, nombres, etc.). L'analyse de séquences (*sequence analysis*) est donc un corpus de techniques traitant les données sous forme de séquences.

Introduites à partir des années 1980 [11], ces méthodes sont à présent largement répandues dans le domaine des parcours de vie (*life course analysis*) et ont bénéficié de nombreuses avancées méthodologiques au cours du temps, permettant ainsi une plus large utilisation [12]–[14]. Elles sont principalement utilisées pour l'étude des trajectoires professionnelles, de la situation maritale et pour l'étude des comportements dans la population, comme l'impact d'un diagnostic sur les parcours de vie.

Les recherches utilisant l'analyse séquentielle tentent principalement de répondre à trois questions selon Andrew Abbott [15] :

- Existe-t-il des modèles (*patterns*), des séquences typiques, parmi un ensemble de séquences donné ?
- Lorsqu'elles existent, comment ces séquences typiques sont-elles produites, quels facteurs les déterminent ?
- Quelles sont les conséquences de ces séquences typiques ?

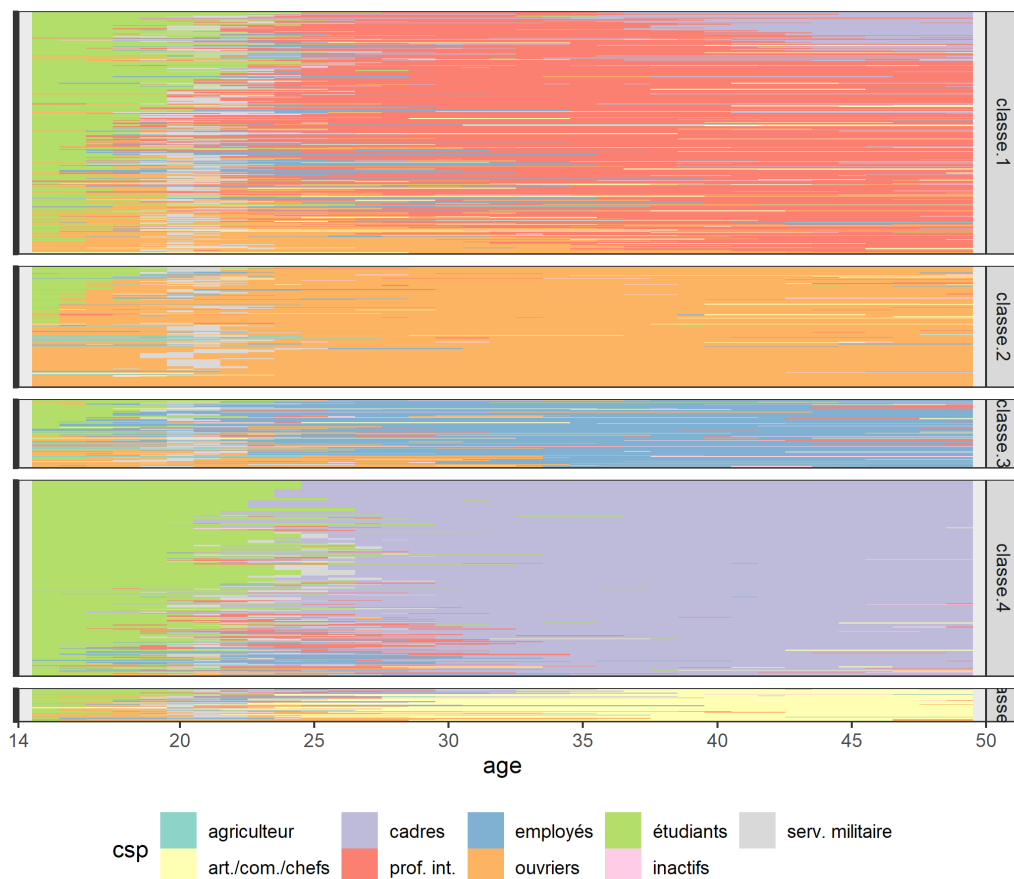
Dans le domaine de la santé, l'analyse de séquences commence à être utilisée par les chercheurs pour étudier, par exemple, la consommation de soins des femmes enceintes [16] ou la prise en charge de la sclérose en plaque [17].

Parmi les différentes méthodes d'analyse de séquences qui ont été proposées dans les dernières années, il y en a une qui a été employée et diffusée plus largement que les autres : l'Optimal Matching Analysis (OM). L'usage de l'OM s'est principalement développé en biologie moléculaire pour l'analyse des protéines et des séquences d'ADN. L'objectif est de rechercher dans d'importantes bases de données des séquences ressemblant à une séquence particulière, par exemple à une protéine donnée.

L'OM ne constitue en pratique qu'une étape d'une analyse de séquences. Son principe consiste à mesurer la dissemblance (ou dissimilarité) entre chaque paire de séquences constituant l'échantillon. Cela répond à la première question de l'analyse séquentielle, à savoir s'il existe des séquences typiques au sein de la base de données. En mesurant la dissimilarité entre différentes séquences calculées sert alors de point de départ à la seconde étape de la démarche d'analyse. Celle-ci consiste le plus souvent à construire une typologie de séquences, en créant des groupes de séquences similaires, à l'aide de techniques telles que les classifications hiérarchiques. La typologie peut ensuite être utilisée comme variable dépendante ou indépendante pour de nouvelles analyses, qui constituent la troisième et dernière étape de l'analyse séquentielle, afin de répondre à la question suivante : quelles sont respectivement les causes et les conséquences de l'existence de séquences typiques [18] ? Enfin, un avantage majeur de cette technique et du développement informatique qui l'a accompagné est la possibilité de visualiser de manière très intuitive les différentes trajectoires.

On montre ici un graphique appelé *index plot* (« tapis »), décrivant l'analyse de parcours de vie. Chaque ligne correspond à un individu, allant de l'âge de 14 ans à l'âge de 50 ans. Les différents parcours de vie ont déjà été catégorisées en 5 classes homogènes [19].

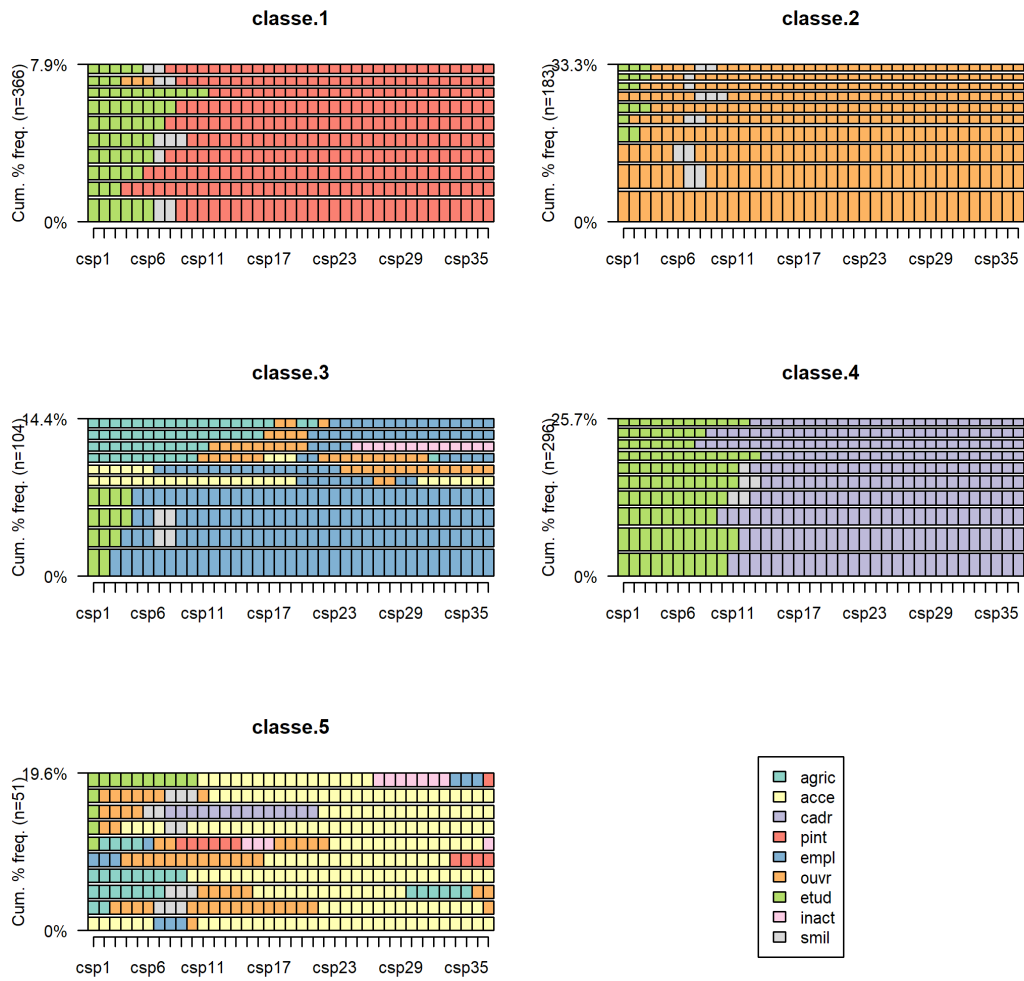
Figure 3. Trajectoires individuelles regroupées par classe (index plot).



La classe 2, par exemple, montre des individus qui étaient initialement étudiants (couleur verte), une partie d'entre eux a ensuite effectué le service militaire (en gris clair) et puis ils ont exercé une profession d'ouvrier pour le reste de leur vie (en orange). Dans la classe n° 4, les cadres (en violet) ont un parcours d'étudiant plus long que les ouvriers, avec un service militaire parfois plus tardif et une entrée dans la vie professionnelle plus tardive aussi.

Un autre type de visualisation est celui du chronogramme, qui présente les différentes classes séparément avec à l'intérieur de chaque classe les 10 séquences les plus fréquentes.

Figure 4. Chronogramme des 10 séquences les plus fréquentes par classe.



Cette méthode a été utilisée lors de l'encadrement de deux étudiants en Master 2 (Dr S. Goro et Dr S. Tzedakis).

RECHERCHE & ENCADREMENT

Année	Etudiant	Projet	Cursus	% d'encadrement
2019	Anaïs Charles-Nelson	Charles-Nelson A, Lazzati A, Katsahian S. Analysis of Trajectories of Care After Bariatric Surgery Using Data Mining Method and Health Administrative Information Systems. <i>Obes Surg.</i> 2020;30:2206–2216.	Doctorat ED 393 : Pierre Louis de santé publique. Sorbonne université	50%
2020-2021	Stylios Tzedakis	Analyse des trajectoires des patients atteints d'un cholangiocarcinome en France	Master 2 Santé publique - Parcours : Données massives en santé. Université de Paris	90%
2020-2021	Seydou Goro	Analyse de la trajectoire des patients ayant subi une chirurgie bariatrique.	Master 2 Santé publique - Parcours : Données massives en santé. Université de Paris	100%
2021-2024	Stylios Tzedakis	Facteurs épidémiologiques associés au pronostic du cancer des voies biliaires	Doctorat ED 393 : Pierre Louis de santé publique. Sorbonne	50%

ETUDE N°1

ANALYSIS OF TRAJECTORIES OF CARE AFTER BARIATRIC SURGERY USING DATA MINING METHOD AND HEALTH ADMINISTRATIVE INFORMATION SYSTEMS

ETUDE PUBLIEE

Charles-Nelson A, Lazzati A, Katsahian S.

Analysis of Trajectories of Care After Bariatric Surgery Using Data Mining Method and Health Administrative Information Systems.

Obes Surg. 2020;30:2206–2216.

RESUME

INTRODUCTION

Le taux de réadmission de 30 jours après la chirurgie bariatrique est considéré comme une mesure de la qualité des soins hospitaliers. Toutefois, le taux de réadmission au-delà de 30 jours est rarement décrit et ne fournit aucune information sur les trajectoires de soins, ce qui aurait un grand intérêt pour la planification des soins. L'objectif de cette étude était d'analyser les trajectoires de soins au cours de la première année suivant une chirurgie bariatrique sur une base de données à l'échelle nationale en utilisant des méthodes d'exploration de données.

METHODE

Il s'agit d'une étude descriptive rétrospective sur les trajectoires de soins dans la première année suivant la chirurgie bariatrique. Les données ont été extraites d'une base de données administrative nationale (la base de données PMSI) et des trajectoires ont été définies comme un diagnostic principal de réadmissions successives. L'analyse formelle de concepts a été réalisée pour trouver des concepts communs de trajectoires de soins.

RESULTATS

Nous avons inclus pour analyse 198.389 procédures bariatriques effectuées sur 196.323 patients. Douze concepts principaux ont été sélectionnés. Environ un tiers des patients (32,4%) ont été réadmis au cours de la première année après la chirurgie. Les trajectoires les plus courantes étaient les suivantes: suivi régulier (14,9%), pathologie lithiasique biliaire (2,2%), douleur abdominale (1,9%) et sepsis abdominale (1,3%). Des différences importantes ont été trouvées dans les trajectoires parmi les différentes procédures bariatriques: 1,8% des patients opérés d'anneau gastrique avaient des événements liés à la grossesse (accouchement ou avortement), tandis que nous observons un taux de réadmission pour le sepsis abdominal de 2,7% et 5,1% des patients opérés de bypass gastrique et sleeve gastrectomy respectivement.

CONCLUSION

Les données médico-administratives peuvent être analysées par l'analyse formelle de concepts afin de classer les trajectoires de soins. Cette approche permet de quantifier les complications postopératoires attendues et d'identifier des événements inattendus.

ABSTRACT

CONTEXT

The 30-day readmission rate after bariatric surgery is considered an important metric of the quality of hospital care. However, readmission rate beyond 30-days is rarely reported and does not provide any information about trajectories of care, which would be of great interest for healthcare planning. The aim of this study was to analyze trajectories of care during the first year after bariatric surgery on a nationwide basis using data mining methods.

METHOD

This was a retrospective descriptive study on the trajectories of care within the first year after bariatric surgery. Data were extracted from a national administrative claims database (the PMSI database) and trajectories were defined as principal diagnosis of successive readmissions. Formal concept analysis was performed to find common concepts of trajectories of care.

RESULTS

We included for analysis 198,389 bariatric procedures performed on 196,323 patients. Twelve main concepts were selected. About one third of patients (32.4%) were readmitted in the first year after surgery. Most common trajectories were: regular follow-up (14.9%), cholelithiasis (2.2%), abdominal pain (1.9%), and abdominal sepsis (1.3%). Important differences were found in trajectories among different bariatric procedures: 1.8% of gastric banding patients had pregnancy-related events (delivery or medical abortion), while we observed a readmission rate for abdominal sepsis in 2.7% and 5.1% of patients operated of gastric bypass and sleeve gastrectomy respectively.

CONCLUSION

Administrative claim data can be analyzed through Formal Concept Analysis in order to classify trajectories of care. This approach permits to quantify expected postoperative complications, and to identify unexpected events.

INTRODUCTION

Obesity is a major public health problem whose prevalence has dramatically increased in the last 25 years. Indeed, the number of obese patients has almost tripled since 1975 with more than 1.9 billion overweighted adults, and about 650 million of obese [20]. Obesity is known for being a major risk factor for some chronic diseases such as cardiovascular disease, type-2 diabetes, dyslipidemia, obstructive sleep apnea syndrome, musculoskeletal disorders, especially osteoarthritis and some cancers (endometrium, breast, ovarian, prostate, liver, gallbladder, kidney and colon). Studies have shown that behavioral and pharmacological therapies have a limited long-term efficacy on weight loss for patients with severe obesity [21]. Bariatric surgery has become a reliable complementary treatment for obesity. In the last two decade, a major development and improvement in bariatric procedures has been observed with about 500,000 procedures per year performed worldwide since 2013 [22].

It is accepted that bariatric surgery has a rate of early [23] complications between 5% and 10% [24], [25]. Several factors have been associated to the risk of major postoperative complications: the type of procedure [26], surgical skills and presence of chronic health problems [27], [28].

The 30-day readmission rate is considered as an important metric of the quality of hospital care. This indicator is of clinical interest for the patient and economical for the institutions. Thus, hospital readmission rates have been widely studied [29]–[31]. However, readmission rate beyond 30-days are less frequently reported. Doumouras et al reported a readmission rate of 6.1% over a period of 3 years [30] and Saunders et al reported 1-year readmission rate of 18% [32]. However, readmission rate provides no information about trajectories of care even though they are of great interest for healthcare planning. Indeed, understanding trajectories of care may help physician to anticipate future care and thus, may potentially anticipate severe complications, leading to a more personalized medicine. Trajectories of care have been studied in different domain such as breast cancer [33] and prenatal care [5]. In the field of obesity, however, only medical weight loss trajectories have been studied [34]–[36], but no study analyzed trajectories of care after bariatric surgery. Thus, as readmission is considered as a metric of quality of hospital care, a better understanding of trajectories may participate to enhance the quality itself.

In the last few years, several studies have used national registries or administrative databases to monitor their outcomes [37]–[39]. In France, the national administrative claims database (*Programme de Médicalisation des Systèmes d'Information* PMSI) includes all reimbursed surgical interventions performed in France in every hospital. It was introduced in the mid-1980s and was first presented as an epidemiological tool before being used for budget allocation. It is also allowed the promotion exchange between hospital partners: doctors, nurses, administrators. The collection and processing of data for the short stay are directly inspired by the American Classification of Diagnosis Related Groups (DRG) [40]. This classification classifies hospital stays in groups with a double homogeneity in medical and economic terms. The main advantage of the PMSI database is that data collected gives comprehensive information from all surgical centers and can be used to follow any single patient across different hospitals. Because of its exhaustiveness, the PMSI database is considered a big dimension dataset. A major challenge concerning big datasets is the identification of pertinent information. In the last few years, data mining methods have been recognized as an efficient tool to treat large volume data [41]. Data Mining involves drilling, exploring or delving into data. Unlike conventional analyses, Data Mining allows to explore associations and relationships between data, which are hidden or not obvious due to large volumes of data. Thus, it provides a summary of big dataset and makes the information usable and understandable for decision-making.

The aim of this study was to analyze trajectories of care within the first year after bariatric surgery on a nationwide basis using data mining methods.

MATERIAL AND METHODS

DATA SOURCE

This was a retrospective descriptive study on the trajectories of care within the first year after bariatric surgery. Data were extracted from the PMSI database. Data are exhaustive as the database includes all reimbursed surgical procedures performed in France, in any hospital regardless of their academic affiliation or ownership (public and private for-profit and private non-profit). In the PMSI database, data are collected as standardized discharge reports and includes patient demographic data (age, gender, zip code, entry, and release dates), primary and associated diagnoses based on the International Classification of Disease, 10th edition (ICD-10), and therapeutic procedures based on the Common Classification of Medical Acts (Classification Commune des Actes Médicaux, CCAM, 11th edition), which is a national standardized classification of medical procedures [42]. A unique and anonymous identifier is assigned to each patient thus making it possible to identify all of his or her hospital stays planned or not in any hospital in France. Patient consent is not required as the individual information is anonymous and publicly available.

PARTICIPANTS

We included all adult patients (≥ 18 years and age ≤ 60 years) with a principal diagnosis of obesity who underwent one of the most common bariatric procedure performed in France: adjustable gastric banding (AGB), gastric bypass (GB), or sleeve gastrectomy (SG). We retrieved information from the PMSI database from January 1, 2013, to December 31, 2018, and included patients operated before 2018 in order to have one year of potential follow-up for each patient.

Patients were selected matching the CCAM codes for bariatric procedures and the ICD-10 code of obesity. The codes HFMA009 and HFMC007 were used to identify AGB, the codes HFCA001 and HFCC003 for GB and the codes HFFA011 and HFFC018 for SG. The codes E66x were used to identify obesity.

We excluded patients that had ambiguous procedures coding and patients presenting a coding error or a missing value on the principal diagnosis information.

OUTCOMES

We were interested in the trajectories of care using the primary diagnosis within one year from discharge after bariatric surgery took place. Trajectories are defined as a sequence of hospitalizations starting within one year after the bariatric surgery, which is the first element of the trajectory irrespective of the length of the initial stay.

CODING OF VARIABLES

Length of stay was categorized into three classes: 0 to 1 day (capturing ambulatory and outpatient surgery patients), 2 to 7 days and > 7 days. One day of hospital stay means that the patient spent one night at the hospital irrespective of the total number of hours. The cut-off of 7 days was selected according to the American Society for Metabolic and Bariatric Surgery (ASMBS) definition [43] which considered a length of stay greater than 7 days as a major complication. However in our study, this cut off was only used to categorize the length of stay into three categories.

Demographic data included age and gender. The body mass index (BMI) is classified into four categories 30 to 40 kg/m², from 40 to 50 kg/m², >50 kg/m², and patients with BMI unspecified were excluded from the analysis. Comorbidities were assessed using the Charlson comorbidity index, using the version of Quan and colleagues [44]. The final score was categorized into three groups (0, 1–2, and ≥ 3). Obstructive sleep apnea syndrome (OSAS) was included as a separate covariate. Missing data about comorbidities or OSAS at the hospital stay where the bariatric took place, were imputed using the last observation carried forward method, as previously reported [45]. Primary diagnoses are coded using the first 3 digits of ICD-10 code with a few exceptions. Firstly, the codes starting with the digit “Z”, which indicate the admission for regular follow-up, were grouped in one category. Secondly, the code K316 (Fistula of stomach and duodenum) was

kept in its full length and grouped with the code K65 (Peritonitis) for better consistency. We will refer to this group as “abdominal sepsis”. Finally, we also kept four digits for the code K910 (Vomiting following gastrointestinal surgery) as other K91x diagnoses were poorly consistent (Table 1).

Table 1. Main principal diagnoses codes.

Grouped Code	Codes	Designation
E66	E660, E6600, E6601, E6602, E6609, E661, E6610, E6611, E6612, E6619, E662, E6620, E6621, E6622, E6629, E668, E6680, E6681, E6682, E6689, E669, E6690, E6691, E6692, E6699, E6603, E6613, E6683, E6693, E6604, E6605, E6606, E6607, E6614, E6615, E6616, E6617, E6624, E6625, E6626, E6627, E6684, E6685, E6686, E6687, E6694, E6695, E6696, E6697	Obesity
I83	I830, I831, I832, I839	Varicose veins of lower extremities
K22	K220, K221, K222, K223, K224, K225, K226, K227, K228 K229	Other diseases of esophagus
K25	K250, K251, K252, K253, K24, K255, K256, K257, K259	Gastric ulcer
K28	K280, K281, K282, K283, K284, K285, K286, K287, K289	Gastrojejunal ulcer
K29	K290, K291, K292, K293, K294, K295, K296, K297, K298, K299	Acute hemorrhagic gastritis
K316- K65	K316, K650, K658, K659	Fistula of stomach and duodenum and Acute peritonitis
K43	K430, K431, K432, K433, K434, K435, K436, K437, K439	Ventral hernia
K56	K560, K561, K562, K563, K564, K565, K566, K567	Paralytic ileus and intestinal obstruction without hernia
K63	K630, K631, K632, K633, K634, K635, K635+0, K635+8, K638, K639	Other diseases of intestine
K80	K800, K801, K802, K803, K804, K805, K808	Cholelithiasis
K910		Vomiting following gastrointestinal surgery
K92	K920, K921, K922, K928, K929	Other diseases of digestive system
L02	L020, L021, L022, L023, L024, L028, L029	Cutaneous abscess, furuncle and carbuncle
N20	N200, N201, N202, N209	Calculus of kidney and ureter
N23		Unspecified renal colic
O04	O04, O040, O041, O042, O043, O044, O045, O046, O047, O048, O049, O0400, O0401, O0402, O0403, O0410, O0411, O0412, O0413, O0420, O0421, O0422, O0423, O0430, O0431, O0432, O0433, O0440, O0441, O0442, O0443, O0450, O0451, O0452, O0453, O0460, O0461, O0462, O0463, O0470, O0471, O0472, O0473, O0480, O0481, O0482, O0483, O0490, O0491, O0492, O0493,	Medical abortion
O80	O800, O801, O802, O808, O809	Single spontaneous delivery
R10	R100, R101, R102, R103, R104	Abdominal and pelvic pain
R11		Nausea and vomiting
R13	T858, T859	Dysphagia
T85	T850, T851, T852, T853, T854, T855, T8550, T8558, T856, T857, T858, T859	Complications of other internal prosthetic devices, implants and grafts
Zx		Regular follow-up

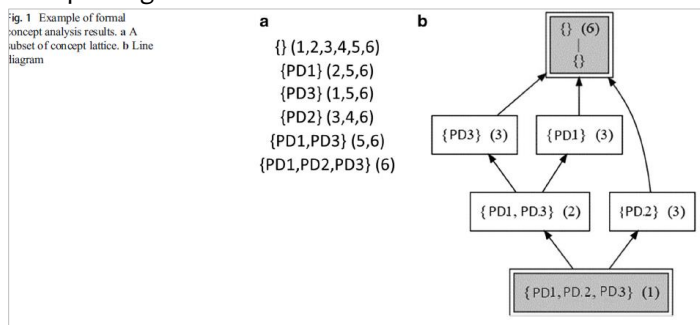
STATISTICAL ANALYSES

Continuous demographic and clinical characteristics are presented as mean \pm one standard deviation if the parameter follows a normal distribution and median with interquartile range if the distribution is not normal for quantitative parameters. Qualitative variables are presented as numbers and proportions.

Formal concept analysis (FCA) was used to hierarchically group objects according to their common attributes [8]. A group of objects sharing some common set of attributes is called a concept. Thus, the FCA method finds and represents all concepts and dependencies in the tabular input data [7]. Input data are in a

Boolean matrix form, called formal context, where each row represents an object, and each column represents one of the defined attributes [7]. It is a matrix fulfilled with 0 and 1, where 1 means that an object presents an attribute. In this study, bariatric surgeries and principal diagnoses (PD) of readmission are objects and attributes respectively. Figure 1 shows an example of a Boolean matrix adapted to this study. FCA method transforms a formal concept into a mathematical structure called concept lattice. It represents all concepts of a formal context with the order relation, representing all possible combinations of diagnoses observed in patient trajectories. 1A shows a subset of an example of a concept lattice. It is illustrated using a line diagram, representing hierarchical relationship of all the found concepts and a list of all found attributes implications in the formal context [7]. 1B represents the line diagram of the concept lattice of the example.

Figure 1. Example of formal concept analysis results: A) A subset of concept lattice; B) Line diagram. PD: Principal diagnosis.



Due to the high number of diagnosis resulting in a high number of concepts, they are filtered according to proportion of own objects and following a stepwise method. Due to a lack of consensus, thresholds for proportion of objects to select concepts are arbitrary chosen. The first step selected concepts combining PD of obesity with a second PD. Concepts were selected if it represented at least 0.4% of the number of surgeries. The second step consists in selecting concepts combining concepts previously selected with another PD. Concepts were selected if it represented at least 0.2% of the number of surgeries. Finally, concepts combining obesity and three other principal diagnoses were selected during the last step. Combinations were selected if they represented at least 0.1% of the number of surgeries.

Concepts observed in trajectories were represented using an organigram where each level corresponds to a different number of combinations of principal diagnosis. The top of the diagram is the diagnosis of obesity. The second level corresponds to diagnosis of obesity combined with one other principal diagnosis. The third level corresponds to combination of obesity with two other principal diagnoses. And the last level corresponds to concept combining obesity with three other PD.

Table 2. Example of Boolean matrix in the event of a bariatric surgery.

Surgery	PD*1	PD2	PD3
1	0	0	1
2	1	0	0
3	0	1	0
4	0	1	0
5	1	0	1
6	1	1	1

*PD, principal diagnosis

Trajectories were analyzed using the Coron System (available at <http://coron.loria.fr>). Other analyses were performed using R version 3.4.4. (R172 Foundation for Statistical Computing, Vienna, Austria). Data are

reported according to the Reporting of studies Conducted using Observational Routinely collected Data statement (RECORD statement) [46].

RESULTS

We identified 240,821 bariatric surgeries between 2013 and 2017. Among these 42,432 (17.6%) did not meet inclusion criteria and were excluded: 13,012 (5.4%) were performed on adolescents or elderly patients, 22,352 (9.3%) presented a coding error on the principal diagnosis or the type of surgery or coding ambiguity, 3,905 (1.6%) were bariatric procedures other than AGB, GB or SG and 3,163 (1.3%) did not have a main diagnosis code of obesity. Hence 198,389 bariatric procedures, performed on 196,323 patients, were included for analysis. Among surgical procedures 13,744 (6.9%) were AGB, 55,945 (28.2%) were GB and 128,700 (64.9%) were SG. Most of patients (90%) had only one surgery. Patients are mostly woman (80.7%) with morbid obesity (70.0% with a BMI greater than 40 kg/m²). The Charlson Comorbidity Index was null for 85.4% of patients. Finally, 30.7% had an obstructive sleep apnea syndrome (Table 3). AGB was mostly used on younger patients (62.6% aged between 18 and 39 years) and with a BMI lower than 40 kg/m² (48.0%). GB and SG were performed on older patients (42.3% and 51.2% are aged between 18 and 39 years respectively) with morbid obesity (respectively 73.9% and 70.3% had a BMI greater than 40 kg/m²).

Table 3. Baseline characteristics at the time of the bariatric surgeries.

Covariate		Whole sample (N=198,389)	AGB (N=13,744)	GB (N=55,945)	SG (N=128,700)
Gender	Female	160,178 (80.7%)	11,663 (84.9%)	46,143 (82.5%)	102,372 (79.5%)
Age (years)	18-29	42,007 (21.2%)	4,542 (33.1%)	8,730 (15.6%)	28,735 (22.3%)
	30-39	56,092 (28.3%)	4,067 (29.6%)	14,922 (26.7%)	37,103 (28.8%)
	40-49	57,661 (29.1%)	3,277 (23.8%)	17,801 (31.8%)	36,583 (28.4%)
	50-60	42,629 (21.5%)	1,858 (13.52%)	14,492 (25.9%)	26,279 (20.4%)
BMI (kg/m ²)	30-40	55,151 (30.0%)	6,290 (48.0%)	13,627 (26.1%)	35,234 (29.7%)
	40-50	109,025 (59.3%)	6,169 (47.0%)	32,601 (62.5%)	70,255 (59.2%)
	≥50	19,815 (10.8%)	654 (5.0%)	5,956 (11.4%)	13,205 (11.1%)
	NA	1,4398	631	3761	10,006
Charlson Comorbidity Index	0	169,492 (85.4%)	12,760 (92.8%)	46,416 (83.0%)	110,316 (85.7%)
	1-2	25,281 (12.7%)	860 (6.3%)	8,099 (14.5%)	16,322 (12.7%)
	≥3	3,616 (1.8%)	124 (0.9%)	1,430 (2.6%)	2,062 (1.6%)
OSAS*		60,832 (30.7%)	1,882 (13.7%)	18,517 (33.1%)	40,433 (31.4%)
Surgical Approach	Laparoscopy	197,266 (99.4%)	13,701 (99.7%)	55,587 (99.4%)	127,978 (99.4%)
	Open	1,123 (0.6%)	43 (0.3%)	358 (0.6%)	722 (0.6%)
Hospital ownership	Private, for profit	120,662 (61.3%)	11,121 (81.6%)	32,921 (59.6%)	76,620 (59.9%)
	Private, nonprofit	10,190 (5.2%)	342 (2.5%)	2,701 (4.9%)	7,147 (5.6%)
	Public	65,943 (33.5%)	2,160 (15.9%)	19,603 (35.5%)	44,180 (34.5%)
	NA	1594	121	720	753
LOS [†] (days)	0-1	11,055 (5.6%)	5,758 (41.9%)	1,718 (3.1%)	3,579 (2.8%)
	2-7	176,470 (89.0%)	7,965 (58.0%)	50,215 (89.8%)	118,290 (91.9%)
	>7	10,864 (5.5%)	21 (0.2%)	4,012 (7.2%)	6,831 (5.3%)
Year of surgery	2013	35,541 (17.9%)	4,614 (33.6%)	10,513 (18.8%)	20,414 (15.9%)
	2014	39,339 (19.8%)	3,678 (26.8%)	11,286 (20.2%)	24,375 (18.9%)
	2015	39,965 (20.1%)	2,373 (17.3%)	11,360 (20.3%)	26,232 (20.4%)
	2016	43,184 (21.8%)	1,880 (13.7%)	11,849 (21.2%)	29,455 (22.9%)
	2017	40,360 (20.3%)	1,199 (8.7%)	10,937 (19.6%)	28,224 (21.9%)
Reoperation	Yes	3,039 (1.5%)	11 (0.1%)	1,230 (2.2%)	1,798 (1.4%)

*OSAS: Obstructive Sleep Apnea Syndrome, BMI: Body Mass Index, †LOS: Length of stay

MAIN CONCEPTS

The formal concept had 1,059 attributes (PD) and the lattice results had 14,617 concepts. Twelve main concepts were selected using the algorithm described above (

Figure 2A). About one third of patients (64,350) have been readmitted at least once in the first year after surgery, among whom 45.8% have been readmitted for regular follow-up, 6.8% for cholelithiasis, 6.0% for abdominal pain, 4.0% for abdominal sepsis, 2.3% for medical abortion, 2.2% for hernias of the anterior abdominal wall, 1.9% for gastritis and duodenitis and 1.3% for dysphagia. Among patients readmitted for regular follow up, 2.0% were successively readmitted for cholelithiasis, 2.8% for abdominal pain and 4.2% for abdominal sepsis (gastric leak or peritonitis). According to the type of surgery, patients who had an AGB had the lowest readmission rate (23.4%) compared to the GB (38.2%) and SG (30.89) patients.

The lattice result has 1,209 concepts for AGB, 6,726 concepts for GB procedure and 8,429 concepts for SG. The number of selected concepts differs according to the type of procedure with 7 concepts for AGB, 10 concepts for the sleeve and 20 for the bypass procedures (2B, 2C, 2D). Four concepts are common to all types of surgery (

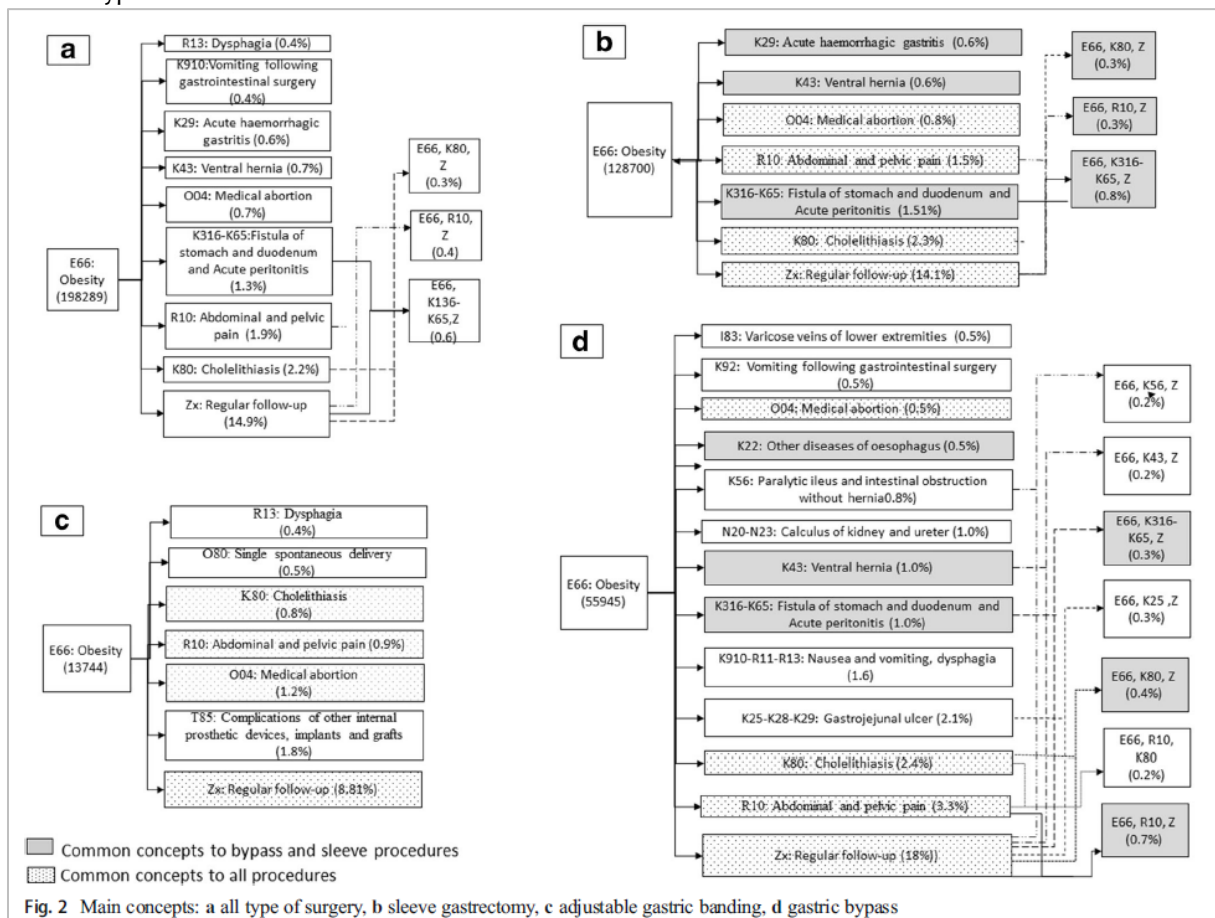
Figure 2). Among readmitted patients, patients operated of GB and SG were more frequently readmitted for regular follow-up, corresponding to the concept (E66, Zx) found in 37.5%, 47.1% and 45.8% for respectively the AGB, GB and SG. Patients with a GB were more commonly readmitted for abdominal pain (8.6%), corresponding to the concept (E66, R10), than other procedures, 3.9% for the AGB and 4.7% for SG.

Among readmitted patients, the concept (E66, K80), corresponding to cholelithiasis, was also present for the three types of surgeries but more common for SG (7.4%) compared to the AGB (3.6%) and GB (6.2%). The rate of medical abortion, corresponding to the concept (E66, O04), was higher in patients with the AGB (4.9%).

The concepts (E66, K43, abdominal wall hernia), (E66, K29, gastro-duodenitis) and (E66, K316-K65, abdominal sepsis) seem to be more specific to both GB and SG, among readmitted patients. However, patients were more readmitted for hernias of the anterior abdominal wall (K43) after GB (2.7%) than after SG (2.0%) while they were more readmitted for abdominal sepsis (K316-K65) for SG (5.0%) than GB (2.7%).

Some concepts were selected by the algorithm for only one type of surgery. For instance, the concepts (E66, O80, delivery) and the concepts (E66, T85, mechanical complication of a device) have been selected only in patients with an AGB device.

Figure 2. Main concepts: A) All type of surgery; B) Sleeve gastrectomy; C) Adjustable gastric banding; D) Gastric Bypass.



PATIENT PROFILE

Patients readmitted for regular follow-up had more comorbidities than the total sample with a Charlson index > 3 of 5.1% versus 1.8% and OSAS of 49.3% versus 30.1% (Table 4). Patients in the abdominal sepsis concept (codes K316-K65) have a longer length of stay than other concepts. Patients with the code K43 (abdominal wall hernia) have the highest proportion of open procedures (3.6%), and a higher rate of reoperation during the initial bariatric procedure (5.7%).

DISCUSSION

We have applied the FCA method on data extracted from the PMSI database to analyze trajectories of care within one year after bariatric surgery. The use of the PMSI database has several advantages. The first is the access to information of readmissions of the entire population in all health French centers leading to reduction of the selection bias. The second is the ability to identify readmissions of patients in the non-index hospital. Finally, an official guide of data coding is used to record and code data, which provides consistency, quality and homogeneity of the data.

Formal Concept Analysis has been applied in many fields such as medicine, biology, genetics, sociology and even to visualize radicalization trajectories over time [32], [47]–[50]. In this study, FCA has been used to find clusters of patients having similarities in trajectories of care. Indeed, a better understanding of trajectories of care may help the clinician to anticipate future complications and thus provide better and earlier medical care. The FCA method is simple to implement and only requires a binary table. The PMSI provides massive amount of data and the FCA method can deal with these kinds of database. Furthermore, being an

unsupervised method, it does not need a training sample. The main advantage of the FCA is that the formal concept can be visualized. However, one limitation of the FCA method is that it provides hierarchical concepts. Thus, concepts are non-disjoint, and a patient can belong to several concepts.

In the analysis of trajectories of care after bariatric surgery, the first result is that almost one third of patients are readmitted in the first postoperative year. Even if we included planned and unplanned readmissions, this rate is considerably higher than previously reported [38], [39]. This observation suggests that complications requiring hospital care occur during the entire first year after surgery and not only in the immediate postoperative period.

We also observed some expected trajectories, as the concept (K316-K65), which included gastric leak and peritonitis, which are typical postoperative complications for GB and SG. Also, the concept (R10), abdominal and pelvic pain, is found for the three bariatric procedures, and it has already been reported as a common cause for readmission [45].

The FCA allowed outlining also a few unexpected trajectories, as the concept (O04), medical abortion, the concept (K80), cholelithiasis and the concept (K43), abdominal wall hernia. The issue around pregnancy raises questions on the preoperative care of female candidates to bariatric surgery. In fact, national guidelines for bariatric surgery include a birth control planning before bariatric surgery in order to prevent pregnancy at least during the first year after surgery [51]. When we consider any PD related to pregnancy, we observe about 2.1% of pregnancies in the first year after bariatric surgery. This information indicates the need to reinforce the strategy of information about contraception in the preoperative care.

The problem of gallstones (concept K80) is well known as obesity itself and rapid weight loss have been identified as risk factors for gallstone formation [52], [53]. Several meta-analysis also suggest that administration of ursodeoxycholic acid after bariatric surgery prevent gallstones formation [54], [55]. In this study we found that more than 2% of patients needed a hospital stay for cholelithiasis, being the first and second reason for unplanned readmission after SG and GB respectively. Even if that goes beyond the aim of present study, it would be of great interest to assess the severity of the disease in this cohort as it has been reported that about 25% of gallstone carriers develop complications [56].

Our analysis also found that about 1% of patients are readmitted for an abdominal wall hernia related problem. This topic, which is quite common in bariatric population, has been recently discussed by the ASMBS with a consensus guideline [57], which conclude that in “patients with severe obesity and ventral hernia, and both being amenable to laparoscopic repair, combined hernia repair and bariatric surgery may be safe and associated with good short-term outcomes and low risk of infection”. Authors also outline that at present “there is a relative lack of evidence, however, about the use of synthetic mesh in this setting”.

LIMITATIONS

This study has several limitations. Coding errors or coding bias on the principal diagnostic of the readmission are encountered in the database. Moreover, the principal diagnosis has been used to determine the reason of the readmission. The principal diagnosis implies only one reason, but cause of readmission may be more complex and multifactorial. In total, 9.3% of surgeries were removed from the analysis because of coding error. Surgeries removed from the analysis were comparable to surgeries that were included except for the type of hospital. Indeed, it seems that there are more errors in private than in public hospitals, hence 7.3%, 0.2% and 1.7% were removed respectively from private for profit hospitals, private not for profit hospital and public hospitals, respectively.

CONCLUSION

In this study we assessed the trajectories of care after bariatric surgery through a Formal Concept Analysis. This method permitted to identify trajectories associated either to postoperative complications (as

cholelithiasis or abdominal wall hernia) or unexpected events like pregnancy. This approach could help health professional to improve the clinical pathway before and after bariatric surgery.

Etude N°2

ANALYSE DE SEQUENCES DE CHIRURGIE ITERATIVE EN CHIRURGIE BARIATRIQUE SUR LE LONG TERME
ETUDE EN COURS (MASTER 2 – DR S. GORO, 2020-2021)

RESUME

INTRODUCTION

La chirurgie bariatrique s'est affirmée comme traitement efficace de l'obésité morbide. Environ 300.000 interventions ont été réalisées en France dans la dernière décennie. L'évolution des techniques chirurgicales et la recherche de meilleurs résultats mènent à des interventions itératives (chirurgie de révision), estimées autour de 14%. Cette étude a le but d'analyser les séquences de chirurgie bariatrique de révision et d'en décrire les facteurs associés.

METHODES

Il s'agit d'une étude de cohorte rétrospective basée sur la base de données des hospitalisations (PMSI). Cette source comprend tous les données de remboursement des séjours hospitaliers en France. Nous avons donc inclus les patients opérés de 3 techniques de chirurgie bariatrique : anneau gastrique ajustable (AGA), sleeve gastrectomy (SG) et bypass gastrique (BPG). La durée de suivi fixée était de 10 ans. L'analyse des actes de chirurgie bariatrique a été réalisée par la technique de l'analyse des séquences (*sequence analysis*) ; l'analyse des facteurs associés a été réalisée un modèle logistique multinomial. De manière exploratoire nous avons appliqué l'analyse des séquences à l'analyse de complications postopératoires.

RESULTATS

Sur un total d'environ 27.000 patients, 29% ont été opérés d'AGA, 43% de SG et 28% de BPG. Au total 4600 patients (17%) ont eu une chirurgie de révision, dont 42% des patients opérés initialement d'AGA, 9% après SG et 1% après BPG. Nous avons analysé les séquences des patients opérés d'AGA et retrouvé 5 trajectoires principales : (1) anneau en place, (2) anneau-ablation, (3) anneau-sleeve, (4) anneau-bypass et (5) anneau-anneau. Parmi les facteurs associés, l'IMC initial semble être le seul associé à une chirurgie itérative après AGA. Chez les patients opérés initialement de SG, nous avons retrouvé 5 trajectoires : (1) sleeve seule, (2) sleeve-bypass, (3) sleeve-sleeve, (4) sleeve-dérivation BP et (5) sleeve-décès. Dans ce groupe de patients, l'âge, le sexe, l'IMC et les comorbidités étaient associés à un risque augmenté de chirurgie de révision. L'analyse des complications par la méthode des séquences n'a pas été contributive car les données étaient peu interprétables.

DISCUSSION

L'analyse des trajectoires des patients ayant bénéficié de la chirurgie bariatrique par la méthode des séquences permet de mieux comprendre l'histoire naturelle de cette chirurgie. La méthode utilisée est intuitive, simple et donne un bon aperçu visuel de la trajectoire du patient. Elle doit donc être plus envisagée dans l'analyse des trajectoires lorsque « l'état » du patient peut être considéré comme un état permanent. En revanche dans le cas des complications elle nous donne des résultats peu informatifs. Cette méthode ne semble donc pas appropriée pour l'analyse d'états ponctuels et transitoires.

INTRODUCTION

Le nombre d'interventions de chirurgie bariatrique a considérablement augmenté ces dernières décennies en France. Chaque année, environ 40.000 interventions sont réalisées.

L'évolution des techniques chirurgicales et la recherche de meilleurs résultats mènent à des interventions itératives. La chirurgie de révision est actuellement estimée à 14% des actes de chirurgie bariatrique. De plus ce type de chirurgie peut être responsable de complications à long terme, peu connues.

Le but de cette étude sera d'identifier les différentes trajectoires de soins des patients ayant été opérés d'une chirurgie bariatrique à l'échelle nationale afin de définir plusieurs clusters de patients, en utilisant la méthode d'analyse des séquences [11]. Dans un second temps nous identifierons les différents facteurs associés à ces trajectoires. Nous testerons également en analyses supplémentaires l'intérêt de l'analyse des séquences pour identifier les trajectoires des patients en termes de complications.

MÉTHODOLOGIE

DESIGN

Il s'agit d'une étude de cohorte rétrospective. La population de cette étude était constituée de patients ayant bénéficié de la chirurgie bariatrique en 2011. Les données recueillies concernent la période 2011-2020, ainsi tout patient aura un suivi théorique de 10 ans (2011-2020).

POPULATION

Les critères d'inclusion :

- Age > =18 ans,
- Patients dont la première de chirurgie bariatrique est l'une des 3 techniques de chirurgie bariatrique: l'anneau gastrique ajustable (AGA), la gastrectomie en gouttière ou sleeve gastrectomy (SG), et le by-pass gastrique (BPG).

Critère de non-inclusion :

- Les patients ayant bénéficié de la dérivation bilio-pancréatique (DBP) comme première intervention et ceux dont le type de chirurgie était indéterminé n'ont pas été inclus.

SOURCE

Les données ont été extraites de la base du Programme de médicalisation des systèmes d'information (PMSI). Les actes de chirurgie bariatrique possèdent des codes CCAM univoques. Nous avons donc utilisé ces codes CCAM pour identifier la population d'intérêt. Anneau gastrique ajustable : HFMA009, HFMC007, HFKC001, and HFKA002. Sleeve gastrectomy: HFMA010, HFMC006, HFFA011, and HFFC018. Bypass gastrique: HFCC003 and HFCA001

OUTCOMES

L'objectif principal de l'étude est descriptif : identification et classification des trajectoires de chirurgie bariatrique itérative. En objectif secondaire nous avons analysé les facteurs associés aux différentes trajectoires. De manière exploratoire nous avons appliqué l'analyse des séquences aux complications chirurgicales.

VARIABLES

Les covariables de notre étude sont des facteurs cliniques (score de Charlson, le syndrome d'apnée de sommeil, l'Indice de Masse Corporelle (IMC), la durée du séjour lié à l'intervention initiale) et sociodémographiques (âge, sexe, Indice développement social). L'indice de comorbidité de Charlson [58] du patient a été calculé en utilisant tous les diagnostics du patient, en utilisant la Classification Internationale des Maladies, 10^{ème} version (CIM-10). L'IMC, et le syndrome d'apnée de sommeil (SAS) ont également été définis à partir des codes CIM-10. L'indice de privation français (FDep) est un indicateur géographique agrégé qui mesure la précarité dans le lieu de résidence du patient [59]. Il s'agit d'un outil construit à partir de quatre éléments : (1) taux de chômage dans la population active (c'est-à-dire, chez les 15-64 ans en France), (2) taux d'ouvriers dans la population active, (3) taux de diplômés du secondaire dans la population non scolarisée de plus de 15 ans et (4) revenu médian déclaré par unité de consommation [59]. Il s'agit d'un proxy du niveau socio-économique du patient.

ANALYSE STATISTIQUE

Nous avons utilisé la méthode d'analyses de séquences pour construire la trajectoire des patients. Parmi les différentes méthodes existantes, nous avons utilisé la méthode d'Optimal Matching (OM) qui est un outil intuitif. Initialement utilisée dans le domaine de biologie pour comparer des séquences d'ADN [60], cette méthode a été étendue à des recherches en sociologie pour l'analyse de trajectoire de vie [11], [61]–[63]. Elle est récemment de plus en plus utilisée dans le domaine de la santé notamment pour analyser les parcours de soins [5], [17].

Le principe de cette analyse est de constituer une séquence d'états, chez chaque patient. Chaque état correspond à une technique de chirurgie bariatrique subit par le patient. Un patient changera d'état lorsqu'il réalisera une nouvelle chirurgie bariatrique (ou chirurgie de révision). L'ensemble des états représente notre « alphabet ». Dans notre cas il sera constitué par les quatre techniques de chirurgie bariatrique (AGA, SG, BPG, DBP), de l'ablation de l'AGA et décès.

Pour les analyses exploratoires sur les trajectoires des complications, les états seront constitués des complications suivantes : (1) les carences nutritionnelles (2) toute complications liées à la grossesse, (3) la lithiase biliaire, (4) les complications liées à l'AGA, (5) les hernies abdominales.

La durée totale du suivi est fixée à 10 ans pour tous les patients. Cette durée de temps est ensuite divisée en unités temporelles qui sont définies comme chacune des unités de temps où observe l'état dans lequel le patient se trouve. Vu que la quasi-totalité des patients n'ont pas plus d'un événement à l'échelle trimestrielle nous avons décidé considérer le trimestre comme unité temporelle. Cela nous permettra d'avoir 40 unités temporelles, sur 10 ans de suivi.

CALCUL DES DISTANCES ENTRE LES SEQUENCES

Une fois les séquences construites, nous utiliserons la méthode d'Optimal Matching pour calculer la distance entre les séquences. Le principe de l'OM est de faire correspondre deux séquences d'états distincts en réalisant 3 opérations élémentaires : insertion, suppression (ou délétion) et substitution. La substitution et la délétion sont toutes les deux regroupées dans un concept appelé *indel*. A chacune des opérations élémentaires est attribué un coût et nous fixerons ces coûts de façon empirique. Le coût *indel* a été fixé comme la moitié du coût de substitution maximal. Le package *TraMiner* utilise l'algorithme de Needleman-Wunsch qui permet d'avoir la distance la plus faible entre deux séquences [64].

CLASSIFICATION ET TYPOLOGIE DES TRAJECTOIRES

La matrice de distance (ou de dissimilarité) ainsi constituée à partir de l'OM servira de base pour classification des trajectoires. Nous utiliserons la classification ascendante hiérarchique (CAH) selon le critère de Ward. Il est l'algorithme le plus utilisé en analyse de séquence selon Doulhy [63]. On s'est basé sur des critères statistiques (notamment le saut d'inertie) et cliniques pour définir le nombre de classes

optimal. Par souci de parcimonie dans ce manuscrit, on décrira dans les détails que les trajectoires des patients opérés initialement d'AGA.

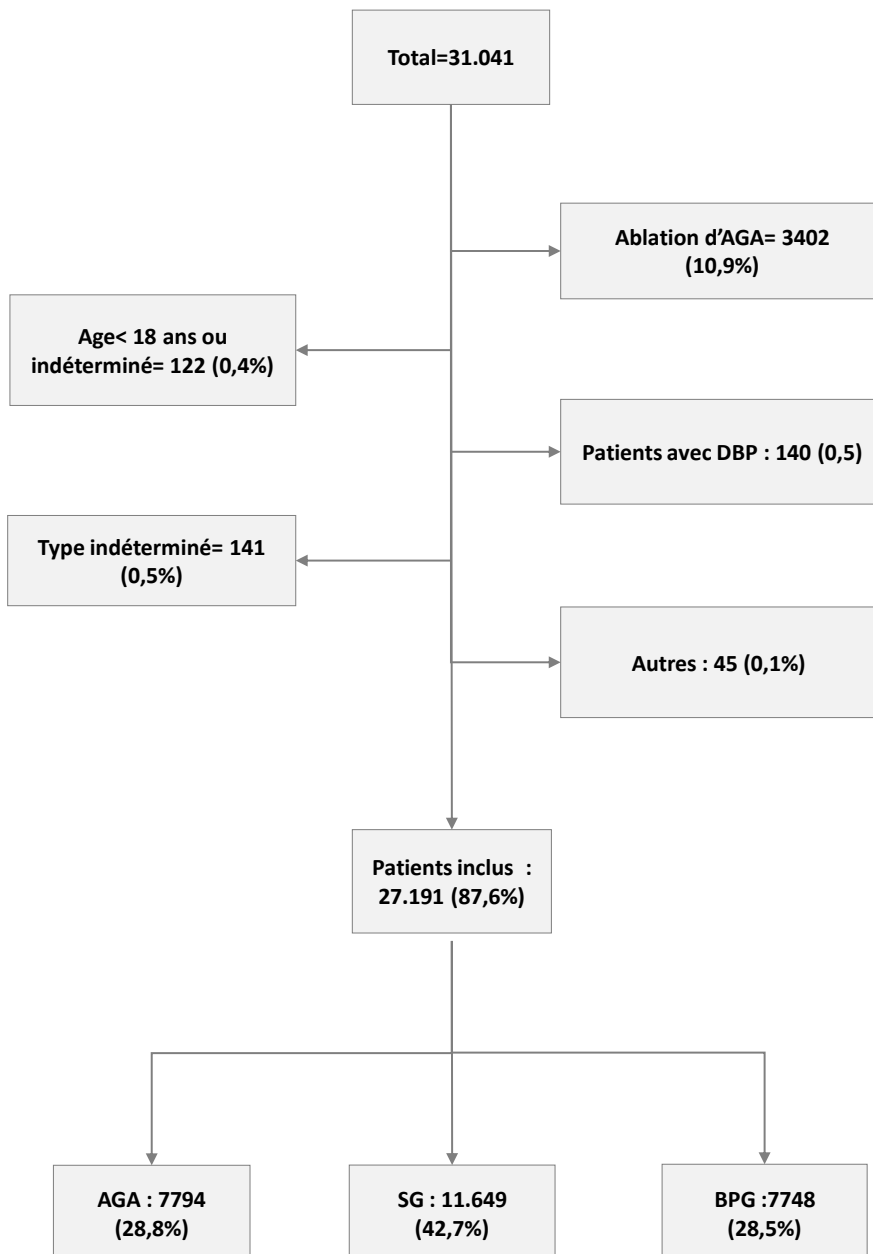
MODELISATION

Puisque notre variable dépendante qu'est la trajectoire est une variable multimodale nous utiliserons le modèle logistique multinomiale ajustée sur toutes covariables pour voir s'il n'y a pas d'association indépendante entre chaque covariable et les trajectoires. Ces associations seront mesurées par les odds ratio. Nous utiliserons les intervalles de confiance à 95% pour évaluer l'importance de ces associations. La fonction *multinom ()* du package *nnet* sera utilisée pour la modélisation. Les sorties du modèle seront représentées par des graphiques pour des simplifications de lectures. Elles seront représentées sous forme tabulaire dans les suppléments pour des besoins de détails.

RESULTATS

Au total 31.041 patients ont initialement bénéficié de la chirurgie bariatrique, dont 3.850 (12,4%) ont été exclus de notre étude, 3.402 pour ablation de l'AGA, 141 patients pour type de chirurgie indéterminé, 122 patients ont été exclus pour âge <18 ans ou indéterminé, 140 patients pour avoir bénéficié de la DBP et 45 patients pour autres raisons. Ce qui nous a laissé avec un effectif de 27.150 (87,7%) patients dans 7.794 (28,8%) cas d'AGA, 11.649 (42,7%) cas de SG et 7.748 (28,5%) cas de BPG (Figure 1).

Figure 1: Flowchart de la sélection des patients de l'étude.



Les caractéristiques initiales de ces patients stratifiées par type de chirurgie sont disponibles dans la table 1.

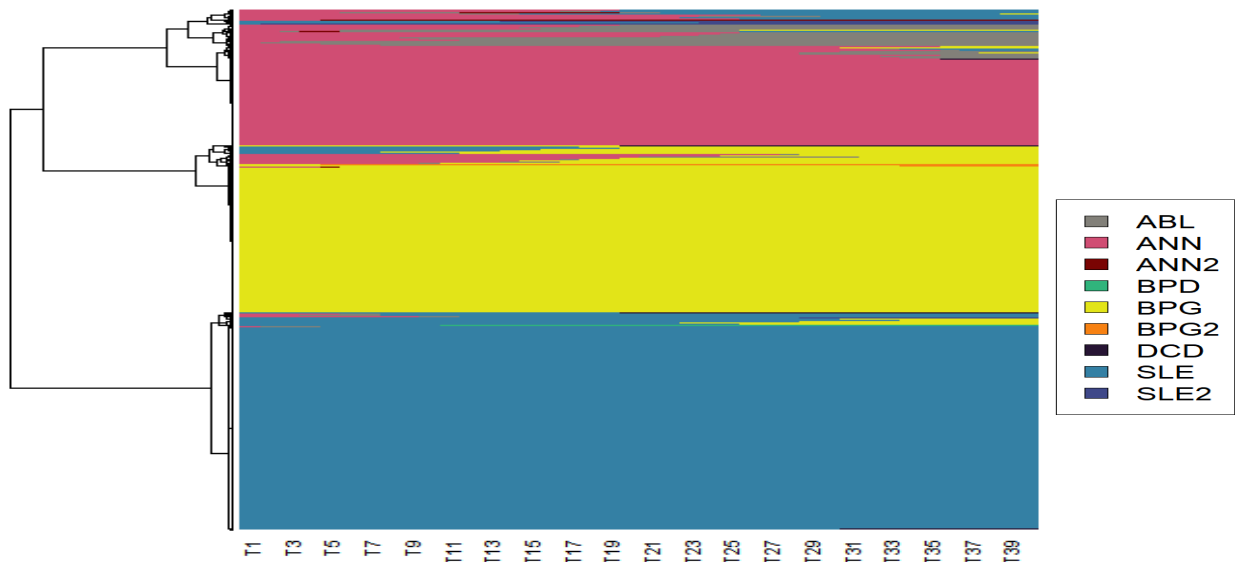
Table 1. Caractéristiques initiales des patients stratifiées par type de chirurgie.

	AGA (N=7794)	BPG (N=7748)	SG (N=11649)	Total (N=27191)	p-value
Sexe					< 0.001
Homme	1132 (14.5%)	1284 (16.6%)	2305 (19.8%)	4721 (17.4%)	
Femme	6662 (85.5%)	6464 (83.4%)	9344 (80.2%)	22470 (82.6%)	
Age					< 0.001
Moyenne (ET)	36.5 (11.517)	40.3 (11.3)	39.8 (11.8)	39.0 (11.7)	
Min-Max	18.0 - 79.0	18.0 - 71.0	18.0 - 82.0	18.0 - 82.0	
IMC					< 0.001
30-40	3041 (40.3%)	1677 (21.8%)	2849 (24.7%)	7567 (28.2%)	
40-50	4062 (53.9%)	5025 (65.2%)	7122 (61.6%)	16209 (60.5%)	
>50	434 (5.8%)	1003 (13.0%)	1584 (13.7%)	3021 (11.3%)	
SAS					< 0.001
Oui	7262 (93.2%)	6213 (80.2%)	9614 (82.5%)	23089 (84.9%)	
Non	532 (6.8%)	1535 (19.8%)	2035 (17.5%)	4102 (15.1%)	
Score de Charlson					< 0.001
0	6390 (82.0%)	5255 (67.8%)	8330 (71.5%)	19975 (73.5%)	
1-2	1286 (16.5%)	2072 (26.7%)	2820 (24.2%)	6178 (22.7%)	
>=3	118 (1.5%)	421 (5.4%)	499 (4.3%)	1038 (3.8%)	
FDep					< 0.001
Niv.1	2034 (26.5%)	2070 (26.8%)	2582 (22.8%)	6686 (25.0%)	
Niv.2	1882 (24.5%)	2060 (26.7%)	2744 (24.2%)	6686 (25.0%)	
Niv.3	1840 (23.9%)	1891 (24.5%)	2955 (26.1%)	6686 (25.0%)	
Niv.4	1930 (25.1%)	1703 (22.0%)	3052 (26.9%)	6685 (25.0%)	
Durée du séjour initial					< 0.001
Moyenne (ET)	2.656 (1.355)	6.294 (4.223)	5.911 (4.054)	5.087 (3.879)	
Min-Max	0.0 - 39.0	1.0 - 134.0	0.0 - 104.0	0.0 - 134.0	

ANALYSE DES TRAJECTOIRES.

Parmi les 27.150 patients qui ont été inclus dans l'étude, 4.605 (16,9%) ont fait un changement d'état (chirurgie de reconversion ou une ablation de l'AGA) pendant les 10 années de suivi, dont 3.306 (12,2%) étaient constitués de patients ayant bénéficié de l'AGA, 1.099 (4%) de la SG et 100 (0,7%) du BPG (Figure 2).

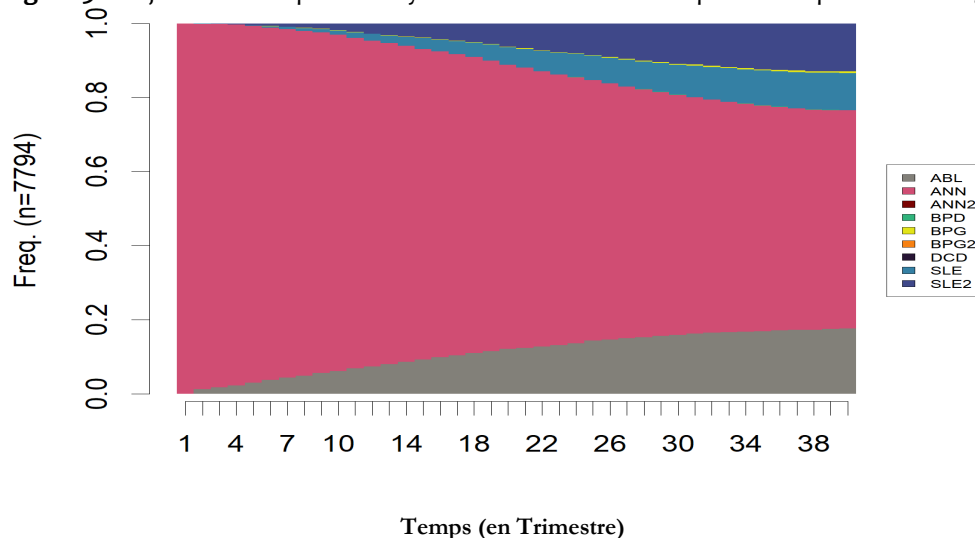
Figure 2. Trajectoires de l'ensemble des patients ayant bénéficié de la chirurgie bariatrique



TRAJECTOIRES DES PATIENTS AYANT BENEFICIE DE L'AGA.

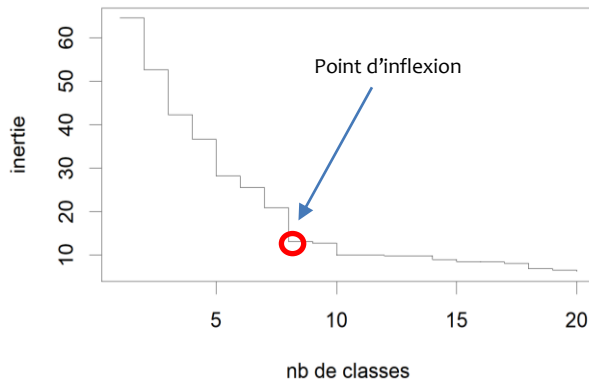
Parmi les 7.794 patients qui ont initialement bénéficié d'un AGA on retrouve que 41% d'entre eux ont changé d'état, dans les 10 ans du suivi, et en particulier, 17,7% ont retiré leurs anneaux, 12.9% ont eu une SG, 10% ont eu un BPG (Figure 3). Le taux de mortalité est très faible autour de 0.4% sur dix ans de suivi.

Figure 3 : Trajectoires des patients ayant bénéficié de l'AGA représentée par un chronogramme.



En utilisant le saut d'inertie (Figure 4) et des critères cliniques nous avons retenu sept classes de trajectoires suffisamment homogènes pour les patients ayant bénéficié de l'AGA.

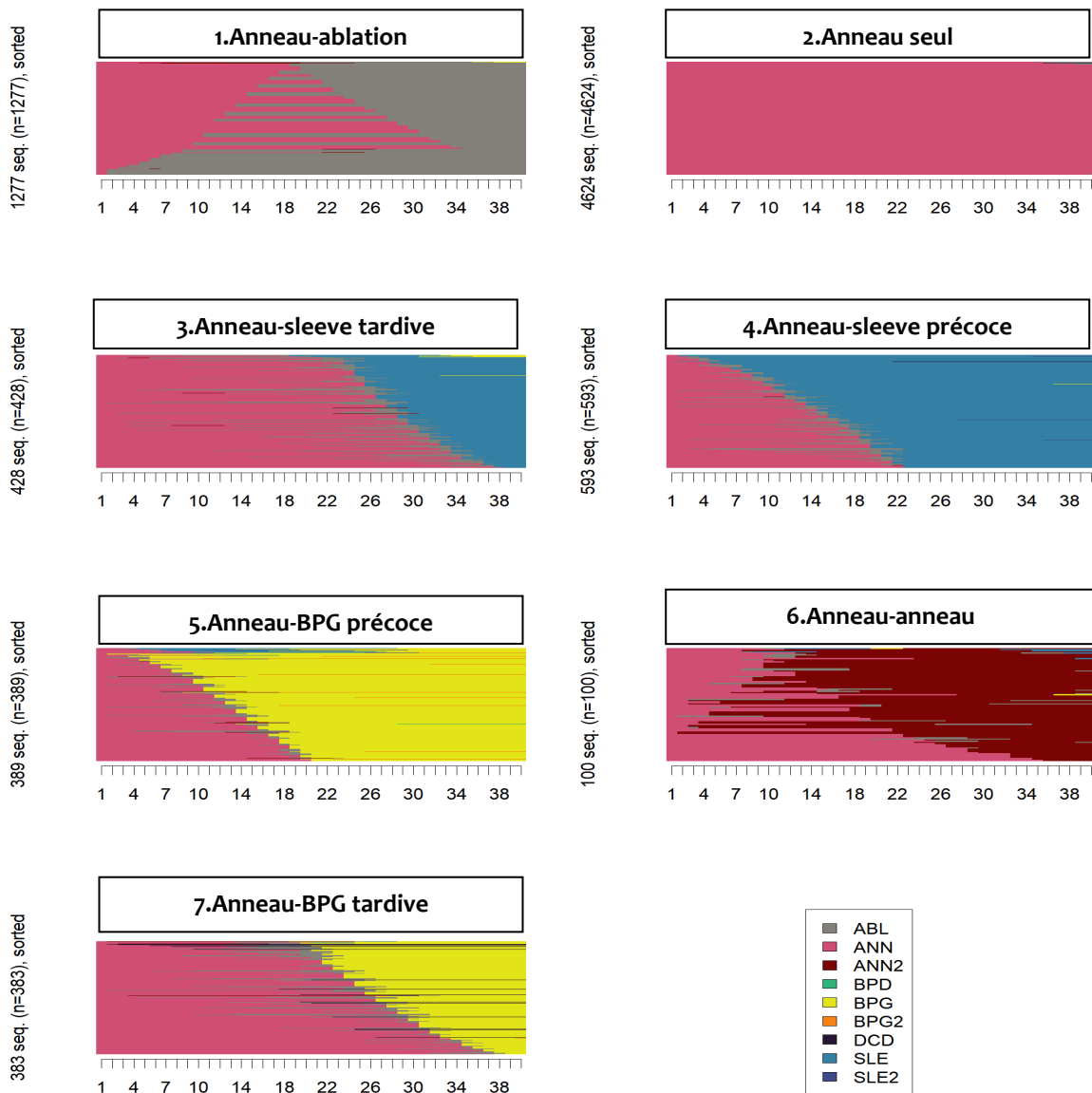
Figure 4 : Saut d'inertie pour la classification des trajectoires des patients ayant bénéficié de l'anneau.



- La première trajectoire est celle des patients ayant fait une simple ablation de l'anneau que nous avons nommée : « Anneau-ablation ».
- Le groupe 2 constitué de patients gardant leur AGA tout au long de leur suivi a été nommé « Anneau seul ».
- La troisième et la quatrième trajectoire sont celles des patients ayant fait respectivement des conversions en SG précocement et tardivement. Ces deux trajectoires seront regroupées pour similarités cliniques.
- Les trajectoires 5 et 7 quant à elles correspondent respectivement au passage de l'AGA au BPG précocement et tardivement. Elles seront également regroupées comme les deux précédentes.
- Finalement la sixième trajectoire est celle des patients ayant eu une deuxième intervention d'AGA.

La classe « Anneau seul » était la plus fréquente, soit 4.624 (59,3%), suivie du groupe « Anneau-ablation » avec 1.412 (16,4%) patients, du groupe « Anneau-sleeve » avec 1057 (13,1%) patients, du groupe « Anneau-BPG » avec 741 (9,9%) patients et enfin du groupe « Anneau-anneau » avec 100 patients (1,2%) (Figure 4).

Figure 5 : Trajectoire des patients ayant bénéficié d'un AGA.

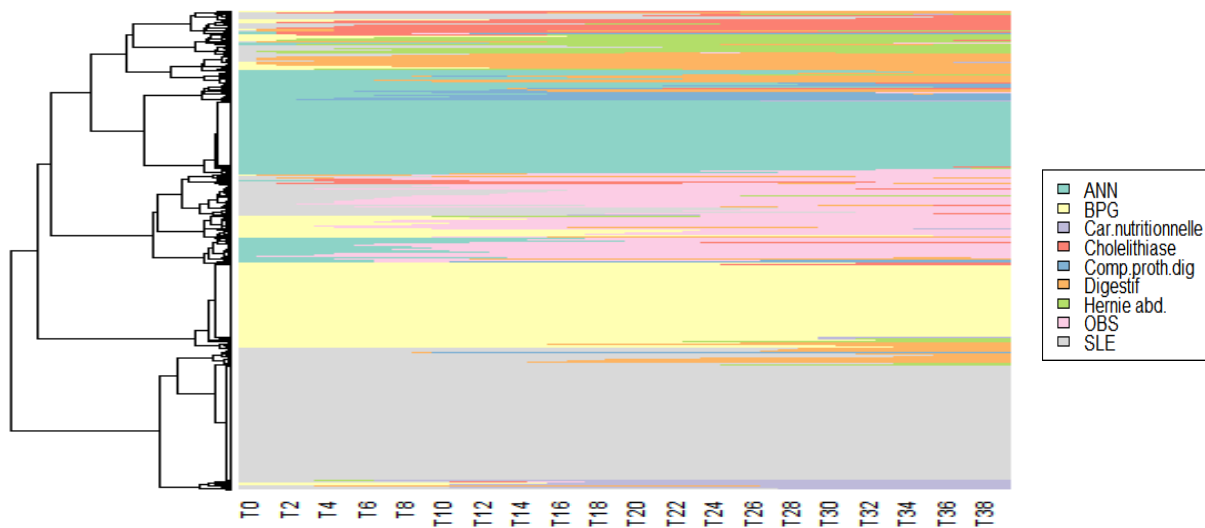


LA DESCRIPTION DES TRAJECTOIRES PAR RAPPORT AUX COMPLICATIONS

De manière exploratoire nous avons appliqué l'analyse de séquences aux complications chirurgicales. Nous avons construit notre alphabet avec les principales complications postopératoires de chirurgie bariatrique : les carences nutritionnelles, toute complications liées à la grossesse, la lithiase biliaire, les complications liées à l'AGA, les hernies abdominales et autres complications digestives. Nous avons gardé la même unité temporelle de 3 mois.

L'analyse des trajectoires en se basant sur les complications donne des résultats difficiles à interpréter (Figure 6). Les complications sont des événements ponctuels dans le temps alors que l'analyse des séquences est un outil plus adapté lorsque l'état du patient reste permanent jusqu'à l'apparition d'un nouvel état. La méthode de l'analyse de séquence semble donc moins adaptée à ce type d'événement, contrairement à la méthode de l'AFC, utilisée dans l'étude précédente.

Figure 6 : Classification des trajectoires des patients en fonction des complications



ANALYSES BIVARIEES DES FACTEURS ASSOCIES AUX TRAJECTOIRES

Nous avons étudié les associations sans ajustement entre les facteurs démographiques (sexe et âge), socioéconomiques (Indice de développement social) et cliniques (score de Charlson, IMC, durée de séjour, SAS) et les différentes trajectoires.

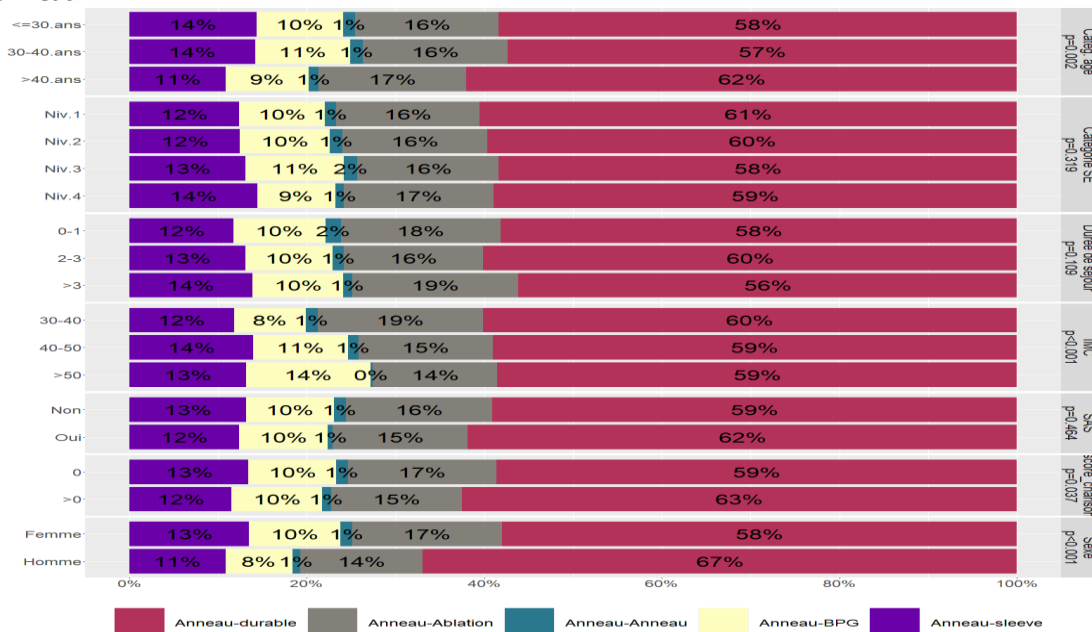
Pour les patients ayant initialement bénéficié de l'AGA nous constatons une association significative entre le sexe et la trajectoire. Être une femme diminue la chance d'être dans la trajectoire anneau seul (58% vs 66% pour les hommes).

L'âge aussi semble avoir un impact sur la trajectoire. Les patients de plus de 40 ans semblent avoir plus de chance d'appartenir à la classe « Anneau seul » que les patients plus jeunes (62% vs 57 et 58% pour respectivement les 30-40 ans et < 30 ans).

Avoir un IMC initial > 50 kg/m² semble légèrement augmenter le risque d'être dans le groupe « Anneau-BPG » (12% vs 10 et 7% pour les IMC à 40-50 et 30-40 kg/m²); tandis que les patients qui ont un IMC initial entre 30-40 kg/m² ont plus tendance à être dans le groupe « Anneau-ablation » (20% vs 17% et 16% pour les IMC à 40-50 et > 50kg/m²).

L'indice de développement social, le syndrome d'apnée de sommeil et le score de Charlson quant à eux, ne semblent pas montrer d'association avec la trajectoire du patient (Figures 7).

Figure 7 : Analyse sans ajustement des facteurs associés à la trajectoire des patients ayant bénéficié de l'anneau.



ANALYSES MULTIVARIEES DES FACTEURS ASSOCIES AUX TRAJECTOIRES

RELATIONS ENTRE LES COFACTEURS ET LES TRAJECTOIRES DES PATIENTS APRES UN AGA

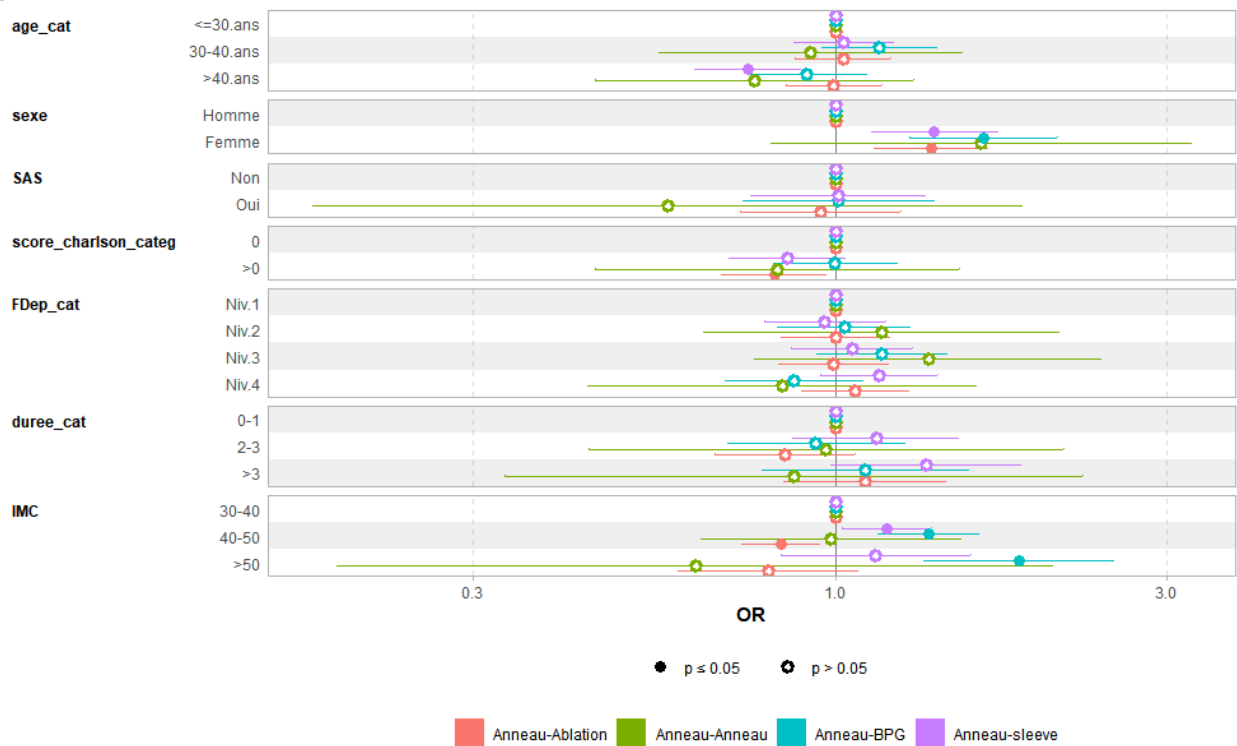
Après avoir ajusté sur les différentes covariables nous pouvons toujours observer une faible association entre le sexe et la trajectoire du patient. Être femme augmente le risque d'être dans les groupes « Anneau-BPG », « Anneau-Sleeve » et « Anneau-Ablation » avec respectivement un OR de 1,63 [1,26-2,10]; 1,43 [1,16-1,76] et 1,32 [1,1-1,58] (Figure 9).

Après ajustement, l'âge ne semble pas influencer sur la trajectoire globalement. Cependant nous pouvons constater qu'avoir plus de 40 ans diminue significativement le risque d'appartenir à la trajectoire « Anneau-sleeve » par rapport aux moins de 30 ans avec un OR à 0,75 [0,63-0,89]

Une augmentation de l'IMC initial augmente le risque d'être dans la trajectoire « Anneau-BPG » tandis qu'elle diminue celui d'être dans la trajectoire « Anneau-ablation ». Les patients avec un IMC à 40-50 kg/m² et ceux avec un IMC >50 kg/m² ont respectivement 1,37 [1,15-1,63] et 1,72 [1,23-2,41] fois plus de risque d'appartenir à la trajectoire « Anneau-BPG » que ceux ayant un IMC à 30-40 kg/m². Ailleurs on constate une légère augmentation du risque d'appartenir à la trajectoire « Anneau-sleeve » pour les patients avec un IMC à 40-50 kg/m² comparés aux patients avec l'IMC à 30-40 kg/m² avec un OR à 1,17 [1,01-1,35]. A l'inverse avoir un IMC entre 40-50 diminue significativement le risque d'appartenir à la trajectoire « Anneau-ablation » par rapport au fait d'avoir un IMC à 30-40 kg/m² avec un OR à 0,86 [0,75-0,97]

Le score de Charlson ne semble pas avoir d'impact sur la trajectoire, bien que paradoxalement les patients ayant les scores les plus élevés ont tendance à être moins dans les trajectoires avec chirurgie de conversion ou ablation d'anneau. Le niveau socioéconomique et la durée du séjour initial ne semblent pas non plus avoir un impact sur la trajectoire du patient.

Figure 9 : Modélisation de la trajectoire des patients ayant bénéficié de l'AGA.



DISCUSSIONS

Dans cette étude rétrospective réalisée à partir des données du PMSI, nous avons identifié les différentes trajectoires de soins des patients ayant bénéficié d'une chirurgie bariatrique, en termes de réintervention et de complication. L'analyse des séquences était la méthode utilisée. Nous avons inclus plus de 27.000 patients, dont les 28,7% étaient des cas d'AGA, les 42,8% des cas de SG, et les 28,5% des cas de bypass gastriques.

Parmi les patients ayant bénéficié de l'AGA, dans environ 40 % des cas, l'AGA sera retiré à 10 ans. La plupart des patients à qui ont retiré l'AGA (60%), seront réopérés d'une nouvelle chirurgie bariatrique de type SG ou BPG. On observe aussi quelques cas de réintervention avec la pose d'un deuxième anneau. De nombreuses études ont également rapporté chez les patients ayant bénéficié de l'AGA un taux élevé d'ablation ou de chirurgie de révision sur le long terme [65].

En ce qui concerne les patients qui ont bénéficié de la SG, plus de 9% d'entre eux auront une chirurgie de révision, principalement en bypass gastrique [66], [67]. Peu d'études existent sur le parcours à long terme des patients ayant bénéficié de la SG, et les résultats de ces études en termes de taux de conversion (ou de révision) diffèrent en raison des effectifs relativement petits [66].

Dans ce travail nous avons utilisé la méthode de l'analyse des séquences pour construire les trajectoires des patients. L'OM a été utilisé pour calculer les distances entre les trajectoires qui ont ensuite été classées par la CAH ce qui permet généralement une classification homogène des trajectoires.

L'enjeu majeur lors de la classification des trajectoires est de trouver un bon compromis entre un nombre faible de classes qui simplifie l'interprétation, et un nombre de classes assez élevé dans le but d'obtenir une homogénéité intra-classe. Dans nos analyses nous avons pu isoler des trajectoires-types avec des bonnes homogénéités, à la fois pour les patients ayant bénéficié de l'AGA et ceux ayant bien bénéficié de la SG.

Environ 20% des patients change dans les 10 ans qui suivent la première intervention. De plus la majorité des patients des patients qui changent d'état ne le font qu'une seule fois (c'est-à-dire une seule chirurgie de révision). Nos trajectoires-type étaient donc constituées d'une ou de deux étapes dans la plupart des cas.

L'une des forces de cette étude est que nous avons travaillé sur un effectif large. L'analyse des séquences que nous avons utilisée pour identifier les trajectoires est un outil très puissant, souple, et intuitif permettant à l'utilisateur d'adapter les paramètres en fonction de la spécificité des états constituant les séquences. Il permet également une lecture très simple de la trajectoire du patient.

LIMITES

Notre analyse montre des limites. La classification des trajectoires en fonction des complications donne des résultats peu concluants. Les complications sont des états ponctuels dans le temps, c'est-à-dire au moment elles ont été diagnostiquées et ne peuvent être considérées comme permanentes, contrairement au type de chirurgie du patient.

L'Analyse Formelle de Concepts, utilisée dans l'étude précédente, semble plus appropriée pour analyse des événements ponctuels comme les complications.

Il est à noter que cette étude est une étude rétrospective menée à partir de base de données médico-administratives. Nous n'avons de l'information sur un patient que lorsqu'il est vu à l'hôpital ; les cas d'éventuels décès ou de réopération dans d'autres hôpitaux ne sont donc pas identifiés.

La méthode d'OM est par ailleurs un algorithme très gourmand en mémoire vive pour la machine. Ce qui nous a contraint à réduire la taille de l'échantillon qui était initialement prévu. Cette même problématique liée à l'OM nous a imposé une réduction de la longueur de séquence, qui s'est traduite par l'augmentation de l'unité temporelle (suivi trimestriel au lieu de mensuel). Cela n'a cependant pas induit de perte d'information, puis quasi-totalité des patients n'ont pas plus d'un évènement par trimestre.

CONCLUSIONS

L'analyse des trajectoires des patients ayant bénéficié de la chirurgie bariatrique par la méthode des séquences permet de mieux comprendre l'histoire naturelle de cette chirurgie. La méthode utilisée semble appropriée pour analyser les actes chirurgicaux itératifs. Elle est intuitive, simple et donne un bon aperçu visuel de la trajectoire du patient. Elle doit donc être plus envisagée dans l'analyse des trajectoires lorsque « l'état » du patient peut être considéré comme un état durable. En revanche dans le cas des complications elle nous donne des résultats peu informatifs et difficilement interprétables. Cette méthode ne semble donc pas appropriée pour l'analyse d'états ponctuels et transitoires.

Etude N°3

ANALYSE DES SEQUENCES DES SOINS DES PATIENTS ATTEINTS DE CHOLANGIOMYOCARCINOME TERME

ETUDE EN COURS (MASTER 2 – DR S. TZEDAKIS)

RESUME

INTRODUCTION

Le cholangiocarcinome (CC) constitue un groupe hétérogène des cancers originaires de l'arbre biliaire. L'épidémiologie du CC et de ses différents sous-types sont mal connus en France, car cette pathologie a été souvent classée dans le passé parmi les tumeurs du foie ou comme métastase. L'étude de la prise en charge de cette pathologie ainsi que les différents parcours de soins reste un sujet important à explorer.

METHODES

Il s'agit d'une étude de cohorte rétrospective basée sur la base de données nationale des hospitalisations (PMSI). Cette source comprend toutes les données de remboursement des séjours hospitaliers en France. Nous avons donc inclus les patients présentant un code de CC entre 2011 et 2020. L'analyse du parcours de soins a été réalisée par la technique de l'analyse des séquences (*sequence analysis*). Nous avons défini comme « état » dans l'analyse des séquences: l'acte chirurgical, la chimiothérapie, les actes interventionnels (radiologiques et endoscopiques), la radiothérapie et le décès. La période de suivi était de 2 ans après inclusion. L'analyse des facteurs associés a été réalisée par un modèle logistique multinomial.

RESULTATS

Au total nous avons inclus 34.000 patients atteints de CC, répartis en 4 groupes anatomiques : 11% avec un CC extra-hépatique, 41% un CC intra-hépatique, 14% un CC de la vésicule biliaire et 33% un CC de localisation non définie. Les patients présentent des caractéristiques de base très différentes parmi les 4 groupes anatomiques. L'analyse des trajectoires a porté sur un sous-groupe de 14.000 patients, retrouvant 6 classes principales : (1) chimiothérapie seule, (2) chimiothérapie - interventionnel, (3) chimiothérapie - décès, (4) chimiothérapie - palliatif (5) chirurgie, et (6) chimiothérapie - radiothérapie. Plusieurs facteurs sont associés à l'appartenance à une classe donnée : l'âge, la localisation de la tumeur, les comorbidités ainsi que le type d'établissement où la prise en charge a lieu.

DISCUSSION

Cette étude a permis d'estimer la prévalence du CC et d'identifier les principales trajectoires des patients en France entre 2011-2020. Nous avons aussi pu analyser les divers facteurs influençant l'appartenance à ces trajectoires qu'ils soient intrinsèques à la personne (sexe et âge), liés à la maladie elle-même (type de tumeur) ou à la filière de prise en charge (caractéristiques de l'établissement). Enfin, elle a permis de mettre en évidence des problématiques d'organisation de la prise en charge de ces patients, notamment sur le délai de prise en charge et le manque de centralisation dans les centres experts.

INTRODUCTION

Le cholangiocarcinome (CC) constitue un groupe hétérogène de cancers originaires de l'arbre biliaire qui, en fonction de leur origine anatomique, peuvent être classés en différentes entités distinctes, d'étiologie et pathogénie différentes. En France, les cancers biliaires représentent environ entre 2.000 et 3.000 nouveaux cas par an avec un taux d'incidence de 1,4 et 0,7 cas/100 000 habitants par an chez l'homme et la femme respectivement [68].

La chirurgie représente le seul traitement curatif du CC. Seuls 20% à 40 % des patients sont éligibles à une résection chirurgicale au moment du diagnostic, les récurrences sont fréquentes et le pronostic reste sombre avec un taux de survie à cinq ans compris entre 20-50% après un traitement curatif.

L'épidémiologie du CC et de ses différents sous-types en pratique courante est mal connue en France. Cette observation est encore plus vraie si nous considérons que cette pathologie n'était pas correctement identifiée dans le passé : elle était classée comme métastase de primitif inconnu ou tumeur maligne des voies biliaires, sans précision. Ce n'est que depuis une dizaine d'années que l'enregistrement du CC est fait correctement.

L'étude de la prise en charge de cette pathologie ainsi que les différents parcours de soins des patients diagnostiqués avec un CC (ex. délai de prise en charge, impact des inégalités socioéconomiques sur la prise en charge, volume du centre de prise en charge, type de traitement) reste un sujet à explorer. Jusqu'à présent, les données dont nous disposons n'étaient pas exhaustives et le rapport le plus complet des patients ayant reçu un traitement curatif reste celui de l'association française de chirurgie (AFC) datant de 2009, basé sur des patients de centres sélectionnés, sans recensement des patients non opérés et avec 70% des observations étant antérieures à 2000 [69]. L'amélioration de la prise en charge de cette pathologie est un enjeu majeur de santé publique en France.

L'objectif principal de ce travail est l'étude de la prise en charge des patients atteints d'un CC en France via les différents parcours de soins sur une période de 10 ans. L'objectif secondaire est d'explorer les différents facteurs influençant l'appartenance à ces parcours.

MATERIELS ET METHODES

DESIGN

Il s'agit d'une étude de cohorte rétrospective. La population de cette étude était constituée de patients avec un code de CC diagnostiqué dans la période 2011-2020.

POPULATION

Les critères d'inclusion :

- Age \geq 18 ans.
- Tous les patients identifiés avec un des codes CIM-10 de cholangiocarcinome : C223, C23, C240, C248, C249, D376.
- Patients ayant un séjour lié au CC entre le 1^{er} Janvier 2011 et le 31 Décembre 2020.

Critère de non-inclusion :

- Des patients avec un code C22.2 (carcinome hépatocellulaire) ou un code C24.1 (tumeur maligne de l'ampoule de Vater) ont été exclus car s'agissant de tumeurs potentiellement d'autre origine (hépatique, intestinale ou pancréatique).

SOURCE

Les données ont été extraites de la base du Programme de Médicalisation des Systèmes d'Information (PMSI).

OUTCOMES

L'objectif principal de l'étude est descriptif : identification et classification des trajectoires des patients atteint de CC. En objectif secondaire nous avons analysé les facteurs associés aux différentes trajectoires.

VARIABLES

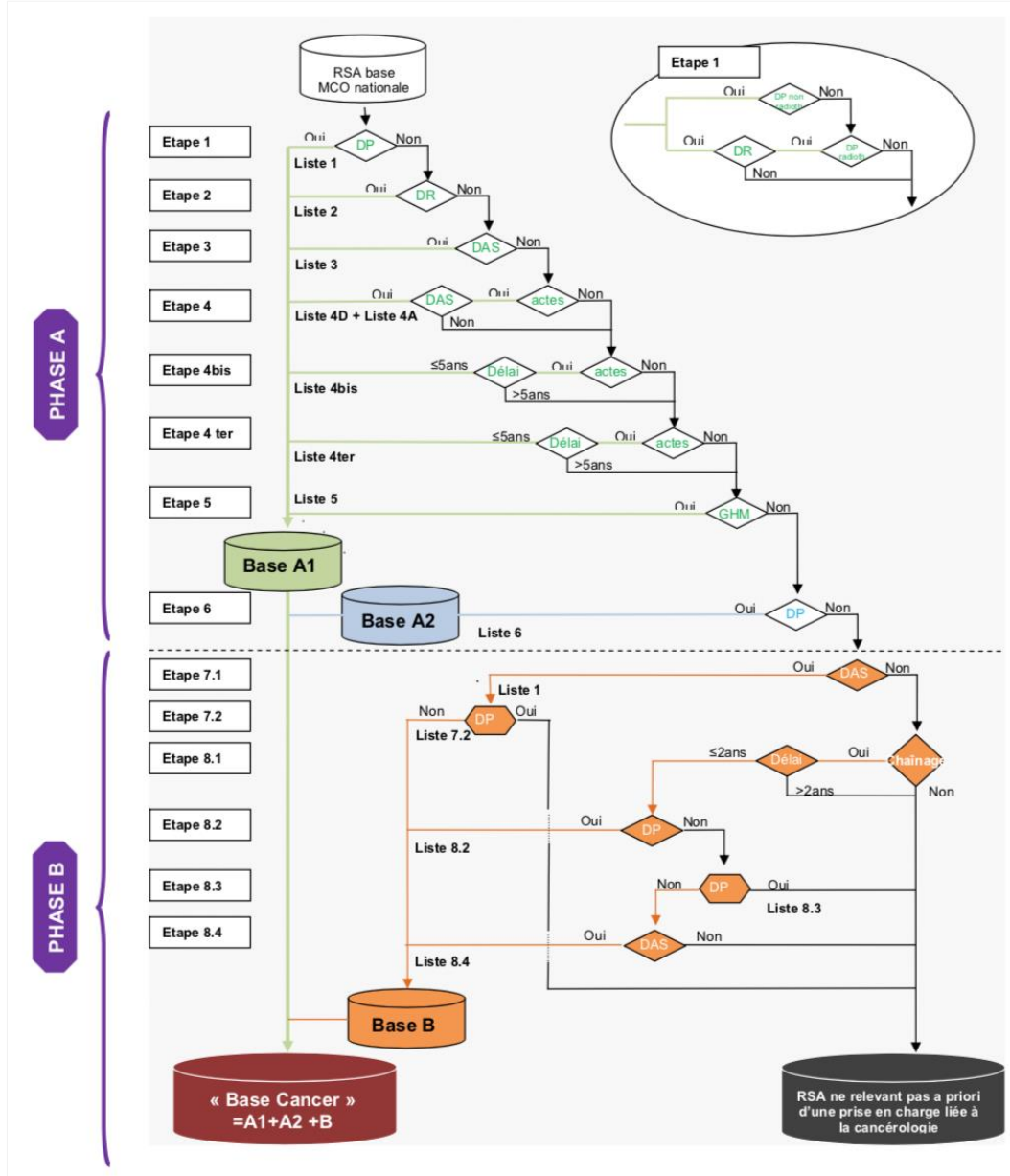
Nous avons inclus comme variable les informations sur l'état de santé des patients, résumées par le score de Charlson [70]. Nous avons utilisé comme variable socio-économique l'indice de défavorisation sociale (FDep), défini par l'Institut National de la Santé et de la Recherche Médicale (Inserm) à partir des données de recensement de l'Institut National de la Statistique et des Etudes Economiques (Insee) [59].

Les patients identifiés avec un code de CC ont été classés selon la localisation anatomique de leur cancer comme suit :

- CC intra hépatique (CCIH) portant le code C22.1,
- CC extra hépatique (CCEH) portant le code C24.0,
- CC de la vésicule biliaire (CCVB) portant le code C23.0,
- CC non défini (CCND) portant le reste des codes.

Afin de repérer les premiers séjours en rapport avec le CC (le code CIM-10 n'apparaissant que rarement au premier séjour), un algorithme développé par l'Institut National du Cancer (INCa) a été utilisé. L'objectif de cet algorithme était de sélectionner au sein de la base PMSI-MCO, selon une méthodologie commune à tous les acteurs, les séjours et séjours hospitaliers en relation avec une prise en charge liée à la cancérologie [71] (Figure 1).

Figure 1: schéma de l'algorithme d'INCa pour l'identification des séjours liés à la prise en charge d'un cancer.



ANALYSE STATISTIQUE

Comme pour l'étude précédente nous avons utilisé la méthode de l'analyse des séquences. Nous avons construit l'alphabet par l'identification de six états: actes interventionnels (radiologiques et endoscopiques), chirurgie, chimiothérapie, radiothérapie, soins palliatifs, et décès.

L'unité temporelle retenue est la semaine, et la durée de suivi totale était de deux ans à partir du premier séjour associé au CC. Par conséquent, chaque séquence sera composée de 104 états (52 x 2).

Lorsque plusieurs états étaient simultanément présents dans la même semaine, leur relation temporelle a été respectée, le premier état était conservé et les états supplémentaires étaient décalés d'une semaine de telle manière qu'il y ait qu'un état par semaine. Enfin, le volume de patients et le nombre d'états ne permettaient pas un traitement informatique sur les machines à notre disposition. Nous avons donc réalisé l'analyse des trajectoires sur un échantillon de patients.

CALCUL DES DISTANCES ENTRE LES SEQUENCES & CLASSIFICATION ET TYPOLOGIE DES TRAJECTOIRES

Nous avons appliqué les mêmes méthodes que dans l'étude précédente, à savoir l'Optical Matching pour le calcul des distances entre les séquences et la Classification Ascendante Hiérarchique pour la classification des classes.

MODELISATION

Dans l'analyse des facteurs potentiellement associés à une séquence donnée, nous avons considéré les facteurs en lien avec les caractéristiques des patients (sexe, âge, score de Charlson, type de CC et score de déprivation) et en lien avec leur prise en charge (type d'établissement, volume d'activité, délai du début de prise en charge). Une première analyse univariée a été réalisée à l'aide du test de Chi². Les facteurs pour lesquels la p-valeur était inférieure à 0.1 ont été inclus dans une analyse multivariée. Puisque la variable dépendante (type de trajectoire) est une variable multimodale, une régression logistique multinomiale a été réalisée, ajustée sur les différentes covariables explicatives. La multicolinéarité a été vérifiée au préalable. Les associations entre le type de trajectoire et les covariables ont été exprimées en odds-ratio (OR) avec leur intervalle de confiance à 95% pour évaluer l'importance de ces associations. L'ensemble de ce travail a été effectué grâce au logiciel R (v.4.0.3) et au package TraMiner (v.2.2-2) [72].

RESULTATS

DESCRIPTIF DE LA POPULATION

Au total 51.123 patients ont été identifiés entre Janvier 2011 et Décembre 2020 avec un code de CC.

Parmi ces patients, 36.003 (70,4%) présentaient un code unique de CC, 11.522 (22,5%) présentaient un double code de CC alors que 3598 (7%) présentaient plus de 2 codes. La classification du type de cholangiocarcinome n'était donc possible que pour 33979 (66,5%) patients.

Les caractéristiques des patients ainsi que des séjours sont présentées dans le tableau 1.

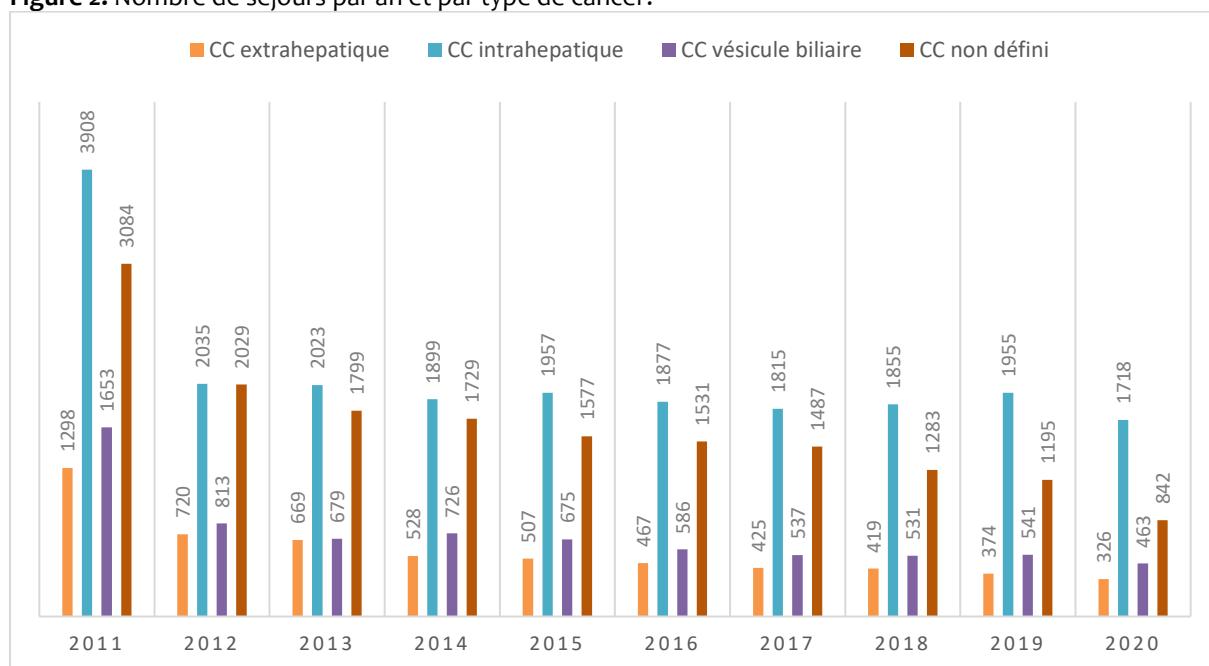
Tableau 1. Caractéristiques des patients au début de la prise en charge.

	CC extrahépatique	CC intrahépatique	CC vésicule biliaire	CC non défini	Total	p
	5733 (11.2)	21042 (41.1)	7204 (14.1)	17144 (33.5)	51123	
Sexe: Femme (%)	2602 (45.4)	9639 (45.8)	4728 (65.6)	8207 (47.9)	25176 (49.2)	<0.001
Age moyenne (SD)	73.96 (12.33)	72.43 (12.42)	73.73 (12.62)	71.79 (12.48)	72.57 (12.48)	<0.001
Catégorie de Charlson score (%)						<0.001
0	2024 (35.3)	6620 (31.5)	2471 (34.3)	6241 (36.4)	17356 (33.9)	
1-2	1579 (27.5)	4675 (22.2)	1359 (18.9)	4036 (23.5)	11649 (22.8)	
3-4	635 (11.1)	1858 (8.8)	513 (7.1)	1526 (8.9)	4532 (8.9)	
>4	1495 (26.1)	7889 (37.5)	2861 (39.7)	5341 (31.2)	17586 (34.4)	
Classe fDEP09 (%)*						<0.001
1er quantile	1085 (19.5)	3489 (17.0)	1185 (17.4)	3415 (20.4)	9174 (18.5)	
2eme quantile	1016 (18.2)	3961 (19.3)	1212 (17.8)	2982 (17.8)	9171 (18.5)	
3eme quantile	1084 (19.4)	4199 (20.5)	1381 (20.3)	3269 (19.5)	9933 (20.0)	
4eme quantile	1144 (20.5)	4434 (21.6)	1481 (21.7)	3437 (20.5)	10496 (21.1)	
5eme quantile	1248 (22.4)	4447 (21.7)	1558 (22.9)	3623 (21.7)	10876 (21.9)	
Durée de suivi, mois (SD)	10.81 (18.81)	7.51 (14.50)	11.15 (20.28)	10.78 (17.20)	9.49 (16.92)	<0.001
Nb moyen des séjours (SD)	9.58 (14.98)	9.63 (14.51)	7.69 (12.49)	12.39 (16.92)	10.29 (15.25)	<0.001
DMS, jours (SD)	8.86 (11.16)	8.77 (11.05)	8.93 (10.39)	7.71 (9.64)	8.44 (10.53)	<0.001

*Classe fDEP09: 1^{er} quantile: communes les plus favorisées) et 5^{eme} quantile (communes les plus défavorisées)

Le nombre de séjours identifiés par an sont décrit dans la figure 2. Le nombre nettement plus élevé de séjours identifiés en 2011 est en lien avec des cas prévalent de CC diagnostiqués avant 2011.

Figure 2. Nombre de séjours par an et par type de cancer.



Les caractéristiques des hôpitaux impliqués dans la prise en charge du CC sont décrites au tableau 2.

Tableau 2. Caractéristiques des établissements 2013-2020

	CH	CHR/U	CLCC	PRIVE	PSPH/EBNL	Total
Nb d'établissement (%)	655 (46)	182 (13)	29 (2)	430 (30)	117 (8)	1423
Nb de séjours						
Petit (< 4 patients/an)	463 (50)	74 (7.9)	1 (0.1)	318 (34.3)	70 (7.5)	926
Moyen (4-34 patients/an)	148 (46)	31 (10)	5 (2)	98 (31)	39 (12)	321
Grand (>34 patients/an)	44 (27)	77 (46)	23 (14)	14 (8)	8 (5)	166
Nb des patients [§] (%)	19261 (31.4)	17792 (29.0)	3898 (6.35)	15459 (25.2)	4924 (8.0)	61334

CH : Centre Hospitalier, CHR/U : Centre Hospitalier Régional/Universitaire, CLCC : Centre de Lutte Contre le Cancer, PRIVE : centre hospitalier privé, PSPH/EBNL : centre hospitalier Participant au Service Public Hospitalier/Établissements à But Non Lucratif.

[§] Certains patients sont pris en charge dans plusieurs types établissements

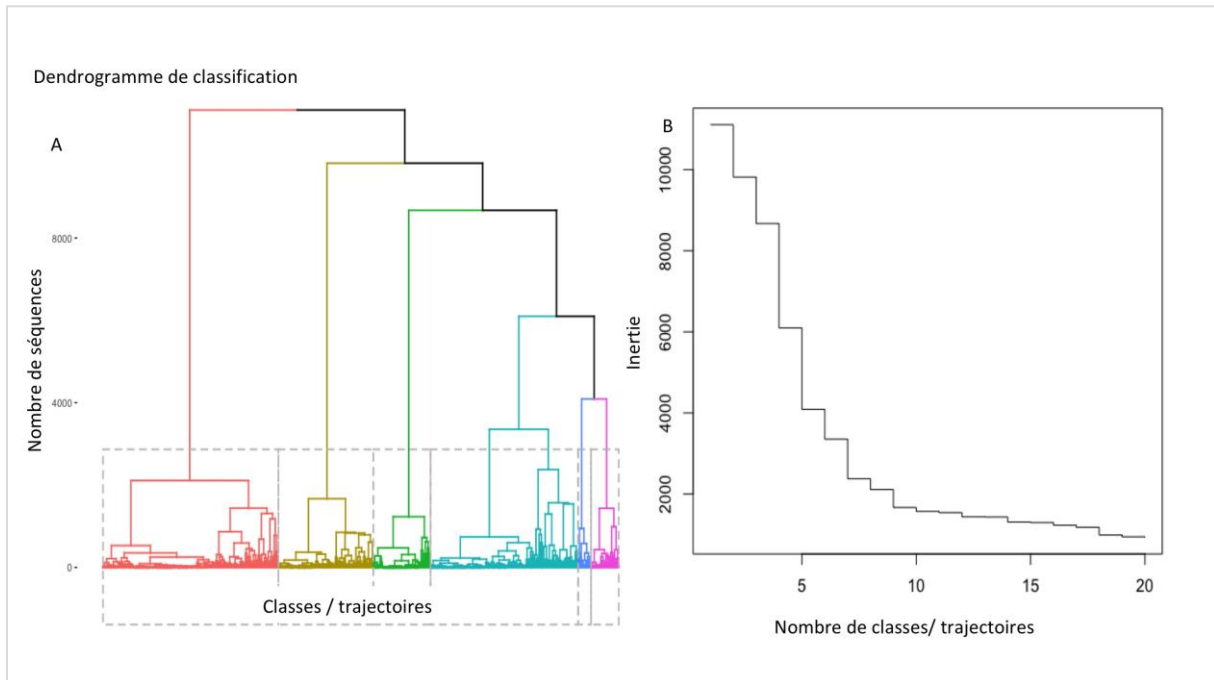
TRAJECTOIRES

Le volume des données étant très grand, l'analyse des séquences et des parcours de soins était contrainte sur le plan informatique. En effet, le temps de calcul des trajectoires dépassait limites du processeur en terme de mémoire.

Dans cette analyse préliminaire un échantillon des patients a donc été analysé. Parmi les 51.123 patients, 35.004 (68%) patients traités après 2013 a été choisis afin de disposer de l'information la plus récente et de s'affranchir des cas prévalents. Enfin, les patients pour lesquels le type de CC était non défini ainsi que les patients qui n'avaient aucun « état » pour la construction de leur trajectoire été exclus. Le parcours des 14.704 patients (40%) a été finalement analysé.

L'*Optimal Matching* a permis d'identifier 9.968 séquences différentes. La CAH a conduit à une partition en six classes (Figure 3). La partition peut être considérée comme étant de bonne qualité.

Figure 3: Dendrogramme de classification (A) et courbe des sauts d'inertie (B) mettant en évidence la flexion de la courbe autour de 6 classes.



Les trajectoires créées semblent regrouper des états relativement stables (Figure 4).

- La trajectoire 1 (appelé **CHIMIO-EXPLO**) avec 5.000 patients (34%) est une trajectoire dominée par des gestes interventionnels (endoscopiques et/ou radiologiques). Une partie de cette population (20%) a reçu de la chimiothérapie au début de sa prise en charge.
- La trajectoire 2 (appelé **CHIMIO-DECES**) avec 2.708 patients (18,4%) est caractérisée par un décès précoce après état d'explorations et de chimiothérapie.
- La trajectoire 3 (appelé **CHIMIO**) avec 4.209 patients (28,6%) est une trajectoire relativement plus hétérogène avec l'état de chimiothérapie occupant la majorité de la trajectoire. Certaines de ces séquences sont interrompues par des nouveaux gestes endoscopiques/ radiologiques et/ou de la chirurgie.
- La trajectoire 4 (appelé **CHIR**) avec 1.631 patients (11,1%) est la classe la plus homogène caractérisée par de la chirurgie.
- La trajectoire 5 (appelé **CHIMIO-PALL**) avec 790 patients (5,3%) débute par un état de chimiothérapie étant suivi par l'état de soins palliatifs.
- La trajectoire 6 (appelé **CHIMIO-RADIO**) avec 366 patients (2,5%) débute par un état de chimiothérapie étant suivi par l'état de radiothérapie.

Parmi les six trajectoires, les trajectoires de CHIMIO-EXPLO et CHIMIO étaient prédominantes (Figure 4 et 5).

Figure 4. Tapis de séquences (index plot) montrant les différentes classes et leur dimension.

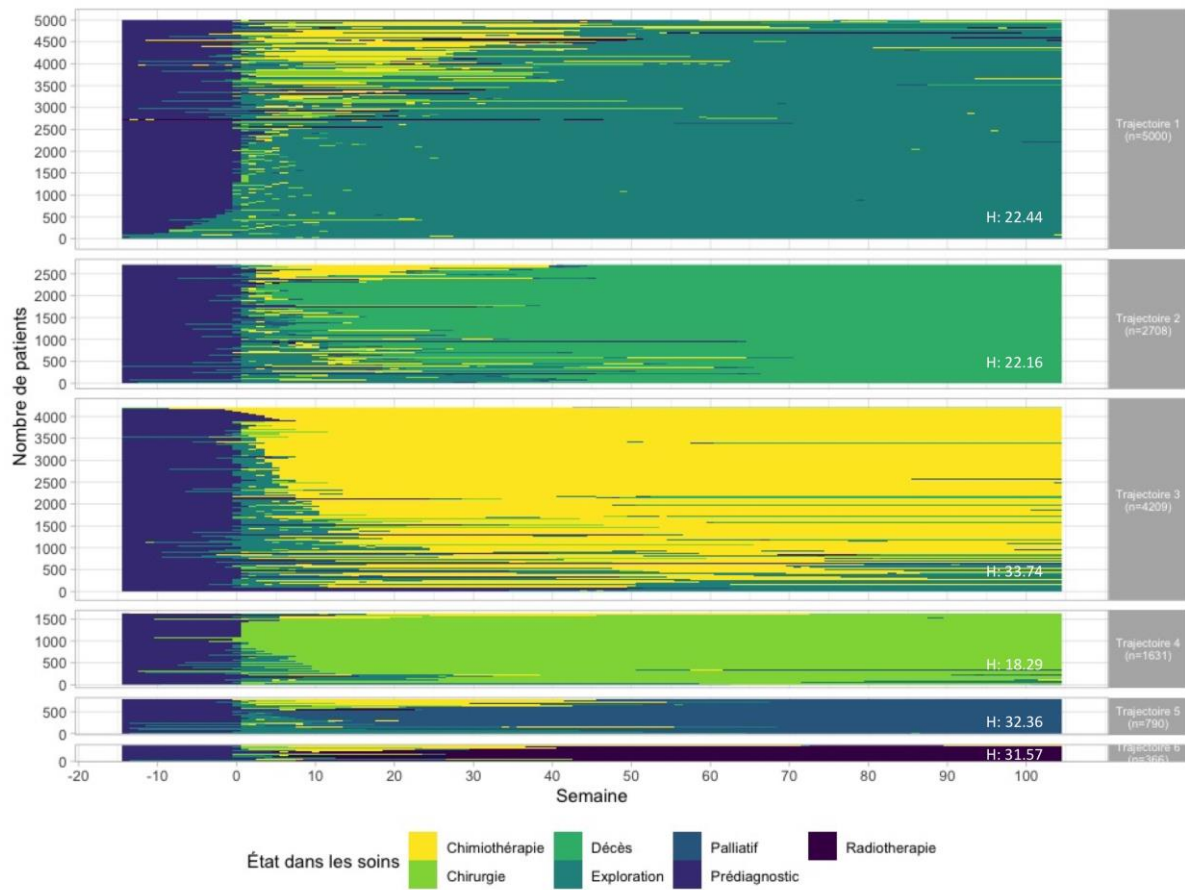
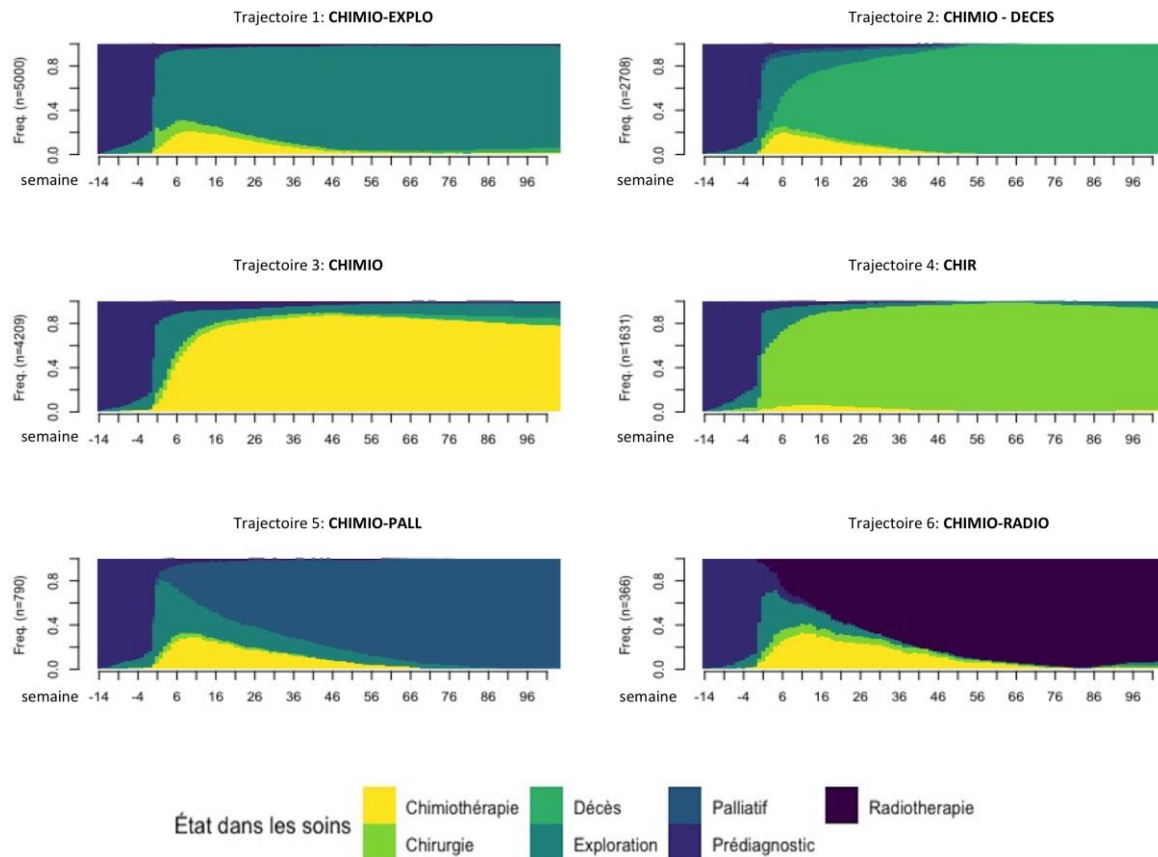


Figure 5. Chronogramme montrant pour chaque trajectoire les proportions d'individus de la classe dans les différents états.



MODELISATION

Après ajustement sur les cofacteurs et par rapport à la trajectoire CHIMIO, nous constatons une forte association entre l'âge élevé (>80 ans) et l'effet d'appartenir à la trajectoire CHIMIO-EXPLO [OR : 1,34-1,80, $p < 0,001$], CHIMIO-DECES [OR : 2,51-3,44, $p < 0,001$], CHIR [OR : 1,46-2,13, $p < 0,001$], et CHIMIO-PALL [OR : 3,53-5,57, $p < 0,001$].

Les patients qui ont un CC intrahépatique sont plus susceptibles d'appartenir à la trajectoire de décès précoce (CHIMIO-DECES) [OR : 1,23-1,66, $p < 0,001$], et à la trajectoire des soins palliatifs (CHIMIO-PALL), Le CC de la vésicule biliaire est fortement associé avec la trajectoire CHIR [OR : 2,88-4,40, $p < 0,001$] par rapport au CC intrahépatique et CC extrahépatique.

Quant au score de Charlson, les patients avec un score élevé (> 4) sont plus à risque d'appartenir à une trajectoire de chimiothérapie comme la CHIMIO-DECES [OR : 1,08-1,38, $p < 0,001$], ou une trajectoire de soins de confort comme la CHIMIO-PALL [OR : 1,23-1,81, $p < 0,001$] et inversement moins de chance d'appartenir à une trajectoire interventionnelle comme la CHIR [OR : 0,27-0,37, $p < 0,001$] ou la CHIMIO-EXPLO [OR : 0,61-0,75, $p < 0,001$].

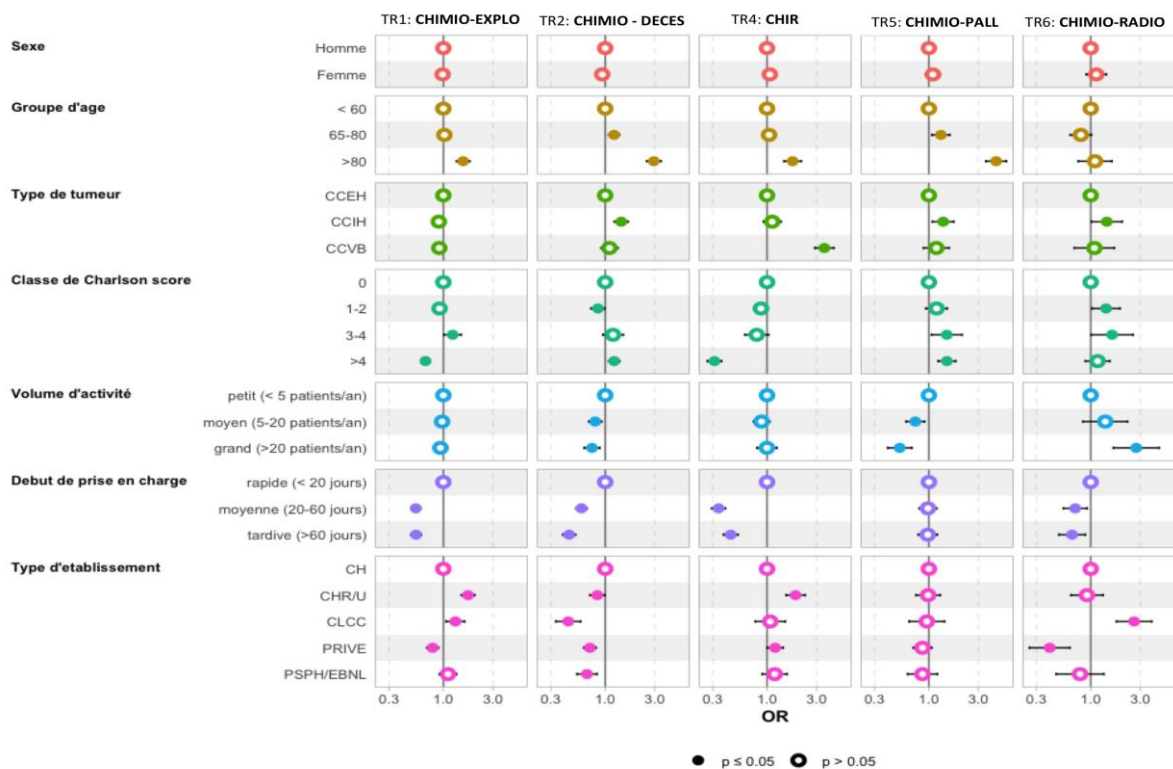
En ce qui concerne le type d'établissement, les patients adressés dans un CHR/U sont plus susceptibles d'appartenir à la trajectoire CHIR [OR : 1,52-2,33, $p < 0,001$] ou la trajectoire de CHIMIO-EXPLO [OR : 1,50-2,00, $p < 0,001$], ceux adressés dans un CLCC sont plus susceptibles d'appartenir à la trajectoire de CHIMIO-RADIO

[OR : 1,77-3,86, $p < 0,001$] et ceux adressés dans un centre privé plus susceptibles d'appartenir à la trajectoire CHIR [OR : 1,00-1,43, $p < 0,049$] et moins dans aux trajectoires CHIMIO-EXPLO [OR : 0,69-0,90, $p < 0,001$], CHIMIO-DECES [OR : 0,62-0,82, $p < 0,001$], et CHIMIO-RADIO [OR : 0,26-0,63, $p < 0,001$].

Pour le volume d'activité, il existe une forte association négative entre les trajectoires CHIMIO-DECES [OR : 0,63-0,89, $p < 0,001$] et CHIMIO-PALL [OR : 0,40-0,69, $p < 0,001$] et un volume d'activité de moyen – grand taille alors qu'une forte association positive existe entre un grand volume d'activité et la trajectoire de CHIMIO-RADIO [OR : 1,65-4,54, $p < 0,001$].

Enfin, les patients qui sont pris en charge tardivement ont moins de chance d'appartenir à la trajectoire CHIR [OR : 0,38-0,53, $p < 0,001$] ou n'importe quelle autre trajectoire par rapport à la trajectoire CHIMIO.

Figure 6. Analyse des OR des facteurs d'appartenance à différentes trajectoires.



DISCUSSION

En France, les cancers des voies biliaires sont rares : ils sont au deuxième rang des cancers primitifs du foie après les carcinomes hépatocellulaires [73]. L'évolution de cette pathologie dans ces 10-15 dernières années ainsi que les parcours de soins de l'ensemble des patients diagnostiqués en pratique courante sont mal connus. Dans cette étude rétrospective réalisée à partir des données du PMSI, nous avons essayé d'analyser la prise en charge actuelle du CC en France et d'identifier les différentes trajectoires thérapeutiques par la méthode d'analyse des séquences. Enfin nous avons essayé d'identifier des facteurs associés à l'appartenance ces trajectoires dans l'objectif de mieux comprendre et d'identifier des éventuelles pistes d'actions pour améliorer l'organisation des soins.

Nous avons inclus 51.123 patients entre 2011-2020 et nous avons retrouvé que l'incidence du CC est plus élevée que celle décrite dans la littérature avec près de 4.000 nouveaux cas par an, contre 2.000-3.000 à la dernière estimation réalisée en 2000 [68]. Cette augmentation du nombre des patients diagnostiqués avec CC a été aussi rapportée par Neuzillet et al. qui ont calculé une incidence de CC intrahépatique augmentée avec 1.824 nouveaux patients pour l'année 2014-2015 [74]. Cette augmentation pourrait être expliquée par le développement de techniques d'imagerie de plus en plus performantes qui permettent une détection plus fines des tumeurs des voies biliaires [75].

En termes de traitement nous avons pu constater que la majorité des patients ont reçu de la chimiothérapie (70%) et de la radiothérapie (73%) alors que seuls 29% des patients ont reçu le traitement curatif, c'est à dire la chirurgie (29).

De plus, le délai moyen de la prise en charge chirurgicale, par chimiothérapie ou par radiothérapie était respectivement de 66 jours (+/- 169), 127 (+/-270,3) et 370 jours (+/-509,8) et donc particulièrement long par rapport à la prise en charge des cancers plus fréquents comme le cancer du sein (22+/-13 jours), du poumon (13+/-10 jours), du colon (19+/-15 jours) ou de la prostate (45+/-30) en France (30).

Notre travail a montré que la majorité des séjours des patients avait lieu dans des CH (31,4%), et des CHU/R (29,0%). La plupart des patients sont donc pris en charge dans des centres autres que les CHR/U et CLCC qui sont des établissements plus susceptibles d'être spécialisés en cancérologie. De plus, 65% des centres dans lesquels le CC est pris en charge sont de petit volume d'activité, et notamment les CH (50%) et les centres privés (34,3%). Ce manque d'optimisation des soins pourrait avoir un effet péjoratif dans la prise en charge du CC et secondairement dans le pronostic de la maladie. Idrees et al. ont récemment rapporté l'importance de centralisation de la prise en charge de cholangiocarcinome en montrant une amélioration de survie des patients lorsqu'ils étaient pris en charge dans des centres à haut volume [76].

La méthode de l'analyse des séquences que nous avons utilisée dans notre travail a permis d'identifier 6 trajectoires de soins dont les 3 trajectoires principales sont des trajectoires palliatives. Parmi les 14.704 patients étudiés, 4.209 (28,6%) patients suivent une trajectoire de chimiothérapie seule (CHIMIO) puisqu'ils restent dans l'état de chimiothérapie tout le long de leur trajectoire. 5.000 patients (34%) suivent une trajectoire dominée par des gestes interventionnels radiologiques et/ou endoscopiques (CHIMIO-EXPLO) très probablement à but palliatif (ex. traitement de l'ictère) puisqu'ils sont accompagnés par de la chimiothérapie seulement dans 20% des cas. Enfin, 2.700 patients (18%) suivent une trajectoire de décès précoce (CHIMIO-DECES) puisque à 8 mois presque la totalité des patients sont décédés. La trajectoire de la chirurgie, seule trajectoire à but curatif, n'est réservée qu'à 1.631 (11%) des patients avec un CC. Il est aussi intéressant de remarquer qu'aucun patient n'a reçu de la chimiothérapie dans cette trajectoire. Cette séquence changera dans le futur puisque depuis 2019 une chimiothérapie adjuvante est administrée en postopératoire [76]. Les trajectoires de soins de confort (CHIMIO-PALL) et de la radiothérapie (CHIMIO-RADIO), ne sont suivies que par une petite partie des patients (7,8%). Une observation intéressante dans l'ensemble de ces trajectoires (hormis le CHIMIO-DECES) est l'absence de l'état de décès au bout de 2 ans d'observation. La découverte de la maladie à un stade souvent avancé et l'absence de moyens de

traitement pourraient conduire les patients de rester à domicile et décéder en dehors de l'hôpital. C'est probablement aussi la raison pour laquelle la durée moyenne de suivi dans notre étude était seulement de 9,5 mois (+/-16,9). L'information du décès en dehors de l'hôpital n'est pas disponible dans le PMSI-MCO.

L'analyse multivariée nous a permis de mettre en évidence des associations importantes entre les facteurs étudiés et certaines trajectoires. Par exemple l'appartenance à une trajectoire curative (trajectoire CHIR) est plus probable pour un patient atteint d'un CC de la vésicule biliaire, en bon état général (score de Charlson <4) et qui est pris en charge rapidement dans un CHR/U ou dans un centre privé. Inversement, les patients qui ont plus de chance d'appartenir à une trajectoire palliative (CHIMIO-DECES, CHIMIO-PALL) sont des patients âgés avec un mauvais état général (score de Charlson >4), atteints d'un CC intrahépatique et dont la prise en charge est faite dans un établissement à faible volume d'activité.

Notre travail est le premier à avoir utilisé des données récentes et exhaustives du PMSI pour construire des trajectoires de patients atteints d'un CC. L'inclusion d'un aussi large effectif fait partie des forces. Une autre force de ce travail est la construction des parcours à l'aide de l'analyse des séquences qui est à la fois une méthode puissante, intuitive et facilite la visualisation des différentes trajectoires.

LIMITES

Les limites de cette étude concernent sa nature rétrospective et le fait qu'elle soit menée à partir de base de données medico-administratives. En effet, le PMSI ne dispose pas d'information clinique (par exemple le stade de la tumeur, l'anatomopathologie, etc.) et la construction de trajectoires en cancérologie est donc limitée. L'absence de codage spécifique pour chaque type de CC ainsi que la présence de multiples codes par patient a rendu l'analyse plus difficile et moins précise. L'interprétation doit donc être prudente. L'absence d'information sur des traitements réalisés en dehors de l'hôpital pourrait conduire à une mauvaise analyse des séquences et construction des trajectoires. Néanmoins, le CC reste une pathologie complexe et sa prise en charge se fait dans la majorité des cas dans le milieu hospitalier. Enfin, nous n'avons pas pu faire une analyse approfondie des trajectoires concernant le devenir des patients puisque l'information sur le statut vital était souvent absente.

CONCLUSION

En conclusion notre étude a permis d'identifier les principales trajectoires des patients atteints d'un CC en France entre 2011-2020 et d'analyser les divers facteurs influençant l'appartenance à ces trajectoires qu'ils soient intrinsèques à la personne (sexe et âge), liés à la maladie elle-même (type de tumeur) ou à la filière de prise en charge (caractéristiques de l'établissement). Enfin, elle a permis de mettre en évidence des problématiques d'organisation de la prise en charge de ces patients, notamment sur le délai de cette prise en charge et le manque de centralisation dans les centres experts.

PERSPECTIVES DE RECHERCHE

L'analyse du parcours de soins est un sujet de recherche assez récent, sans doute favorisé par la disponibilité des données de volume important, organisées de la sorte qu'il est possible de suivre la consommation de soins d'un patient de manière longitudinale.

L'accessibilité à des bases des données nationales permet de ne pas se limiter à des données produites dans un seul établissement, dans une cohorte ou dans une région, mais de pouvoir analyser l'exhaustivité de la population française quel que soit l'endroit où le soin est administré.

Ce type d'analyse, comme nous l'avons montré, semble approprié aux parcours complexes comme celui du cancer et de l'obésité.

L'analyse des trajectoires de patients opérés de chirurgie bariatrique a été centrée dans la première étude sur les événements postopératoires précoces et dans la deuxième sur la chirurgie itérative.

L'analyse des événements qui se produisent dans l'année qui suit la chirurgie a permis à la fois de retrouver des complications postopératoires connues (hémorragie digestive, dysphagie, douleur abdominale, fistule etc.) mais aussi d'identifier des événements moins attendus, comme la grossesse (fortement déconseillée dans l'année qui suit la chirurgie), les complications lithiasiques biliaires et les hernie/éventrations abdominales.

Dans la deuxième étude sur la chirurgie bariatrique nous avons analysé le volet itératif de cette chirurgie. En effet, l'obésité étant une maladie chronique et la chirurgie étant un traitement efficace mais imparfait, une partie de cette population est opérée à plusieurs reprises. Il est extrêmement difficile d'appréhender avec précision cette succession d'interventions à partir des données issues d'études interventionnelles ou observationnelles à cause du délai parfois long entre deux événements et de la variabilité géographique. A notre connaissance aucune autre base de données au monde ne permet de chaîner une entière population sur tout le territoire national sur une durée de dix ans.

L'analyse que nous avons menée sur la prise en charge du CC a été construite autour de différents axes thérapeutiques (chirurgie, chimiothérapie, radiothérapie). Cela génère des parcours de soins divers qui peuvent être réunis en quelques trajectoires principales. Nous avons donc pu estimer la réelle prévalence de cette tumeur (à priori sous-estimée en France jusqu'au présent) et la prévalence des différentes classes, caractérisées par des profils de patients différents, ainsi que par de délais de prise en charge et un taux de survie qui sont assez variables.

Pour la suite de nos recherches, nous avons identifié quelques points de réflexion.

1. **Des nouvelles méthodes.** Nous avons testé deux méthodes d'analyse des trajectoires (l'analyse formelle de concept et l'analyse des séquences). Il en existe d'autres qui mériteraient d'être explorées, telles que l'analyse symbolique des données ou l'analyse des motifs séquentiels contextuels).
2. **Choix de la méthode.** Le choix de la bonne méthode n'est pas aujourd'hui bien codifié et il est encore basé sur l'intuition, et testé de manière empirique comme nous l'avons fait. En effet nous avons analysé les complications chirurgicales dans l'AFC et l'analyse des séquences (étude 1 et 2). La seconde méthode s'est révélée complètement inadaptée à la thématique. Il serait donc intéressant de recenser toutes les méthodes existantes et de proposer un algorithme décisionnel en fonction de la nature des données et de la question posée. Sur cet axe un projet de thèse de doctorat a été déposé en septembre 2021.
3. **Validation de l'information.** Un sujet problématique majeur que nous avons rencontré concerne la validité de l'information des BDMA. Cela se décline à plusieurs niveaux de la construction de la base

de données : l'identification de la population d'intérêt, la caractérisation de la population, et l'identification des événements d'intérêt. Ce point rejoint le deuxième axe de recherche sur la validation des algorithmes d'identification de l'information pour lequel une thèse de doctorant est déjà en cours.

4. **Données supplémentaires.** Nos analyses ont porté sur les données des hospitalisations, à partir des données PMSI-MCO. D'autres sources de données existent et pourraient enrichir l'information existante : les données sur les causes de mortalité, les données sur la consommation des soins en extrahospitalier, les entrepôts des données hospitaliers. Le SNDS a récemment rendu plus simple l'accès aux données du CépiDc (cause de mortalité), et du SNIIRAM (consommation des soins libéraux). Nous comptons intégrer ces nouvelles sources à nos différents projets de recherche.

AXE II : INFERENCE CAUSALE ET DONNEE OBSERVATIONNELLES

CONTEXTE GENERAL

Pour répondre à des questions causales sur l'efficacité ou la sécurité comparatives entre deux groupes, les essais randomisés contrôlés sont le *gold standard* en recherche biomédicale.

Lorsqu'on ne peut pas mener une expérience randomisée, pour des raisons éthiques, logistiques ou économiques, on peut se tourner vers les données observationnelles.

Nous abordons deux méthodes de recherche qui visent à réaliser de l'inférence causale à partir des données d'observation : l'émulation d'un essai cible et l'utilisation des variables instrumentales.

EMULATION D'UN ESSAI RANDOMISE A PARTIR DES DONNEES OBSERVATIONNELLES

La première méthode que nous avons explorée consiste à émuler un essai randomisé (l'essai cible ou « *target trial* ») qui répondrait à la question d'intérêt à partir de grandes bases de données médicales. Lorsque l'objectif est de guider les décisions entre plusieurs stratégies, les analyses causales des données observationnelles peuvent être évaluées en fonction de leur efficacité à émuler un essai cible.

Les essais randomisés sont utilisés pour identifier et quantifier les effets causaux moyens parce que l'assignation aléatoire du traitement conduit à une comparabilité entre groupes.

Les études observationnelles, en revanche, peuvent être beaucoup moins convaincantes et présenter des biais de sélection importants. Une des principales hésitations à attribuer une interprétation causale aux associations observationnelles est, donc, l'absence de l'assignation aléatoire du traitement.

En effet, l'association entre le traitement et le résultat dans une étude observationnelle n'est pas nécessairement un effet causal du traitement sur le résultat. La notion que « association » n'est pas égale à « causalité », est en effet un principe fondamental en épidémiologie.

Tout en reconnaissant que les études randomisées ont des avantages intrinsèques pour l'inférence causale, parfois nous ne disposons que d'informations issues d'études d'observation pour répondre à des questions causales.

Dans ce cas, nous pouvons tenter d'analyser les données comme si le traitement avait été assigné de manière aléatoire en fonction des covariables mesurées, même si nous savons qu'il s'agit au mieux d'une approximation.

L'inférence causale à partir des données d'observation tourne alors autour de l'espoir que l'étude observationnelle puisse être considérée comme une expérience aléatoire. Ce cadre méthodologique se base sur la théorie contrefactuelle (« *counterfactual theory* ») qui ne sera pas rediscutée ici [77].

Une étude observationnelle peut être conçue comme une expérience randomisée aux trois conditions suivantes :

1. « *Consistency* » : les valeurs de traitement comparées correspondent à des interventions bien définies qui, à leur tour, correspondent aux versions de traitement dans les données.
2. « *Exchangeability* » : la probabilité conditionnelle de recevoir chaque valeur du traitement, bien que non décidée par les chercheurs, dépend uniquement des covariables mesurées.

3. « *Positivity* » : la probabilité conditionnelle de recevoir chaque valeur de traitement est supérieure à zéro, c'est-à-dire positive. Cela signifie que chaque patient inclus dans l'étude pourrait potentiellement être alloué à chaque bras de l'étude, et qui ne présente donc pas des caractéristiques qui dans un essai randomisé contrôlé le rendrait non éligible à un des bras de l'étude.

Ces conditions sont souvent difficiles à appréhender, ce qui explique pourquoi les inférences causales issues des études observationnelles sont considérées avec méfiance.

Notre premier travail d'inférence causale à partir des données observationnelles a voulu estimer l'effet de la perte de poids induite par la chirurgie bariatrique sur l'incidence de cancer. Les patients en état d'obésité ont donc été répartis dans un groupe chirurgie bariatrique et dans un groupe sans chirurgie.

Nous allons utiliser cet exemple pour expliquer la construction d'un protocole pour une étude observationnelle voulant émuler un essai randomisé. Ceci doit tenir compte de conditions de validité (*consistency, positivity, exchangeability*) et les intégrer dans les différentes étapes: les critères d'éligibilité, les stratégies de traitement comparées, les procédures d'assignation, la période de suivi, l'objectif d'intérêt, les contrastes causals, et le plan d'analyse.

CRITERES D'ELIGIBILITE.

L'analyse observationnelle doit appliquer les mêmes critères d'éligibilité de l'essai cible. De plus, il faut considérer que les critères d'éligibilité ne peuvent pas inclure des exclusions basées sur des événements postérieurs au point de départ (*post baseline*) qui peuvent donc introduire un biais dans l'analyse. En effet des informations plus tardives ne pourraient pas s'appliquer au moment d'une réelle randomisation d'un vrai essai randomisé. Si toutes les informations qu'on utiliserait dans un essai randomisé ne sont pas disponibles dans les données observationnelles, cela peut être problématique. Il faut donc décider s'il est possible de procéder à l'émulation d'un essai cible différent et si cela aborde aussi une question d'intérêt.

Dans notre exemple nous avons appliqué comme restriction les critères d'indication à la chirurgie bariatrique : IMC > 40 kg/m² ou IMC 35-40 kg/m² en présence d'une comorbidité susceptible de s'améliorer après la perte de poids. Parmi les comorbidités retenues pour valider l'indication, nous avons considérés les plus fréquentes : le diabète de type 2, l'hypertension artérielle, la dyslipidémie, et le syndrome d'apnée du sommeil. Ils existent aussi des contre-indications à la chirurgie bariatrique, essentiellement d'ordre psychiatrique et associées à certaines addictions. Nous les avons prises en compte parmi les critères d'exclusion. Tous ces critères d'éligibilité ont été appliqués aux deux groupes, en conséquence, les patients du groupe chirurgie qui ne remplissaient pas les critères d'inclusion ont aussi été exclus. Cette attitude permet ainsi de respecter la condition de « *positivity* », qui implique que chaque patient présente les caractéristiques pour être potentiellement inclus dans chaque bras.

TRAITEMENTS COMPARES.

L'essai cible émulé à l'aide de données observationnelles sera typiquement un essai pragmatique, c'est-à-dire un essai dans lequel les stratégies de traitement sont comparées dans les conditions de vie réelle. Par exemple, on ne peut pas imiter un essai contrôlé par placebo avec une surveillance étroite.

Pour émuler notre essai cible, nous identifions dans la base de données les personnes qui répondent à tous les critères d'éligibilité. Nous les attribuons ensuite à la stratégie d'essai qui est cohérente avec leurs données de base. Dans notre étude nous avons identifié les patients par la présence d'une obésité et nous les avons attribués à un bras ou à l'autre en fonction de leurs caractéristiques de base (chirurgie ou pas chirurgie). Le bras interventionnel est donc défini par l'acte chirurgical standardisé, ce qui respecte donc la condition de « *consistency* ».

PROCEDURE D'ASSIGNATION.

On ne peut émuler que des essais cibles sans affectation en aveugle, ce qui est la conception standard des essais pragmatiques. En effet, les individus dans l'ensemble de données et leurs médecins sont généralement avertis des traitements que les participants reçoivent. Ce n'est pas nécessairement une limitation si l'objectif est de comparer les stratégies de traitement en vie réelle chez des personnes conscientes de leurs soins. Pour émuler l'assignation aléatoire des stratégies au départ, nous devons ajuster tous les facteurs de confusion nécessaires pour assurer la comparabilité (« *exchangeability* ») des groupes [77]. L'ajustement pour les facteurs de confusion peut être effectué via l'appariement (sur le score de propension par exemple), la stratification, la régression, la standardisation ou la pondération de probabilité inverse, etc.

Si la base de données observationnelle ne contient pas suffisamment d'informations sur les facteurs de confusion ou si on ne parvient pas à les identifier, il ne sera pas possible d'obtenir une émulation réussie de la randomisation. Les biais de confusion peuvent être particulièrement graves lors de l'émulation d'essais cibles qui comparent un traitement à l'absence de traitement (ou aux soins habituels) plutôt qu'à un autre traitement actif [78], [79].

Bien qu'il soit généralement impossible de déterminer si l'émulation a échoué en raison d'un facteur de confusion non mesuré, plusieurs approches indirectes peuvent alerter sur ce biais. On rappellera ici les « *negative control outcomes* », dont la définition formelle indique qu'il s'agit d'un objectif qui partage les mêmes sources potentielles de biais avec le *outcome* principal, mais ne peut pas être lié au traitement d'intérêt [80]. L'idée est qu'on ne devrait pas voir d'effet sur les « *negative control outcome* » en absence de facteurs de confusion non mesurés. Dans notre étude nous avons analysé séparément les cancers qui sont associés à l'obésité et ceux qui ne le sont pas [81]. On pourrait s'attendre logiquement que la perte de poids induite par la chirurgie et donc la réduction voire la disparition de l'obésité ait un effet sur les cancers associés à l'obésité mais pas un effet sur les autres cancers. En conséquent, nous avons considéré l'incidence des cancers non associés à l'obésité comme notre « *contrôle outcome* ».

PERIODE DE SUIVI

Dans un essai randomisé, le début naturel du suivi est le moment où la stratégie de traitement est attribuée par randomisation, ce qui se produit souvent peu de temps avant ou en même temps que le début de la stratégie de traitement. Commencer après la randomisation pourrait entraîner un biais de sélection car tous les événements d'intérêt entre la randomisation et le temps zéro seraient exclus de l'analyse.

Avec des données observationnelles, la meilleure façon d'émuler le temps zéro de l'essai cible est de définir le temps zéro comme étant le moment où une personne éligible initie une stratégie de traitement. Les critères d'éligibilité doivent être remplis à ce stade, mais pas plus tard ; les objectifs de l'étude commencent à être comptés après ce point, mais pas avant.

Dans notre exemple, les patients du groupe chirurgie sont donc inclus lors de la chirurgie elle-même et les patients non chirurgicaux, lors de leur premier séjour disponible.

OBJECTIF

Une évaluation de l'objectif indépendante et à l'aveugle est certainement l'idéal afin de garantir que la connaissance du bras de traitement n'influence pas l'évaluation de l'objectif. Dans notre exemple, une telle détermination différentielle peut entraîner une incidence accrue de diagnostic de cancer chez les patients opérés de chirurgie bariatrique, même en l'absence d'effet biologique. Néanmoins, étant donné que les médecins sont généralement au courant du traitement reçu par l'individu, nous ne pouvons pas utiliser les données observationnelles pour émuler un essai cible avec une détermination systématique et aveugle des résultats, sauf lorsque la détermination des résultats ne peut pas être affectée par les antécédents de traitement (par exemple, si le résultat est le décès et est déterminé de manière indépendante à partir d'un registre des décès).

CONTRASTES CAUSAUX D'INTERET & PLAN D'ANALYSE

Deux effets sont habituellement évalués dans les vrais essais randomisés : l'effet en intention de traiter (ITT, c'est-à-dire l'effet comparatif d'être affecté aux stratégies de traitement au départ, que les individus continuent ou non à suivre les stratégies après le départ) et l'effet per protocole (PP, c'est-à-dire l'effet comparatif : effet de suivre les stratégies de traitement spécifiées dans le protocole de l'étude). Si les effets en intention de traiter et per protocole présentent un intérêt dans l'essai cible, nous essaierons d'estimer les analogues des deux effets à partir de nos données d'observation. Dans notre cas, certains patients opérés de chirurgie bariatrique (notamment d'anneau gastrique) pourraient ne pas rester dans le groupe chirurgie : l'anneau gastrique peut être retiré par exemple. Dans ce cas on pourrait imaginer une analyse en ITT et une en PP. Le plan d'analyse prévoit une comparaison en intention de traiter des objectifs des deux groupes. Dans notre exemple nous allons comparer l'incidence de cancer dans les deux groupes par une analyse de survie.

Les estimations des effets à partir des données observationnelles sont bien définies lorsque l'on est capable de construire l'analyse d'observation sur un essai cible particulier. Cependant, on est rarement en mesure d'imiter l'essai idéal qui nous intéresse le plus. Ainsi un certain nombre de compromis sont souvent faits concernant les critères d'éligibilité, les stratégies à comparer, etc. La spécification du protocole de l'essai cible sera généralement un processus itératif au cours duquel nous apprendrons quels essais cibles particuliers peuvent être raisonnablement étayés par les données observationnelles [82]. De tous ces essais cibles possibles, nous choisirons celui qui se rapproche le plus de l'essai idéal que nous aurions aimé mener pour répondre à notre question. Nous pourrons ensuite décrire un protocole, présenter un arbre décisionnel, résumer comment l'ensemble de données observationnelles est utilisé pour émuler l'essai cible et expliquer en quoi l'essai cible diffère de l'essai idéal.

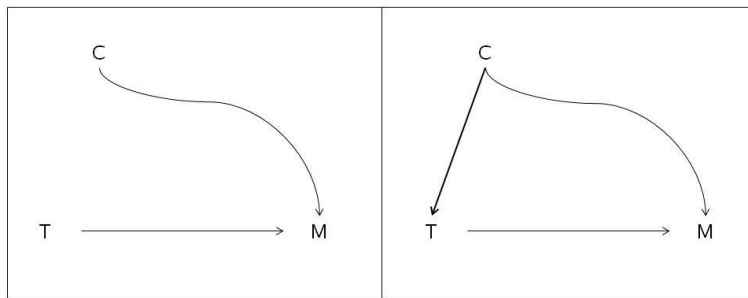
VARIABLES INSTRUMENTALES

Lorsque des facteurs de confusion ne sont pas mesurés, la comparaison entre les groupes risque d'être biaisée et empêche une analyse causale entre exposition/traitement et effet.

Il existe donc une autre approche possible de l'inférence causale à partir de données d'observation : espérer qu'un facteur de prédiction du traitement, appelé variable instrumentale (« *instrumental variable* »), soit alloué de manière aléatoire. Les variables instrumentales prennent donc en compte à la fois les facteurs de confusion mesurés et non mesurés. Elles sont largement utilisées dans le domaine de l'économie et des sciences sociales et ont été introduites que récemment dans la recherche biomédicale [83].

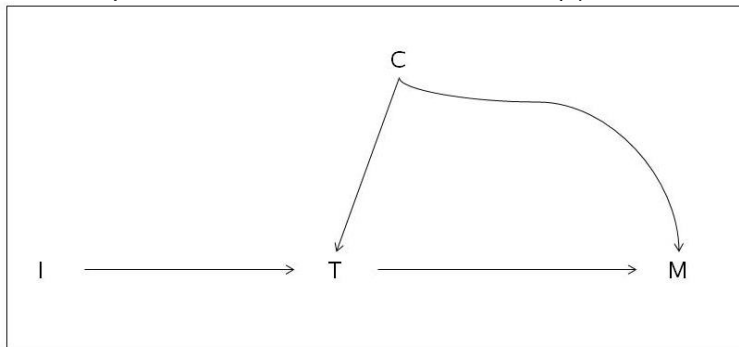
On peut représenter les relations d'association avec un « *direct acyclic graph* » (DAG). Dans un essai randomisé contrôlé, le traitement (T) est alloué entre les deux groupes de manière indépendante des facteurs de confusion (C), tels que les caractéristiques des patients, alors que dans une étude observationnelle le risque est que les facteurs de confusion soient associés au traitement T représentent donc un biais de sélection (figure 1).

Figure 1. DAG de la relation entre exposition/traitement (T), effet/maladie (M) et facteurs de confusion (C), à droite dans un essai randomisé contrôlé, à gauche dans une étude observationnelle.



Une variable instrumentale est un phénomène naturel qui randomise les sujets pour une exposition ou un traitement dans les différents groupes. De ce fait, les patients seraient assignés à chaque groupe de manière équilibrée à la fois pour les facteurs de confusion mesurés et pour les facteurs de confusion non mesurés, sans avoir aucun effet sur l'objectif de l'étude. Dans la représentation graphique de la figure 2 on observe que la variable instrumentale (I) est associée au traitement (T) mais pas aux facteurs de confusion.

Figure 2. La variable I semble être une variable instrumentale appropriée car elle est associée au traitement T, et n'est pas associée aux facteurs de confusion (C).



Une variable instrumentale, pour être plausible, doit remplir 3 conditions : (1) elle doit être associée à l'exposition/traitement ; (2) elle doit être sans rapport avec les facteurs de confusion (au moins après ajustement pour les facteurs de confusion mesurés) ; (3) elle ne doit pas avoir d'effet direct sur le résultat, sauf par son effet sur l'exposition/traitement [84].

Parmi les types les plus communs de variable instrumentale, on peut citer:

1. **Génotype.** Un gène qui soit associé à l'exposition/traitement mais pas à l'effet/maladie, pourrait être utilisé comme variable instrumentale. Par exemple le gène ALDH2 est utilisé pour analyser l'association entre consommation d'alcool et pathologies cardiovasculaires. En effet ce gène est impliqué dans le métabolisme de l'alcool; les personnes qui portent les copies inactives du gène subissent des rougeurs faciales et d'autres symptômes indésirables après avoir consommé de l'alcool, et donc ils présentent une consommation d'alcool faible voire nulle. Ce gène par contre n'a aucune association avec les pathologies cardiovasculaires [85].
2. **Géographie.** Pour étudier l'impact de l'admission dans un « stroke center » sur la mortalité par AVC, il est possible d'utiliser la distance différentielle jusqu'à un stroke center comme variable instrumentale. Par distance différentielle on entend la distance entre la résidence d'un patient et le stroke center le plus proche moins la distance entre la résidence d'un patient et l'hôpital le plus proche de quelque sorte que ce soit[86].
3. **Pratiques médicales.** Pour examiner l'impact de la rééducation précoce sur la mortalité hospitalière due à la pneumonie par inhalation, on peut, par exemple, s'appuyer sur la proportion de patients subissant une rééducation précoce dans chaque hôpital comme variable instrumentale. La

fréquence à laquelle les hôpitaux fournissent ce service reflète davantage les modèles de pratique hospitalière que les caractéristiques des patients [87].

4. **Tendances de prescription dans le temps.** Pour étudier l'impact de l'hormonothérapie substitutive sur le risque d'accident vasculaire cérébral et de crise cardiaque, des chercheurs ont utilisé le calendrier comme variable instrumentale. L'utilisation de l'hormonothérapie substitutive chez les femmes ménopausées était répandue avant 2002, mais a fortement baissé en 2002 en raison des résultats de l'essai randomisé de la *Women's Health Initiative* qui a révélé une augmentation inattendue du risque de maladie cardiovasculaire chez les utilisatrices d'hormones [88].

Cependant toute la validité de l'analyse par variables instrumentales réside dans la validité même de l'outil. Il est donc primordial d'essayer de vérifier les conditions de validité.

1. **Association entre la variable instrumentale et l'exposition/traitement.** Ceci est la première difficulté, car le choix de la variable instrumentale doit être plausible et justifiable. La justification doit être à la fois empirique et statistique. En effet une justification solide doit être fournie pour expliquer pourquoi la variable instrumentale choisie peut être considérée comme fortement liée à l'attribution du traitement. Ces justifications théoriques doivent être étayées par des preuves empiriques d'une forte association avec le traitement dans la première étape de la modélisation.
2. **Absence d'association entre la variable instrumentale et les facteurs de confusion.** Tester si l'instrument est lié à des facteurs de confusion est plus difficile. Il est possible de vérifier que les facteurs de confusion mesurés soient équilibrés entre les groupes, mais il n'est pas possible de faire de même avec les facteurs de confusion non mesurés. Pour vérifier l'équilibre des facteurs de confusion mesurés, il est souhaitable d'utiliser les différences standardisées (*standard mean difference*, SMD) plutôt que le *p-value*. Une différence standardisée <10 % est considérée comme acceptable. Par contre, s'il n'y a aucun moyen de vérifier l'équilibre dans les facteurs de confusion non mesurés, il est cependant possible d'effectuer des analyses de sensibilité pour examiner l'impact potentiel des facteurs de confusion non mesurés sur les résultats.
3. **La variable instrumentale n'a pas d'impact direct sur l'outcome.** Encore une fois cela est difficile à tester empiriquement : une réflexion approfondie, des preuves provenant d'autres publications et des analyses de sensibilité peuvent aider à évaluer l'impact potentiel d'effets non observés.

En pratique, l'utilisation des variables instrumentales passe par un modèle de régression en deux étapes : la première modélise la probabilité de recevoir un traitement donné à partir de la variable instrumentale ; la deuxième étape modélise l'effet de l'exposition basé sur la variable instrumentale sur l'outcome. Si des facteurs de confusion connus sont inclus, ils doivent être présents pour les deux étapes.

Modèle 1 : exposition/ traitement = variable instrumentale + facteurs de confusion

Modèle 2 : outcome = exposition/ traitement estimé (modèle 1) + facteurs de confusion

Enfin, il faut s'interroger sur quelle population les effets des variables instrumentales s'appliquent. Pour comprendre cela il faut considérer que dans un essai on observe deux types d'individus: les observants (ceux qui sont tirés au sort dans le groupe A et qui assument effectivement le traitement A) et les non-observants (ceux qui sont tirés au sort dans le groupe A et qui assument effectivement le traitement B). L'analyse par variables instrumentales ne prend pas en compte les individus non-observants. Cela implique que les résultats de cette analyse ne sont pas généralisables aux individus non-observants. Ce problème ne subsiste pas dans les essais randomisés car l'analyse en intention de traiter évalue l'effet sur toute la population (observants et non-observants) et l'analyse per-protocole analyse les résultats pour les individus observants.

VALIDATION DE DONNEES

Toute cette réflexion méthodologique sur l'inférence causale à partir de données observationnelles se base sur le présupposé que les données soient de bonne qualité. Cela implique que la base de données ait fait l'objet d'études de validation qui confirment que l'outcome recherché correspond au vrai diagnostic. L'absence de validation pourrait en effet amener à un biais d'estimation de l'effet comme déjà décrit dans la littérature [89]–[91]. Dans le cas du PMSI, de nombreuses études ont porté sur la validation des données [92]–[95].

Néanmoins, la qualité du codage reste hétérogène, en particulier pour les codes des diagnostics selon la CIM-10. Plusieurs solutions sont possibles pour diminuer ce biais potentiel. L'évaluation des comorbidités est important dans l'analyse de la morbi-mortalité, car elles agissent comme facteurs de confusion et il fait désormais partie de la pratique courante d'ajuster l'analyse sur les comorbidités dans l'analyse des *outcomes*. Un mauvais codage de ces comorbidités peut avoir un impact sur l'analyse statistique. L'accès à des données longitudinales sur plusieurs années du PMSI permet d'améliorer la description des comorbidités. En effet il est possible d'établir une période d'analyse rétrospective (« *lookback period* ») pour renforcer l'identification des données [96], [97].

La validation des données est donc d'importance majeure, mais ce sujet est souvent négligé dans les publications sur des vastes bases de données. Par exemple, dans des récentes études sur le même sujet (impact de la chirurgie bariatrique sur l'incidence de cancer), aucune étude de validation n'est mentionnée [98]–[100]. Et pourtant cet aspect est bien inclus parmi les étapes importantes des publications basées sur des données observationnelles collectées de manière automatique (i.e. les bases des données médico-administratives). Les recommandations du *RECORD Statement* à ce sujet indiquent en effet que « *Any validation studies of the codes or algorithms used to select the population should be referenced* » [46]. Comme les recommandations PRISMA sont devenues un référentiel incontournable pour les revues systématiques, permettant une amélioration de la qualité de celles-ci, il serait important que les études basées sur BDMA s'inspirent davantage de ces guidelines.

RECHERCHE & ENCADREMENT

Année	Etudiant	Projet	Cursus	% d'encadrement
2020-2021	Salomé Epaud	Impact de la chirurgie bariatrique sur l'incidence de cancer	Data science – Centre de recherche Clinique du CHI de Créteil – Kaduceo SAS	100%
2021-2022	Seydou Goro	Comparaison de deux stratégies chirurgicales de chirurgie bariatrique de révision	Data science – Centre de recherche Clinique du CHI de Créteil	100%
2019-	Alexandre Challine	Un score de Morbidité postopératoire « Clavien-Dindo » adapté à la recherche sur le PMSI en chirurgie digestive oncologique	Doctorat ED 393 : Pierre Louis de santé publique. Sorbonne université	90%
2019-2020	Hugo Perrichon	Un score de Morbidité postopératoire « Clavien-Dindo » adapté à la recherche sur le PMSI en chirurgie digestive oncologique	Stage de 2 ^{ème} année ENSAI - Ecole nationale de la statistique et de l'analyse de l'information	30%

ETUDE N°1

“IMPACT OF BARIATRIC SURGERY ON OCCURRENCE OF CANCER: RESULTS FOR AN EMULATED TARGET TRIAL USING DATA FROM 1.1 MILLION PATIENTS ON A NATIONWIDE SCALE”

ETUDE EN COURS. SUBMITTED TO THE BR J SURG, OCT 2021

RESUME

INTRODUCTION

L'impact de la perte de poids intentionnelle induite par la chirurgie bariatrique sur l'incidence du cancer est encore controversé. Afin d'identifier l'effet causal de la chirurgie bariatrique sur le risque de cancer à partir d'une base de données observationnelles, nous avons adopté une méthodologie visant à émuler un essai randomisé contrôlé.

METHODES

Les données sur les patients admis entre 2010 et 2019 avec un diagnostic d'obésité ont été extraites d'une base de données nationale d'hospitalisations. Les critères d'inclusion comprenaient les critères d'éligibilité pour l'obésité et l'absence de cancer dans les 2 ans suivant l'inclusion. Les bras d'intervention étaient la chirurgie bariatrique versus l'absence de chirurgie. Les objectifs étaient la survenue de tout cancer lié à l'obésité, tandis que les cancers non liés à l'obésité ont été utilisés comme « *negative control outcome* ».

RESULTATS

Un total de 1.140.300 patients potentiellement éligibles pour une chirurgie bariatrique ont été inclus dans l'étude. Parmi ces patients, 288.600 (25,3 %) patients ont subi une chirurgie bariatrique. Au total, 48.411 cancers ont été identifiés, dont 4.483 chez des patients chirurgicaux et 43.928 chez des patients n'ayant pas reçu de chirurgie bariatrique. La chirurgie bariatrique est associée à une diminution du risque de cancer lié à l'obésité de 0,89 (IC à 95 %, 0,83-0,95), alors qu'aucun effet significatif de la chirurgie n'a été identifié en ce qui concerne les cancers non liés à l'obésité (risque relatif était de 0,96 (95 % IC, 0,91-1,01)).

CONCLUSIONS

Lors de l'émulation d'un essai cible à partir de données d'observation, nos résultats suggèrent une réduction de 11 % du cancer lié à l'obésité après une chirurgie bariatrique. L'absence d'association observée pour le résultat du contrôle négatif suggère que le biais de sélection pourrait être contrôlé.

ABSTRACT

BACKGROUND

The impact of intentional weight loss induced by bariatric surgery on cancer occurrence is still controversial. In order to identify the causal effect of bariatric surgery on risk of cancer from an observational database, we adopted a methodology aimed at emulating a target randomized controlled trial.

METHODS

Data on patients admitted between 2010 and 2019 with a diagnosis of obesity were extracted from a national discharge database. Criteria for inclusion included eligibility criteria for bariatric and the absence of cancer in the 2 years following inclusion. The intervention arms were bariatric surgery versus no surgery. The outcomes were the occurrence of any cancer and obesity-related cancer, while cancers not related to obesity were used as negative control outcomes.

RESULTS

A total of 1,140,300 patients potentially eligible for bariatric surgery were included in the study. Among these patients, 288,600 (25.3%) patients underwent bariatric surgery. A total of 48,411 cancers were identified, including 4,483 in surgical patients and 43,928 in patients not receiving bariatric surgery. Bariatric surgery is associated with a decrease in the risk of obesity-related cancer 0.89 (95% CI, 0.83-0.95), while no significant effect of surgery was identified with regard to cancers not related to obesity (risk ratio was 0.96 (95% CI, 0.91-1.01)).

CONCLUSIONS

When emulating a target trial from observational data, our results suggest a reduction of 11% for obesity-related cancer after bariatric surgery. The absence of an association observed for the negative control outcome suggests that selection bias could be controlled.

INTRODUCTION

Obesity is a rapidly growing global health issue [101]. In 2016, the World Health Organization estimated that 1.9 billion adults were overweight, and of these, 650 million were affected by obesity, which corresponds to 13% of the adult population [20].

It has been demonstrated that obesity increases the risk of overall mortality for cardiovascular, respiratory and liver diseases [102]. Moreover, the causal association between obesity and cancer is nowadays strongly documented, in particular for hormone-related cancers (breast, corpus uteri and prostate), colorectal cancer (CRC), oesophageal cancer and gastric cancer [81]. Several mechanisms are advocated to explain the effect of obesity on cancer development, in particular insulin resistance, chronic inflammation, and alterations in the metabolism of adipokines and sexual hormones [103].

Obesity (bariatric) surgery is currently considered as the most effective treatment for sustained weight loss and remission of obesity-related conditions [104]. Several studies have also shown a reduction in the incidence of cancer after bariatric surgery [105], [106].

Nevertheless, most of these studies have methodological limitations. The control of bias, in particular indication bias, represents a major issue, as patients who undergo bariatric surgery may not be comparable to patients who do not receive bariatric surgery. In addition, small sample sizes result in limited power to examine the differences in risk by cancer type. As a result, existing studies show conflicting results (in particular for CRC) or counterintuitive effects with regard to the reduction of risks for cancers not related to obesity after bariatric surgery [99], [105], [107], [108].

The gold standard for evaluating the causal effects of an intervention is a randomized controlled trial, with the unique design allowing for identification of causal effects without bias, since randomization provides groups that differ only with regard to the intervention. Emulating a target randomized trial from observational databases allows, under specific assumptions, to identify the causal effects of an intervention [109]. In this study, we estimated the effect of bariatric surgery on cancer incidence using a nationwide routinely collected database, applying methods that help to minimize bias in the analyses of observational data through consideration of the target trial.

METHODS

STUDY DESIGN AND SETTING

Data were extracted from a national discharge database, the “Programme De Médicalisation des Systèmes d’Information” (PMSI), which collects all discharge reports from all hospitals in France, irrespective of facility ownership or academic affiliation. Discharge reports are mandatory and represent the basis for hospital funding. For this reason, the database is comprehensive for all reimbursed surgical interventions in the country.

We defined the target trial designed to compare a potential candidate for bariatric surgery who indeed underwent bariatric surgery and those who did not. Three assumptions are needed to conceptualize an observational study as a randomized experiment: consistency (defined as the proper definition of treatment under comparison, corresponding to the version of treatment in the data), exchangeability (the conditional probability of receiving every value of treatment, depending only on the measured covariates) and positivity (the conditional probability of receiving every value of treatment is greater than zero) [77].

PARTICIPANTS

In order to minimize potential bias and ensure comparability (exchangeability and positivity assumptions), we applied patient selection criteria that would be used in a target trial. Eligible patients were 18-65 years old, admitted with a diagnosis of obesity between 2010 and 2017, with a BMI ≥ 40 kg/m² or with a BMI 35-40 kg/m² associated with an obesity-related comorbidity (hypertension, type 2 diabetes, dyslipidaemia, or

obstructive sleep apnoea syndrome). The identification of obesity and comorbidities was made on the basis of the International Classification of Disease, 10th edition (ICD-10). Patients with a history of alcohol abuse and psychoactive substance use were excluded, as addictions represent a contraindication to bariatric surgery. We also excluded patients with a history of cancer at baseline and patients who were diagnosed as having cancer in the 2 years after inclusion in order to avoid detection bias (i.e., misdiagnosis of cancer in the preoperative assessment for bariatric surgery).

Patients were then divided in two groups: patients who underwent bariatric surgery and patients who did not. Patients in the surgical group were identified through the medical acts classification (“Classification Commune des Actes Médicaux”, CCAM, 11th edition), which is a national standardized classification of surgical, endoscopic and interventional procedures [110]. The unambiguous identification of the intervention permits researchers to assure the consistency assumption.

The inception time of the bariatric cohort was the hospital stay for bariatric surgery. The inception time for the control group was the date of admission when a diagnosis of obesity firstly appeared, irrespective of the cause of admission. All included patients were followed from the first diagnosis of obesity until cancer appearance, death or censoring in December 2019.

OUTCOMES

The main outcome was cancer incidence, for any type of cancer and for the subgroups of obesity-related cancers. Obesity-related cancers included 14 types of cancer (oesophageal adenocarcinoma, postmenopausal breast cancer and cancers of the kidney, colon, rectum, gastric cardia, liver, gallbladder, pancreas, ovary, corpus uteri, thyroid and meningioma, as well as multiple myeloma) having sufficient evidence for being obesity-related as reported by the International Agency for Research on Cancer (IARC) Working Group [81].

In order to verify whether the relation among bariatric surgery and cancer risk was correctly inferred, we employed the risk of cancer not related to obesity as a “negative control” outcome [111]. Negative controls are used in epidemiological studies to detect confounding, selection and measurement biases [80]. The formal approach is the repetition of the experiment under conditions in which it is expected to produce a null result. In our case, in patients affected by cancers that are not related to obesity, intentional weight loss is not expected to provide any effect. Hence, a null effect on negative controls would suggest that unmeasured confounders have been ruled out.

Incident cancers were identified in the database through the ICD-10 codes, which is in general based on the topography/anatomical sector rather than on the histology of the tumour. Hence, for instance, we can identify an oesophageal cancer but the information about the histological type is not available. For the identification of postmenopausal breast cancer, we adopted the definition of cancers diagnosed at age ≥ 55 as previously reported in studies using electronic data sources [108], [112].

In the PMSI database, each patient is assigned a unique identifier, which remains unchanged over time. Thus, linkage between different hospital stays in different hospitals is possible. Since the identifier is anonymous, patient consent was not required. Access to the database was submitted for authorization by the National Commission on Informatics and Liberty (CNIL, authorization n° 01947391).

STATISTICAL METHODS

WEIGHTING

Using a propensity score approach, we applied inverse probability of treatment weighting (IPTW) to balance baseline characteristics between exposure groups. The propensity score was derived using a logistic regression model, employing the following pre-treatment variables: patient’s age, gender, body mass index (BMI), time of follow-up, comorbidities and tobacco consumption. Each observation was

weighted by the inverse of the probability of a patient receiving bariatric surgery, given observed confounders identified on the index date. The IPTW approach creates a pseudo-population in which the exposure group (bariatric surgery versus no bariatric surgery) was independent of measured confounders [113].

In order to avoid the impact of extreme weights on estimations, weights were truncated at a maximum threshold of 15. The balance among the covariates of patients was assessed using standardized mean difference (SMD); a difference of ≤ 0.10 was considered a well-balanced result [114].

MULTIVARIABLE

In order to compare the risk of cancer in patients receiving or not receiving bariatric surgery, we used a Cox proportional hazard regression model adjusted for IPTW. The model was calculated to assess the risk of any cancer, of obesity-related cancer and of cancer not related to obesity. We also assessed the risk for type-specific cancers presenting at least 50 cases. As the weighting does not allow for perfectly similar groups, the adjustment makes it possible to take into account the residual differences between the two cohorts.

SENSITIVITY ANALYSIS

We conducted several sensitivity analyses to assess the robustness of our findings. First, we excluded patients who underwent adjustable gastric banding from this analysis, as this procedure has become anecdotal among bariatric procedures [22]. Secondly, we applied an alternative 1:2 propensity score-matched design, consisting of associating one patient receiving bariatric surgery with two patients not receiving bariatric surgery based on the propensity score.

STATISTICAL PROGRAM

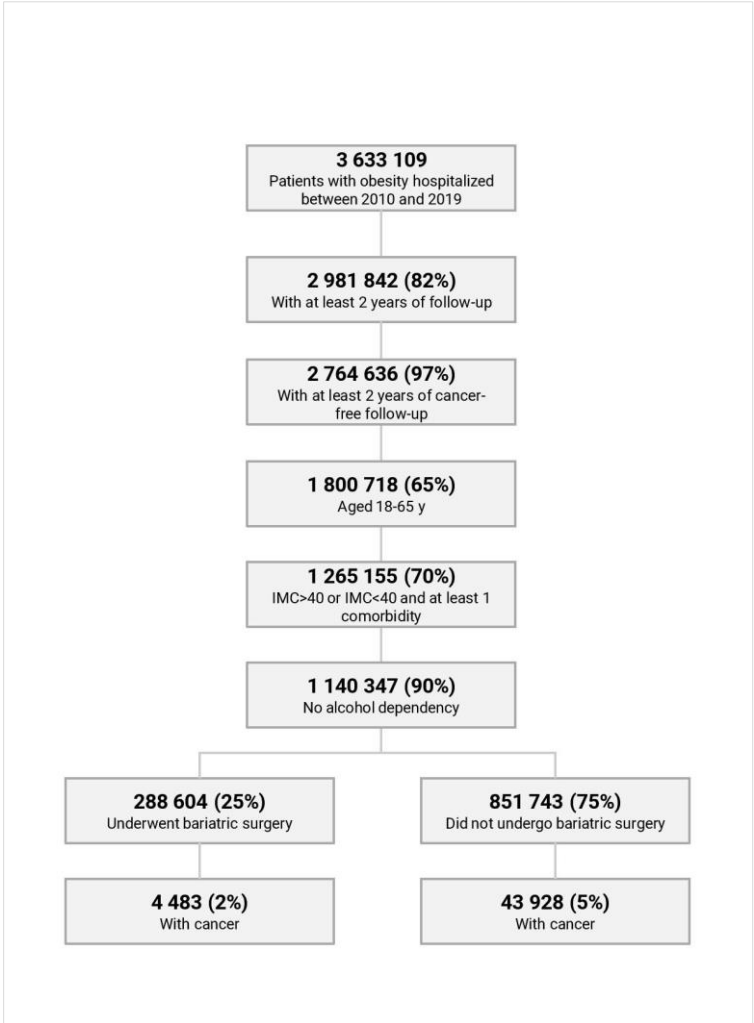
All the statistical analyses were performed with R software (R Foundation for Statistical Computing, Vienna, Austria).[115] Data are reported according to the reporting of studies conducted using the observational routinely-collected health data (RECORD) statement [46].

RESULTS

PARTICIPANTS

Between 2010 and 2019, 3,633,019 patients were admitted with a diagnosis of obesity in France. Among these, 2,764,636 (76.1%) had at least 2 years of follow-up free from cancer. We then selected patients aged 18-65 years, with a BMI > 40 kg/m² or a BMI 35-40 kg/m² and at least one obesity-related comorbidity. At the end, 1,140,347 patients were eligible for the study. Among the eligible patients, 288,604 (25.3%) underwent bariatric surgery and 851,743 (74.7%) did not undergo bariatric surgery and were included in the control group (Figure 1).

Figure 1. Attrition diagram.



DESCRIPTIVE DATA

The baseline characteristics of patients are reported in Table 1. There were several important differences between the bariatric group and the control group, in particular for gender (81.6% female versus 55.2% female), age (mean 39.8 ± 11.5 years versus 51.8 ± 11.1 years), and BMI (> 40 kg/m², 58.3% versus 26.6%). Patients who underwent bariatric surgery also had fewer comorbidities, except for peptic ulcer disease (3.6% versus 2.1%) and mild liver disease (9.3% versus 7.4%). Mean follow-up time was 5.7 ± 2.2 years in the bariatric group and 6.5 ± 2.3 years in the control group.

Table 1. Baseline characteristics of patients.

	Bariatric group n = 288604	Control group n = 851743	SMD
Sex (men)	53194 (18.4)	381997 (44.8)	0.592
Age, mean (SD)	39.82 (11.53)	51.77 (11.09)	1.056
Age (years)			1.068
18-30	71733 (24.9)	55320 (6.5)	
30-40	80712 (28.0)	86238 (10.1)	
40-50	76502 (26.5)	169699 (19.9)	
50-60	49077 (17.0)	322799 (37.9)	
60-65	10580 (3.7)	217687 (25.6)	
BMI (kg/m ²)			0.685
< 40	119326 (41.3)	624852 (73.4)	
40-50	145426 (50.4)	198481 (23.3)	
> 50	23852 (8.3)	28410 (3.3)	
Tobacco use	32746 (11.3)	144309 (16.9)	0.161
Sleep apnoea syndrome	84775 (29.4)	233643 (27.4)	0.043
Hypertension	75829 (26.3)	560604 (65.8)	0.864
Hypercholesterolaemia	37832 (13.1)	321642 (37.8)	0.59
Myocardial infarction	2688 (0.9)	65101 (7.6)	0.336
Congestive heart failure	5087 (1.8)	118034 (13.9)	0.463
Peripheral vascular disease	3042 (1.1)	79779 (9.4)	0.381
Cerebrovascular disease	4615 (1.6)	72963 (8.6)	0.321
Dementia	179 (0.1)	5540 (0.7)	0.099
Chronic pulmonary disease	25269 (8.8)	110736 (13.0)	0.137
Rheumatic disease	2006 (0.7)	18702 (2.2)	0.126
Peptic ulcer disease	10313 (3.6)	18101 (2.1)	0.087
Mild liver disease	26956 (9.3)	62813 (7.4)	0.071
Moderate or severe liver disease	657 (0.2)	5569 (0.7)	0.064
Diabetes without chronic complication	39282 (13.6)	337340 (39.6)	0.615
Diabetes with chronic complication	7578 (2.6)	142017 (16.7)	0.49
Hemiplegia or paraplegia	2906 (1.0)	31678 (3.7)	0.179
Renal disease	2720 (0.9)	58233 (6.8)	0.308
HIV	494 (0.2)	2925 (0.3)	0.034

OUTCOME DATA

In total, 48,411 incident cancers were identified (Table 2). In the bariatric group, 4,483 (1.6%) patients developed a cancer, while in the control group, a cancer was observed in 43,928 (5.2%) patients. The four most common cancers were colorectal (n = 4,763; 9.8%), postmenopausal breast (n = 4,468; 9.2%), prostate (n = 4,376; 9.0%) and lung cancer (n = 3,730; 7.7%).

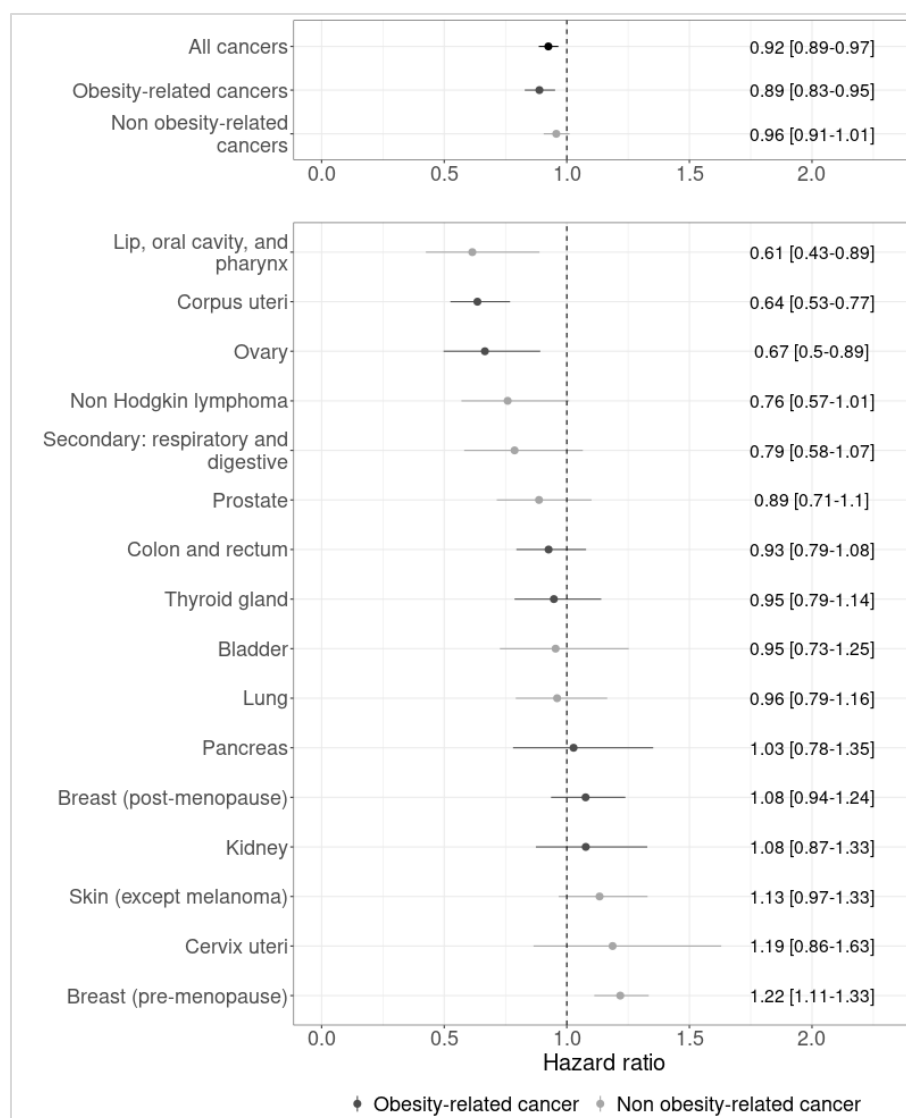
Table 2. Types of cancers.

Type of cancer	TOTAL n=48411	Bariatric group n=4483	Control group n=43928
Colon and rectum	4763 (9.84)	329 (7.34)	4434 (10.09)
Breast (postmenopausal)	4468 (9.23)	331 (7.38)	4137 (9.42)
Prostate	4376 (9.04)	119 (2.65)	4257 (9.69)
Lung	3730 (7.7)	183 (4.08)	3547 (8.07)
Breast (premenopausal)	3605 (7.45)	1029 (22.95)	2576 (5.86)
Skin (except melanoma)	3466 (7.16)	314 (7)	3152 (7.18)
Bladder	2478 (5.12)	104 (2.32)	2374 (5.4)
Kidney	2451 (5.06)	210 (4.68)	2241 (5.1)
Corpus uteri	1895 (3.91)	203 (4.53)	1692 (3.85)
Thyroid gland	1517 (3.13)	296 (6.6)	1221 (2.78)
Pancreas	1426 (2.95)	98 (2.19)	1328 (3.02)
Secondary: respiratory and digestive	1301 (2.69)	86 (1.92)	1215 (2.77)
Non-Hodgkin lymphoma	1287 (2.66)	100 (2.23)	1187 (2.7)
Secondary: other	1037 (2.14)	79 (1.76)	958 (2.18)
Liver and intrahepatic bile ducts	954 (1.97)	40 (0.89)	914 (2.08)
Lip, oral cavity and pharynx	854 (1.76)	50 (1.12)	804 (1.83)
Skin: cutaneous melanoma	683 (1.41)	84 (1.87)	599 (1.36)
Central nervous system (except meningioma)	635 (1.31)	58 (1.29)	577 (1.31)
Ovary	599 (1.24)	87 (1.94)	512 (1.17)
Gastric, noncardia	575 (1.19)	31 (0.69)	544 (1.24)
Multiple myeloma	555 (1.15)	33 (0.74)	522 (1.19)
Lymphoid leukaemia	512 (1.06)	42 (0.94)	470 (1.07)
Secondary: lymph nodes	483 (1)	49 (1.09)	434 (0.99)
Cervix uteri	478 (0.99)	108 (2.41)	370 (0.84)
Oesophagus	457 (0.94)	22 (0.49)	435 (0.99)
Other cancers	431 (0.89)	35 (0.78)	396 (0.9)
Myeloid leukaemia	361 (0.75)	45 (1)	316 (0.72)
Larynx	305 (0.63)	14 (0.31)	291 (0.66)
Connective and soft tissue	255 (0.53)	40 (0.89)	215 (0.49)
Small intestine	220 (0.45)	19 (0.42)	201 (0.46)
Gastric cardia	218 (0.45)	15 (0.33)	203 (0.46)
Hodgkin lymphoma	214 (0.44)	32 (0.71)	182 (0.41)
Extrahepatic biliary tract	194 (0.4)	10 (0.22)	184 (0.42)
Myeloid leukaemia	190 (0.39)	13 (0.29)	177 (0.4)
Immunoproliferative diseases	134 (0.28)	11 (0.25)	123 (0.28)
Vulva	132 (0.27)	20 (0.45)	112 (0.25)
Bone and articular cartilage	119 (0.25)	26 (0.58)	93 (0.21)
Eye and adnexa	112 (0.23)	18 (0.4)	94 (0.21)
Testis	101 (0.21)	20 (0.45)	81 (0.18)
Adrenal gland	96 (0.2)	13 (0.29)	83 (0.19)
Leukaemia, other	90 (0.19)	11 (0.25)	79 (0.18)
Retroperitoneum and peritoneum	83 (0.17)	11 (0.25)	72 (0.16)
Meningioma	81 (0.17)	7 (0.16)	74 (0.17)
Gallbladder	77 (0.16)	2 (0.04)	75 (0.17)
Other endocrine glands	76 (0.16)	5 (0.11)	71 (0.16)
Uterus, part unspecified	75 (0.15)	10 (0.22)	65 (0.15)
Breast (male)	68 (0.14)	1 (0.02)	67 (0.15)
Lymphoid, haematopoietic and related tissues	67 (0.14)	8 (0.18)	59 (0.13)
Heart, mediastinum and pleura	64 (0.13)	8 (0.18)	56 (0.13)
Mesothelioma	63 (0.13)	4 (0.09)	59 (0.13)

MAIN RESULTS

The pseudo-population emulated by IPTW included 2,067,384 patients, including 897,674 (43%) patients who were operated on and 1,169,710 (57%) patients who were not operated on (see supplementary data). The average of the SMD decreased from 0.35 to 0.09 after weighting. In the pseudo-population, the causal hazard risks (HRs) associated with bariatric surgery were 0.92 (95% CI, 0.88-0.97) for all cancers (Figure 2), 0.89 (95% CI, 0.83-0.95) for obesity-related cancers and 0.96 (95% CI, 0.91-1.01) for cancers not related to obesity.

Figure 2. Forest plot of multivariable Cox proportional-hazards models for cancers on the weighted population. The box represents the hazard ratio and the error bars depict the 95% confidence interval.



OTHER ANALYSES

In the first sensitivity analysis, we modified the study design, performing a 1:2 propensity score matching. We obtained two similar populations with a mean SMD of 0.06. The patient characteristics of the matched population are available in the supplementary data. In the matched population, the HR was 0.87 (95% CI, 0.84-0.91) for all cancers, 0.77 (95% CI, 0.72-0.82) for obesity-related cancers and 0.93 (95% CI, 0.89-0.97) for cancers not related to obesity. In the subgroup of patients receiving sleeve gastrectomy or gastric bypass, the HR for obesity-related cancers decreased to 0.86 (95% CI, 0.80-0.93).

DISCUSSION

We applied a national discharge database to the framework for comparative effectiveness based on the emulation of a randomized control trial, in order to estimate the causal effects of bariatric surgery on the incidence of cancer in patients with obesity. The size of our population was unprecedented, as we included more than 280,000 bariatric surgery patients who developed 4,400 cancers, which is considerably higher than the greatest series published up to now [108], [116]. As a result, we assessed a reduction of 11% for obesity-related cancers in patients that received bariatric surgery compared to those who did not receive bariatric surgery. These results strengthen the evidence of the beneficial effect of weight loss for the prevention of cancer. Moreover, we did not find any association with cancers not related to obesity, which represented our negative control outcome, suggesting that confounders were properly controlled.

Our results are consistent with a few previous studies on the impact of bariatric surgery on the risk of cancer. In 2009, Adams et al. published a 12-year follow-up study on more than 6,000 bariatric surgery patients compared to an obese control group and found a 24% reduction in overall cancer incidence and a 38% reduction in obesity-related cancers [105]. This study used a Cox regression model that was adjusted for three covariates (age, sex and BMI).

The same year, the Swedish Obese Subjects (SOS) Study reported on the incidence of cancer in a cohort of 2,000 patients matched with obese controls. In this case, matching was done using 18 clinical variables [106]. After a mean follow-up exceeding 10 years, bariatric surgery was associated with a 33% reduction of cancer incidence.

More recently, Schauer et al. (2018), in a cohort of more than 22,000 bariatric surgery patients compared to 66,000 obese individuals, confirmed the beneficial impact of obesity surgery on cancer incidence for any cancer (HR 0.67, 95% CI 0.60-0.74, $p < 0.001$) and for obesity-related cancers (HR 0.65, 95% CI 0.51-0.69, $p < 0.001$) [98]. In this case, the mean follow-up was shorter (3.5 years), and the two matched populations were compared through a Cox proportional hazard model. This group also surprisingly found an effect of bariatric surgery on cancer not associated with obesity and confirmed this finding in premenopausal breast cancer (not related to obesity) in another publication [117].

In 2020, a study from Bailly et al., using the same data source as our study (the PMSI database), reported a benefit of bariatric surgery on the risk of CRC (HR 0.68, 95% CI 0.60-0.77) [118]. Although other studies have reported on the beneficial impact of surgery on CRC risk, this type of cancer has been a source of controversy on the impact of bariatric surgery, as a few studies from Sweden and the UK have reported an increased risk of CRC in a post-bariatric surgery population [107], [116], [119]–[121]. It is interesting to observe that all the studies that found an increased risk of CRC in a post-bariatric surgery population did not directly compare a bariatric surgery population with an equivalent obese population, and they used the method of standardized incidence ratio (SIR), comparing bariatric surgery patients to the general population. On the contrary, studies in favour of bariatric surgery used different methods of matching, directly comparing obese patients receiving bariatric surgery versus obese patients not receiving bariatric surgery.

This methodological difference highlights the importance of an adequate control group, which by definition, can only be provided by randomization. As a randomized controlled trial is not always feasible for ethical, logistical or financial reasons, epidemiologists have developed methods to emulate this type of trial using observational data. If the emulation is successful, the analysis for observational data could yield the same effect estimates as a target trial, responding to the initial causal question [109].

The inadequacy of controlling for confounders may yield contradictory or counterintuitive results, where bariatric surgery may increase the risk of obesity-related cancers (as in the case of CRC) or decrease the risk of cancers that are not related to obesity.

LIMITATIONS

The main limitations of our study relate to its observational nature, as unmeasured confounding may persist between the bariatric group and the control group. Nevertheless, the specification of a target trial permitted to build our analysis in order to minimize bias.

We included only patients that were potentially eligible for bariatric surgery (exchangeability assumption). With this method, some patients that actually received bariatric surgery did not meet the inclusion criteria for our study and were excluded. This highlights the fact that the national discharge database does not capture all the clinical details leading to the indication for bariatric surgery. However, the same criteria were also applied to the control group in order to improve comparability.

Admission bias, also known as Berkson bias, could be a concern in the interpretation of our results. All patients, including those in the control group, were identified during hospital admissions. Hospital cases might have higher risk exposures or disease than cases from the population at large [122]. For this reason, we excluded patients with a history of cancer in the 2 years after inclusion.

Another limitation is the assessment of the history of smoking through the ICD-10 codes, which is probably underestimated as the prevalence of smoking in the French population is estimated around 30% [123]. However, it is unlikely that this represents confusion bias, because the prevalence of smoking was similar in the two groups in the weighted and in the matched populations.

CONCLUSION

In this study emulating a randomized controlled trial on a nationwide scale, the occurrence of obesity-related cancer decreased after bariatric surgery, with an 11% reduction magnitude. The absence of the effect of bariatric surgery on cancer not related to obesity could reflect control of the bias inherent in observational data. However, since age is a major risk factor for numerous cancers, a database including longer follow-up would be useful for identifying a larger effect.

SUPPLEMENTARY DATA

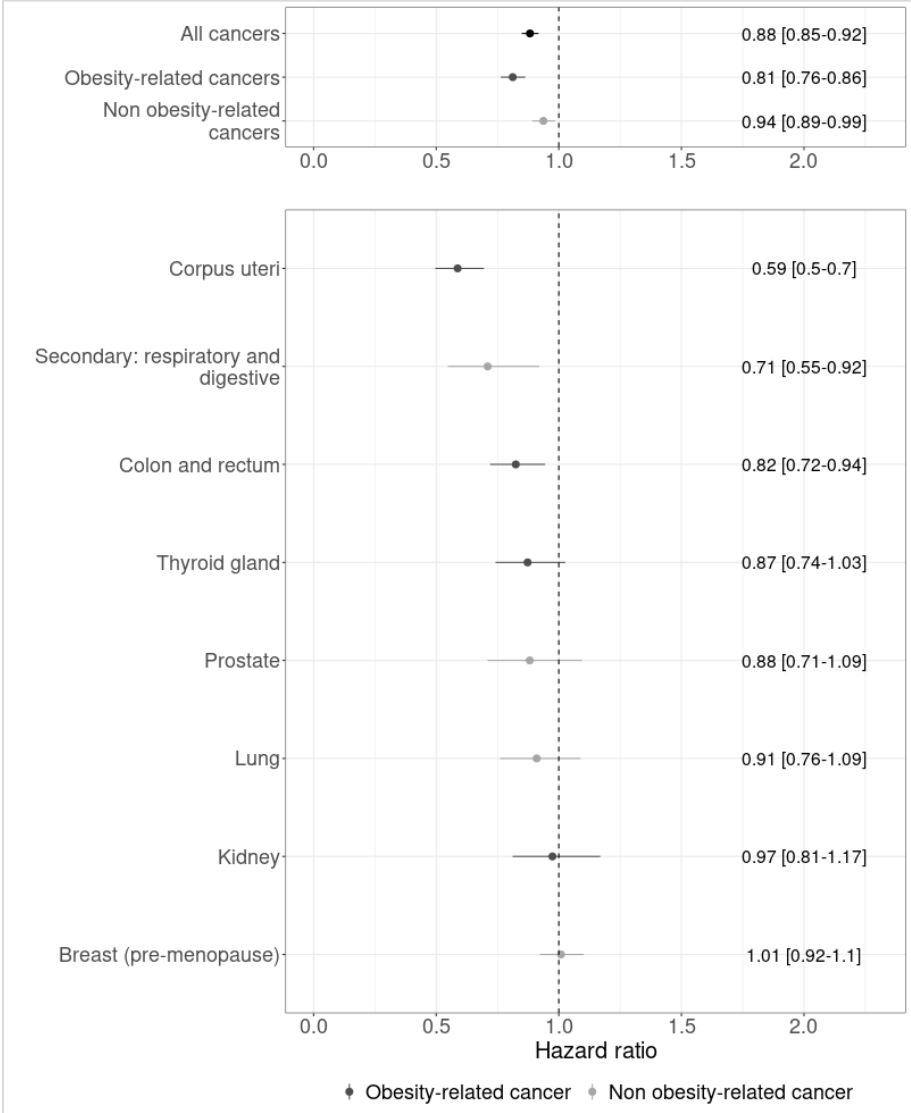
Annexe 1. Baseline characteristics of patients in pseudo-population in IPTW analysis.

	Bariatric group n = 897 674	Control group n = 1 169 710	SMD
Sex (men)	270254.8 (30.1)	434774.7 (37.2)	0.592
Age, mean (SD)	45.61 (11.54)	48.00 (13.17)	1.056
Age (years)			1.068
18-30	114369.9 (12.7)	169552.6 (14.5)	
30-40	174606.3 (19.5)	163793.9 (14.0)	
40-50	258289.5 (28.8)	228470.4 (19.5)	
50-60	269483.1 (30.0)	371619.8 (31.8)	
60-65	80925.2 (9.0)	236274.3 (20.2)	
BMI (kg/m ²)			0.685
< 40	461111.8 (51.4)	732794.9 (62.6)	
40-50	377368.8 (42.0)	377114.9 (32.2)	
> 50	59193.5 (6.6)	59801.2 (5.1)	
Tobacco	122146.9 (13.6)	179260.8 (15.3)	0.161
Sleep apnoea syndrome	295974.0 (33.0)	327577.7 (28.0)	0.043
Hypertension	425851.7 (47.4)	633230.5 (54.1)	0.864
Hypercholesterolaemia	227958.9 (25.4)	359146.2 (30.7)	0.59
Myocardial infarction	28502.8 (3.2)	67841.6 (5.8)	0.336
Congestive heart failure	53761.9 (6.0)	123334.8 (10.5)	0.463
Peripheral vascular disease	31915.3 (3.6)	82878.9 (7.1)	0.381
Cerebrovascular disease	38022.1 (4.2)	77758.5 (6.6)	0.321
Dementia	1864.1 (0.2)	5722.8 (0.5)	0.099
Chronic pulmonary disease	99823.6 (11.1)	137716.6 (11.8)	0.137
Rheumatic disease	15209.8 (1.7)	20794.3 (1.8)	0.126
Peptic ulcer disease	23924.4 (2.7)	28143.8 (2.4)	0.087
Mild liver disease	82096.4 (9.1)	93817.6 (8.0)	0.071
Moderate or severe liver disease	4006.0 (0.4)	6282.5 (0.5)	0.064
Diabetes without chronic complication	264861.3 (29.5)	378451.4 (32.4)	0.615
Diabetes with chronic complication	78250.1 (8.7)	149995.1 (12.8)	0.49
Hemiplegia or paraplegia	19299.1 (2.1)	34757.8 (3.0)	0.179
Renal disease	27109.2 (3.0)	61082.8 (5.2)	0.308
HIV	2499.9 (0.3)	3438.9 (0.3)	0.034

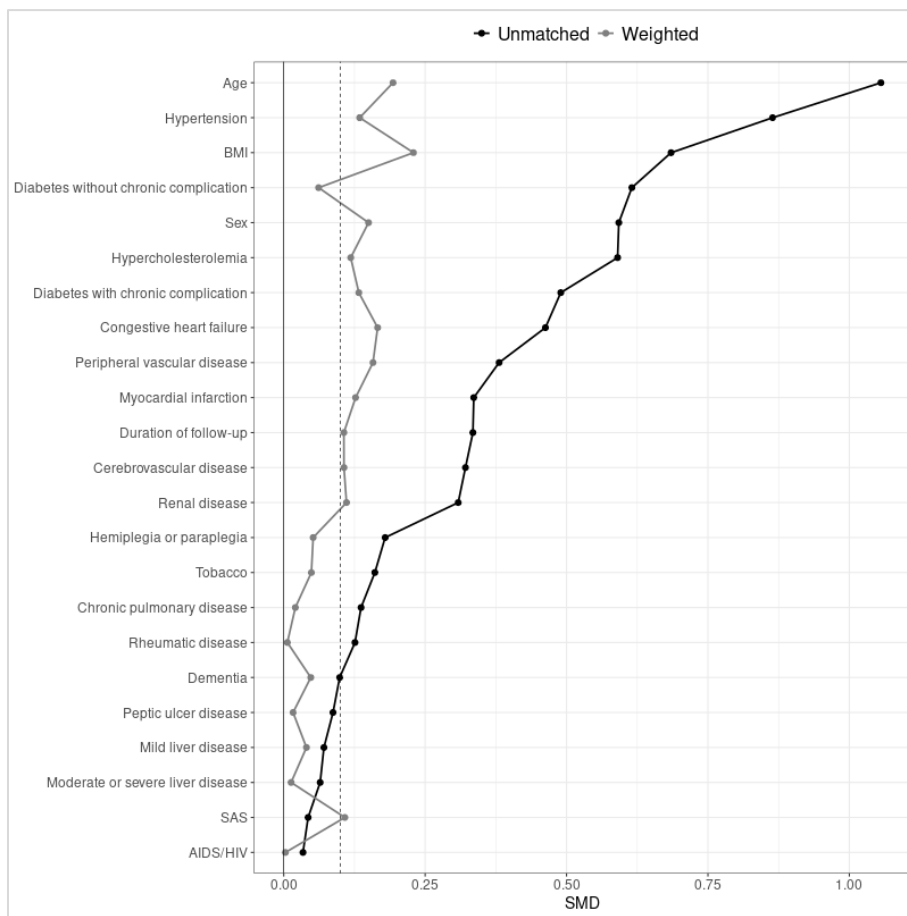
Annexe 2. Baseline characteristics of patients in the matched population.

	Bariatric group n = 212686	Control group n = 298663	SMD
Sex (men)	47539 (22.4)	73387 (24.6)	0.052
Age, mean (SD)	41.94 (11.71)	44.01 (11.77)	0.176
Age (years)			0.167
18-30	42764 (20.1)	47117 (15.8)	
30-40	51898 (24.4)	66710 (22.3)	
40-50	60402 (28.4)	83757 (28.0)	
50-60	47190 (22.2)	81153 (27.2)	
60-65	10432 (4.9)	19926 (6.7)	
BMI			0.192
< 40	78778 (37.0)	137953 (46.2)	
40-50	114124 (53.7)	140339 (47.0)	
> 50	19784 (9.3)	20371 (6.8)	
Tobacco	25812 (12.1)	37754 (12.6)	0.015
Sleep apnoea syndrome	69968 (32.9)	98378 (32.9)	0.001
Hypertension	72340 (34.0)	124365 (41.6)	0.158
Hypercholesterolaemia	35717 (16.8)	60599 (20.3)	0.09
Myocardial infarction	2669 (1.3)	5207 (1.7)	0.04
Congestive heart failure	5049 (2.4)	10085 (3.4)	0.06
Peripheral vascular disease	2997 (1.4)	5772 (1.9)	0.041
Cerebrovascular disease	4512 (2.1)	8608 (2.9)	0.049
Dementia	176 (0.1)	340 (0.1)	0.01
Chronic pulmonary disease	22449 (10.6)	30972 (10.4)	0.006
Rheumatic disease	1995 (0.9)	4122 (1.4)	0.041
Peptic ulcer disease	6249 (2.9)	7440 (2.5)	0.028
Mild liver disease	22692 (10.7)	27185 (9.1)	0.053
Moderate or severe liver disease	38627 (18.2)	71840 (24.1)	0.145
Diabetes without chronic complication	609 (0.3)	1043 (0.3)	0.011
Diabetes with chronic complication	7531 (3.5)	15811 (5.3)	0.085
Hemiplegia or paraplegia	2799 (1.3)	4874 (1.6)	0.026
Renal disease	2696 (1.3)	5471 (1.8)	0.046
HIV	477 (0.2)	837 (0.3)	0.011

Annexe 3. Forest plot of multivariable Cox proportional-hazards models for cancers on the matched population. The box represents the hazard ratio and the error bars depict the 95% confidence interval.



Annexe 4. SMD distribution in the matched and unmatched population.



ETUDE N°2

COMPARAISON DE DEUX STRATEGIES CHIRURGICALES DE CHIRURGIE BARIATRIQUE DE REVISION ETUDE EN COURS (CENTRE DE RECHERCHE CLINIQUE DU CHI DE CRETEIL – DR SEYDOU GORO)

RESUME

La chirurgie de révision représente environ 14% des actes de chirurgie bariatrique. L'anneau gastrique ajustable, qui a été très pratiqué en France pendant une vingtaine d'années (on estime à 170.000 le nombre d'anneaux posés) est particulièrement grand pourvoyeur de chirurgie de deuxième intention [124].

En effet cette intervention peut soit donner un résultat modeste en terme de perte de poids soit provoquer une complication invalidante et, dans les deux cas, une ablation de l'anneau peut être envisagée. La plupart des patients restent dans la perspective de poursuivre leur traitement contre leur obésité avec un nouveau traitement chirurgical, en particulier avec d'autres techniques telles que la *sleeve gastrectomy* ou le *bypass gastrique* [124].

La transition entre anneau gastrique et nouvelle technique (*sleeve* ou *bypass*) peut se faire avec un timing différent : ou on enlève l'anneau et on réalise la deuxième procédure dans le même temps chirurgical (stratégie en un temps), ou on réalise deux interventions distinctes : d'abord l'ablation de l'anneau et dans un deuxième temps, la nouvelle intervention (stratégie en deux temps).

Il n'existe aucune étude randomisée sur la comparaison des deux stratégies en terme de *safety* et *efficacy* et les études observationnelles donnent des résultats contradictoires [125]–[129].

Le choix entre une stratégie ou une autre appartient au jugement des équipes chirurgicales. Celles-ci montrent généralement une grande cohérence dans leur méthode, certaines pratiquant exclusivement une stratégie en un temps et d'autres exclusivement en deux temps, et ce quel que soient les facteurs de risques des patients. Nous pouvons donc considérer le choix des équipes comme un phénomène de randomisation naturelle, et l'utiliser comme variable instrumentale.

L'idée du projet sera donc de comparer les *outcomes* chirurgicaux de morbi-mortalité classiques (selon la classification de Clavien-Dindo) entre les centres qui adoptent de manière quasi-exclusive une stratégie plutôt qu'une autre.

ETUDE N°3

SCORE DE CLAVIEN-DINDO SUR PMSI : ELABORATION D'UN ALGORITHME DE DETECTION

ETUDE EN COURS (DOCTORAT - DR CHALLINE)

RESUME

L'objectif de ce premier travail de développement d'indicateurs de qualité en chirurgie était le développement d'un algorithme de classification de Clavien-Dindo à partir des données d'hospitalisation du PMSI. La validation de l'algorithme a été faite en utilisant les données cliniques issues de 3 essais randomisés français, financés par le Programme Hospitalier de recherche Clinique (PHRC : MIRO, GRECCAR 5 et 6). Afin de préserver l'anonymat des patients, les données des bases (PMSI vs PHRC) ont été appariées sur 6 variables (âge, sexe, durée d'hospitalisation, année, mois et centre d'intervention). Deux algorithmes ont été calculés : un algorithme « clinique », construit à priori, sur des variables cliniquement pertinentes en utilisant les codes CCAM et les codes CIM-10 ; un algorithme « *data-driven* » à partir d'un modèle de prédiction, construit par modélisation.

L'appariement a permis de retrouver 85% patients des essais cliniques dans le PMSI. La méthode clinique sur la base de construction a permis de classer correctement les complications sévères de 91.5% patients, avec une sensibilité de 76% et une spécificité de 91%. Par la méthode *data-driven*, le taux d'erreur était 0.08 avec une sensibilité de 90% et une spécificité de 91%.

Une limite importante de ce travail réside dans l'incapacité d'apparier la totalité des patients entre les deux sources de données (PMSI vs PHRC). Ces algorithmes pourraient être donc optimisés par l'apport de plus de patients. Des démarches sont actuellement en cours pour intégrer des patients traités par chirurgie pancréatique dans l'essai PREFIPS.

Au final, l'algorithme « clinique » de détection du score de Clavien-Dindo sur le PMSI permet d'identifier une complication sévère des grades 3-4-5 correctement dans 91% des cas. L'algorithme basé sur des méthodes *data-driven* de création de variables semble avoir une performance très proche, voire légèrement supérieure.

INTRODUCTION

Le développement et la validation d'indicateurs de qualité, de pertinence et de sécurité des soins issus des systèmes d'information de routine sont considérés par les instances de santé comme une priorité pour les financements des établissements de santé basés sur la qualité. Actuellement peu d'indicateurs sont développés à partir des données médico-administratives en particulier en chirurgie digestive, alors qu'en pratique clinique on dispose davantage d'outils pour évaluer la qualité et la sécurité de soins. A partir des pratiques cliniques courantes, nous souhaitons développer des indicateurs de qualité des soins en chirurgie.

L'objectif de ce premier travail de développement d'indicateurs de qualité en chirurgie est le développement d'un algorithme de classification de Clavien-Dindo à partir des données d'hospitalisation du PMSI. La validation de l'algorithme a été faite en utilisant les données cliniques issues de 3 essais randomisés français, financés par le Programme Hospitalier de recherche Clinique : les essais MIRO, GRECCAR 5 et GRECCAR 6.

MATERIEL ET METHODES

Il s'agit d'une étude de validation comparant les données de la base PMSI à une source de données qui a été prise comme référentiel. Les données de référence sont issues des trois essais randomisés contrôles multicentriques, financés par le dispositif PHRC.

Ces essais ont concerné des patients opérés de cancer de l'œsophage (essai « MIRO ») et du rectum (essai « GRECCAR 5 » et « GRECCAR 6 »).

Les données administratives à valider ont été extraites de la base des hospitalisations PMSI-MCO. Les données issues des PHRC ont été appariées aux données de la base PMSI de manière complètement anonyme sur la base de six variables : âge, sexe, durée d'hospitalisation, année, mois et établissement de santé.

L'objectif principal de l'étude était de construire un algorithme d'identification des complications sévères. Celles-ci étaient définies par la classification de Clavien-Dindo. Cette classification concerne toute complication chirurgicale et peut s'appliquer à toute spécialité chirurgicale. Elle comporte 5 niveaux de sévérité, dont les deux premiers sont considérés comme des complications mineures et les 3 derniers comme des complications sévères. Cette classification a l'avantage majeur de ne pas être construite sur un diagnostic, mais sur le traitement nécessaire pour prendre en charge la complication. Le niveau 3, par exemple, implique une réintervention chirurgicale ou endoscopique, le niveau 4 une hospitalisation en réanimation.

L'objectif secondaire est de construire un algorithme qui permette de classer les complications selon les 5 niveaux de la classification de Clavien-Dindo.

Tableau 1. Classification des complications chirurgicales de Clavien-Dindo

	Définition	Exemples
Grade I	Tout évènement postopératoire indésirable ne nécessitant pas de traitement médical, chirurgical, endoscopique ou radiologique. Les seuls traitements autorisés sont les antiémétiques, antipyrétiques, antalgiques, diurétiques, électrolytes et la physiothérapie.	Iléus, abcès de paroi mis à plat au chevet du patient
Grade II	Complication nécessitant un traitement médical n'étant pas autorisé dans le grade 1.	Thrombose veineuse périphérique, nutrition parentérale totale, transfusion
Grade III	Complication nécessitant un traitement chirurgical, endoscopique ou radiologique.	
IIIa	Sans anesthésie générale	Ponction guidée radiologiquement
IIIb	Sous anesthésie générale	Reprise chirurgicale pour saignement ou autre cause
Grade IV	Complication engageant le pronostic vital et nécessitant des soins intensifs	
IVa	Défaillance d'un organe	Dialyse
IVb	Défaillance multi-viscérale	
Grade V	Décès	
Suffixe d	Complication en cours au moment de la sortie du patient nécessitant un suivi ultérieur (d = discharge)	

Comme variables explicatives nous avons considéré les informations concernant les diagnostics, par la Classification internationale des maladies (CIM) et les actes thérapeutiques et diagnostiques, via la Classification commune des actes médicaux (CCAM). En particulier les comorbidité des patients ont été prise en compte selon le score de comorbidité de Charlson [70].

ANALYSE STATISTIQUE

Pour la construction de l'algorithme d'identification, nous avons utilisé deux approches :

1. Une approche *data-driven* à partir d'un modèle de prédiction, construit par modélisation.
2. Une approche « clinique », construite à priori, sur des variables cliniquement pertinentes en utilisant les codes CCAM et les codes CIM-10.

Dans l'approche *data-driven*, les variables explicatives, en grand nombre, ont été initialement réduites par une analyse de correspondances multiples (ACM). Une régression Lasso a aussi été réalisée dans le même but, à titre de confirmation.

La prédiction des complications sévères a été réalisée par une régression logistique avec sélection pas à pas (*stepwise*) des variables. Une base d'apprentissage a été constituée avec 70% des individus, et une base dédiée à la prédiction avec 30% des individus. Les modèles ont été comparés par le critère d'information d'Akaike (AIC) et par une courbe ROC (AUC). A titre exploratoire, on a aussi réalisé une modélisation par arbre de décision et par forêt aléatoire. La prédiction de la Classification de Clavien-Dindo a été réalisée avec un modèle polytomique ordonné et cumulatif.

RESULTATS

Tout d'abord, l'appariement a permis d'identifier 85% patients des essais cliniques dans le PMSI. Ainsi 15% des patients, ne retrouvant pas de correspondance, ont été exclus. Finalement l'analyse a porté sur 667 patients, répartis en deux groupes: une base d'apprentissage de 467 individus (70% de l'échantillon) et une base dédiée à la prédiction de 200 individus (30% de l'échantillon).

Les individus ont été tirés aléatoirement dans l'échantillon, avec pour contrainte de créer des bases qui conservent globalement la proportion d'individus ayant eu des complications. Au total 61 (9,1%) complications sévères ont été identifiées.

L'ACM, utilisée dans l'approche *data-driven*, a permis d'identifier 6 variables (3 codes CIM-10 et 3 codes CCAM) pertinentes pour la prédiction des complications sévères. Les variables retenues après la sélection pas à pas (*stepwise*) sont décrites dans le tableau 2.

Tableau 2. Codes utilisés dans la construction des algorithmes.

Code	Classification	Libellé
T81	CIM-10	Complications d'actes à visée diagnostique et thérapeutique, non classées ailleurs.
K65	CIM-10	Péritonite
K63	CIM-10	Autres maladies de l'intestin
ZCQH001	CCAM	Scanographie de l'abdomen et du petit bassin [pelvis], avec injection intraveineuse de produit de contraste
YYYY467	CCAM	Supplément pour injection intraveineuse de produit de contraste au cours d'un examen radiographique ou scanographique.
ZBQH001	CCAM	Scanographie du thorax, avec injection intraveineuse de produit de contraste.

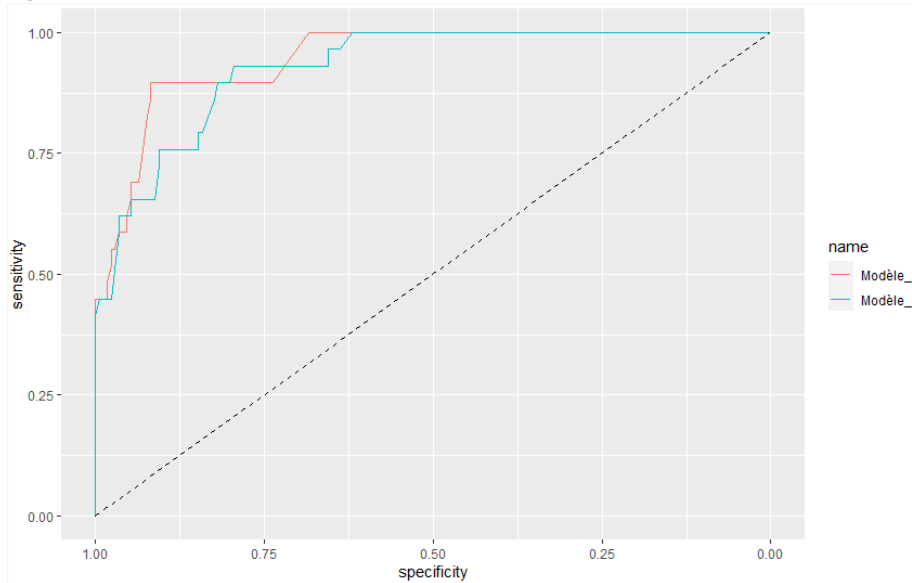
A partir de ces variables nous avons réalisés deux régressions logistiques avec et sans inclusion du score de Charlson (modèle 1 et 3). Dans l'approche « clinique », nous avons aussi réalisé deux modèles, sans score de Charlson (modèle 2) et avec score de Charlson (modèle 4). Le tableau 3 montre la comparaison entre les quatre modèles.

Tableau 3. Comparaison des différents modèles

	Data Driven	Clinique	Data Driven	Clinique
	Modèle 1 – sans Charlson	Modèle 2 – sans Charlson	Modèle 3 –avec Charlson	Modèle 4 –avec Charlson
AIC	0.923	0.913	0.947	0.929
AUC	258	241	197	199
Taux de bien classés	0.89	0.895	0.92	0.915
Seuil choisi	0.15	0.15	0.2	0.28
Sensibilité	0.83	0.86	0.9	0.76
Spécificité	0.87	0.8	0.91	0.91

Ce tableau montre que les modèles 3 et 4 surpassent les deux premiers modèles, ce qui confirme l'importance d'avoir introduit la variable « score de comorbidité de Charlson ». Le modèle 3 possède les meilleures caractéristiques (meilleure AUC, AIC la plus faible, meilleur taux de bien classés le plus élevé, sensibilité et spécificité les plus élevées au point optimal de la courbe ROC) et nous laisse donc penser qu'il s'agit ici du meilleur modèle de prédiction parmi les 4 proposés (figure 1).

Figure 1. Courbe ROC des modèles 3 et 4.



En comparaison avec ces modèles, le pouvoir de prédiction de l'arbre de décision et de la forêt aléatoire est nettement inférieur à celui des modèles *logit*.

Enfin, pour modéliser le score de Clavien-Dindo, on a testé un modèle polytomique ordonné et cumulatif. Ce modèle ne possède qu'une faculté de prédiction limitée sur le score Clavien-Dindo. En effet, le taux de bonne prédiction est seulement de 0.5.

DISCUSSION

Parmi tous les modèles nous paraissant pertinents que nous avons réalisés, nous obtenons finalement un modèle très performant sur la prédiction des complications. En effet, les modèles avec de tels taux prédictions sont rares en médecine. Nous sommes satisfaits de la performance des modèles car ils permettent d'avoir une spécificité et une sensibilité relativement élevées.

Sur le thème de la chirurgie pour cancer de l'œsophage et du rectum, nous avons construit des algorithmes qui identifient assez bien les complications, notamment les complications sévères. Néanmoins dans ces algorithmes, nous retrouvons des codes CIM-10 et CCAM qui sont étroitement liés à la pathologie d'analyse, par exemple le code K65 (péritonite) ou le code ZBQH001 (scanner thoracique). En conséquent, cela pose la question de la généralisation de ces algorithmes à la chirurgie d'autres secteurs anatomiques, qui présentent des complications complètement différentes.

En effet, l'intérêt (et en fin de compte, le coup de génie) de la classification de Clavien-Dindo est qu'elle n'est pas basée sur la pathologie, ni sur la région anatomique. Cette classification se base sur l'action nécessaire pour répondre au problème : administration de médicaments, réintervention, soins intensifs etc. Cela permet donc une généralisation à toute discipline chirurgicale.

L'enjeu de nos recherches sera donc d'aller au-delà des algorithmes spécifiques à une pathologie et rechercher un algorithme plus généralisable.

Parmi les limites de ce travail, il faut mentionner l'incapacité d'apparier la totalité des patients entre les deux sources de données (PMSI et PHRC). Ces algorithmes pourraient être donc optimisés par l'apport d'un plus grand nombre de patients. Des démarches sont actuellement en cours pour intégrer des patients traités par chirurgie pancréatique dans l'essai PREFIPS. En effet, tout algorithme doit être validé sur un

référentiel considéré comme fiable. L'identification et l'accessibilité à un référentiel de qualité sera un enjeu majeur dans la validation des travaux sur les BDMA.

CONCLUSION

Au final, l'algorithme *data-driven* permet d'identifier une complication sévère dans 92% des cas chez des patients opérés pour cancer de l'œsophage ou du rectum. L'algorithme basé sur l'approche « clinique » de création de variables a aussi une performance très satisfaisante.

PERSPECTIVES DE RECHERCHE

L'inférence causale à partir des données observationnelles est un champ d'application de l'épidémiologie encore peu exploré.

Comme dans l'axe I, l'accessibilité à des données massives ouvre les possibilités d'utilisation de cette méthode, en particulier dans les cas où un essai randomisé ne serait pas faisable pour des raisons économiques, éthiques ou de temps.

Compte tenu du volume de données existantes, un grand nombre de questions cliniques qui ne pourraient pas être traitées par un essai randomisés, pourraient faire l'objet d'une analyse.

Le cancer et l'obstétrique sont deux thématiques qui, en effet, se prêtent bien à cet exercice.

Pour la suite de nos recherches, nous avons des projets de recherche sur les sujets suivants :

1. **Cancer œsogastrique et chirurgie bariatrique.** Ce projet, mené avec le CHU de Bordeaux, est une étude complémentaire de l'étude sur l'incidence de cancer après chirurgie bariatrique. Nous nous focalisons sur le cancer de l'œsophage et de l'estomac et nous allons étudier l'incidence de ces cancers après chirurgie de l'obésité, en stratifiant par technique chirurgicale (anneau gastrique, sleeve gastrectomy et bypass gastrique). Le sujet est d'extrême actualité car l'hypothèse a été émise que la chirurgie bariatrique serait associée à une augmentation du cancer de l'œsophage, via le reflux gastro-œsophagien et au cancer de l'estomac, via la modification du circuit alimentaire. Or, très peu des données existent sur le sujet et aucune cohorte nationale n'a été publiée.
2. **Obésité et complications materno-fœtales.** L'obésité est un facteur de risque connue de complication de la grossesse, de l'accouchement et de malformation du fœtus. La perte de poids induite par la chirurgie pourrait amener à une réduction de ces complications. Ce projet, mené avec le CHU Antoine Beclère (APHP), a pour but de constituer une cohorte de patients en état d'obésité, opérée de chirurgie bariatrique et non opérés, et de comparer les effets de la perte de poids post chirurgie sur les résultats obstétricaux.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] M. Prodel, "Process discovery , analysis and simulation of clinical pathways using health-care data," Université de Lyon, 2017.
- [2] C. Pascal, J.-B. Capgras, N. Guilhot, and J.-P. Claveranne, "L'information médico-économique entre décision stratégique et action publique : tout savoir pour peu pouvoir !," *Rev. Polit. Manag. Public*, vol. 28, no. 1, pp. 57–82, 2011, doi: 10.3166/pmp.28.57-82.
- [3] A. Lazzati, G. Chatellier, and S. Katsahian, "Readmissions After Bariatric Surgery in France, 2013–2016: a Nationwide Study on Administrative Data," *Obes. Surg.*, vol. 29, no. 11, 2019, doi: 10.1007/s11695-019-04053-6.
- [4] A. Lazzati, G. Chatellier, L. Paolino, S. Batahei, and S. Katsahian, "Postoperative care fragmentation in bariatric surgery and risk of mortality: a nationwide study," *Surg. Obes. Relat. Dis.*, pp. 1–7, 2021, doi: 10.1016/j.soard.2021.03.004.
- [5] N. Le Meur, F. Gao, and S. Bayat, "Mining care trajectories using health administrative information systems: The use of state sequence analysis to assess disparities in prenatal care consumption," *BMC Health Serv. Res.*, vol. 15, no. 1, Dec. 2015, doi: 10.1186/S12913-015-0857-5.
- [6] V. Vogt, S. M. Scholz, and L. Sundmacher, "Applying sequence clustering techniques to explore practice-based ambulatory care pathways in insurance claims data," *Eur. J. Public Health*, vol. 28, no. 2, pp. 214–219, Apr. 2018, doi: 10.1093/EURPUB/CKX169.
- [7] B. B. F. Škopljanač-Mačina, "Formal Concept Analysis – overview and applications," *Procedia Eng*, vol. 69, pp. 1258–1267, 2014, doi: 10.1016/j.proeng.2014.03.117.
- [8] D. I. Ignatov, "Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Fields," *Commun. Comput. Inf. Sci.*, vol. 505, pp. 42–141, Mar. 2017, doi: 10.1007/978-3-319-25485-2_3.
- [9] A. Napoli, "An Introduction to Formal Concept Analysis," in *Tutorial at ICCS 2020 An Introduction to Formal Concept Analysis (FCA) Bolzano Summer of Knowledge, September 2020*, 2020, no. September.
- [10] R. Wille, "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts," *Ordered Sets*, pp. 445–470, 1982, doi: 10.1007/978-94-009-7798-3_15.
- [11] A. Abbott and J. Forrest, "Optimal Matching Methods for Historical Sequences," *J. Interdiscip. Hist.*, vol. 16, no. 3, p. 471, 1986, doi: 10.2307/204500.
- [12] M. Studer, G. Ritschard, A. Gabadinho, and N. S. Müller, "Discrepancy Analysis of State Sequences:," <http://dx.doi.org/10.1177/0049124111415372>, vol. 40, no. 3, pp. 471–510, Aug. 2011, doi: 10.1177/0049124111415372.
- [13] G. Pollock, "Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis," *J. R. Stat. Soc. Ser. A (Statistics Soc.)*, vol. 170, no. 1, pp. 167–183, Jan. 2007, doi: 10.1111/J.1467-985X.2006.00450.X.
- [14] M. Studer, E. Struffolino, and A. E. Fasang, "Estimating the Relationship between Time-varying Covariates and Trajectories: The Sequence Analysis Multistate Model Procedure:," <https://doi.org/10.1177/0081175017747122>, vol. 48, no. 1, pp. 103–135, Jan. 2018, doi: 10.1177/0081175017747122.
- [15] H. MACINDOE and A. ABBOTT, "Sequence Analysis and Optimal Matching Techniques for Social Science Data," in *Handbook of Data Analysis*, SAGE Publications, Ltd, 2012, pp. 386–406.
- [16] L. Fonteneau *et al.*, "Apport des bases médico-administratives en épidémiologie et santé publique des maladies infectieuses The use of administrative health databases in infectious disease epidemiology and public health," 2017, doi: 10.1016/j.respe.2017.03.131.

- [17] J. Roux, "Parcours de soins des patients atteints de sclérose en plaques à partir des données médico-administratives en France," 2018.
- [18] N. Robette, "L'étude des parcours de vie, démarche descriptive ou causale ?," *Idées économiques Soc.*, vol. N° 177, no. 3, p. 8, 2014, doi: 10.3917/idee.177.0008.
- [19] J. Larmarange, "Analyse de séquences," *analyse-R*. <https://larmarange.github.io/analyse-R/> (accessed Nov. 01, 2021).
- [20] World Health Organization, "Obesity and overweight," 2020. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> (accessed Nov. 25, 2020).
- [21] S. K. KD Hall, "Maintenance of lost weight and long-term management of obesity," *Med Clin North Am*, vol. 102, no. 1, pp. 183–197, Jan. 2018, doi: 10.1016/j.mcna.2017.08.012.
- [22] L. Angrisani *et al.*, "IFSO Worldwide Survey 2016: Primary, Endoluminal, and Revisional Procedures," *Obes. Surg.*, vol. 28, no. 12, pp. 3783–3794, 2018, doi: 10.1007/s11695-018-3450-2.
- [23] J. D. Birkmeyer *et al.*, "Surgical Skill and Complication Rates after Bariatric Surgery," *N. Engl. J. Med.*, vol. 369, no. 15, pp. 1434–1442, 2013, doi: 10.1056/NEJMsa1300625.
- [24] S. B. W. K. A. W. P. B. DR Flum, "Perioperative safety in the longitudinal assessment of bariatric surgery," *N Engl J Med*, vol. 361, no. 5, pp. 445–454, Jul. 2009, doi: 10.1056/nejmoa0901836.
- [25] D. B. D. D. C. S. WE Encinosa, "Recent improvements in bariatric surgery outcomes," *Med Care*, vol. 47, no. 5, pp. 531–535, May 2009, doi: 10.1097/mlr.0b013e31819434c6.
- [26] W. I. M. G. A. P. JP Regan, "Early experience with two-stage laparoscopic Roux-en-Y gastric bypass as an alternative in the super-super obese patient," *Obes Surg*, vol. 13, no. 6, pp. 861–864, Dec. 2003, doi: 10.1381/096089203322618669.
- [27] J. D. Birkmeyer *et al.*, "Surgical Skill and Complication Rates after Bariatric Surgery for the Michigan Bariatric Surgery Collaborative," *N Engl J Med*, vol. 369, pp. 1434–42, 2013, doi: 10.1056/NEJMsa1300625.
- [28] H. Buchwald, R. Estok, K. Fahrbach, D. Banel, and I. Sledge, "Trends in mortality in bariatric surgery: a systematic review and meta-analysis," *Surgery*, vol. 142, no. 4, pp. 621–32; discussion 632-5, Oct. 2007, doi: 10.1016/j.surg.2007.07.018.
- [29] Z. M.-G. A. C.-B. A. M.-C. F. C.-G. J. P.-G. A. García-Ruiz-de-Gordejuela, "Evaluation of bariatric surgery patients at the emergency department of a tertiary referral hospital," *Rev Esp. Enfermedades Dig Organo Soc Esp. Patol Dig*, vol. 107, no. 1, pp. 23–28, 2015.
- [30] A. G. Doumouras, F. Saleh, and D. Hong, "30-Day readmission after bariatric surgery in a publicly funded regionalized center of excellence system," *Surg. Endosc. Other Interv. Tech.*, vol. 30, no. 5, pp. 2066–2072, 2016, doi: 10.1007/s00464-015-4455-8.
- [31] L. M. A. A. I. A. E. L. S. S. RJ Rosenthal, "Readmission rates of bariatric procedures," *J Am Coll Surg*, vol. 221, no. 4, pp. e47–e48, Oct. 2015, doi: 10.1016/j.jamcollsurg.2015.08.020.
- [32] J. Saunders *et al.*, "One-year readmission rates at a high volume bariatric surgery center: Laparoscopic adjustable gastric banding, laparoscopic gastric bypass, and vertical banded gastroplasty-Roux-en-Y gastric bypass," *Obes. Surg.*, vol. 18, no. 10, pp. 1233–1240, 2008, doi: 10.1007/s11695-008-9517-8.
- [33] N. Jay, G. Nuemi, M. Gadreau, and C. Quantin, "A data mining approach for grouping and analyzing trajectories of care using claim data: The example of breast cancer," *BMC Med. Inform. Decis. Mak.*, vol. 13, no. 1, 2013, doi: 10.1186/1472-6947-13-130.
- [34] A. S. I. H.-K. A. S. A. H. S. F. U Fennig, "Pre-surgical weight loss predicts post-surgical weight loss trajectories in adolescents enrolled in a bariatric program," *Obes Surg*, vol. 29, no. 4, pp. 1154–1163,

Apr. 2019, doi: 10.1007/s11695-018-03649-8.

- [35] Y. H. P. B. A. P. G. W. C. S. MR Lent, "Demographic, clinical, and behavioral determinants of 7-year weight change trajectories in Roux-en-Y gastric bypass patients," *Surg Obes Relat Dis Off J Am Soc Bariatr Surg*, vol. 14, no. 11, pp. 1680–1685, Nov. 2018, doi: 10.1016/j.soard.2018.07.023.
- [36] P.-B. A, de L. M, B. I, M. PPP, and C. EM, "Weight loss trajectories and psychobehavioral predictors of outcome of primary and reoperative bariatric surgery: a 2-year longitudinal study," *Surg. Obes. Relat. Dis.*, vol. 15, no. 7, pp. 1104–1112, Jul. 2019, doi: 10.1016/J.SOARD.2019.04.018.
- [37] E. R. Berger *et al.*, "Prevalence and Risk Factors for Bariatric Surgery Readmissions," *Ann. Surg.*, vol. 267, no. 1, pp. 122–131, 2018, doi: 10.1097/SLA.0000000000002079.
- [38] G. Bruze, J. Ottosson, M. Neovius, I. Näslund, and R. Marsk, "Hospital admission after gastric bypass: a nationwide cohort study with up to 6 years follow-up," *Surg. Obes. Relat. Dis.*, vol. 13, no. 6, pp. 962–969, 2017, doi: 10.1016/j.soard.2017.01.004.
- [39] D. A. Telem *et al.*, "Rates and risk factors for unplanned emergency department utilization and hospital readmission following bariatric surgery," *Ann. Surg.*, vol. 263, no. 5, pp. 956–960, 2016, doi: 10.1097/SLA.0000000000001536.
- [40] Y. S. J. F. R. A. J. T. RB Fetter, "Case mix definition by diagnosis-related groups," *Med Care*, vol. 18, no. 2 Suppl, p. iii, 1980.
- [41] B. Ristevski and M. Chen, "Big Data Analytics in Medicine and Healthcare," *J. Integr. Bioinform.*, vol. 15, no. 3, May 2018, doi: 10.1515/JIB-2017-0030.
- [42] G. Moulis, M. Lapeyre-Mestre, A. Palmaro, G. Pugnet, J. L. Montastruc, and L. Sailler, "French health insurance databases: What interest for medical research?," *Revue de Medecine Interne*, vol. 36, no. 6, pp. 411–417, 2015, doi: 10.1016/j.revmed.2014.11.009.
- [43] J. K. M. C. P. P. D. E. A. R. SA Brethauer, "Standardized outcomes reporting in metabolic and bariatric surgery," *Surg Obes Relat Dis Off J Am Soc Bariatr Surg*, vol. 11, no. 3, pp. 489–506, May 2015, doi: 10.1016/j.soard.2015.02.003.
- [44] H. Quan *et al.*, "Updating and validating the charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries," *Am. J. Epidemiol.*, vol. 173, no. 6, pp. 676–682, 2011, doi: 10.1093/aje/kwq433.
- [45] A. Lazzati, G. Chatellier, and S. Katsahian, "Readmissions After Bariatric Surgery in France, 2013–2016: a Nationwide Study on Administrative Data," *Obes. Surg.*, vol. 29, no. 11, pp. 3680–3689, 2019, doi: 10.1007/s11695-019-04053-6.
- [46] E. I. Benchimol *et al.*, "The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement," *PLoS Med.*, vol. 12, no. 10, 2015, doi: 10.1371/journal.pmed.1001885.
- [47] J. A. J. B. E Lounkine, "Formal concept analysis for the identification of molecular fragment combinations specific for active and highly potent compounds," *J Med Chem*, vol. 51, no. 17, pp. 5342–5348, Sep. 2008, doi: 10.1021/jm800515r.
- [48] S. M. U. F. C. F. R. S. J Gebert, "Identifying genes of gene regulatory networks using formal concept analysis," *J Comput Biol*, vol. 15, no. 2, pp. 185–194, Mar. 2008, doi: 10.1089/cmb.2007.0107.
- [49] G. M. Z. P. D. P. L. Y. F Hao, "\$K\\$\\$-clique community detection in social networks based on formal concept analysis," *IEEE Syst J*, vol. 11, no. 1, pp. 250–259, Mar. 2017, doi: 10.1109/jsyst.2015.2433294.
- [50] P. Elzinga, J. Poelmans, S. Viaene, G. Dedene, and S. Morsing, "Terrorist threat assessment with formal concept analysis," *ISI 2010 - 2010 IEEE Int. Conf. Intell. Secur. Informatics Public Saf. Secur.*, pp. 77–82, 2010, doi: 10.1109/ISI.2010.5484773.

- [51] C. Ciangura, D. Nocca, and V. Lindecker, “[Guidelines for clinical practice for bariatric surgery].,” *Presse Med.*, vol. 39, no. 9, pp. 953–9, Sep. 2010, doi: 10.1016/j.lpm.2010.03.025.
- [52] K. M. G. C. J. M. W. W. MJ Stampfer, “Risk of symptomatic gallstones in women with severe obesity,” *Am J Clin Nutr*, vol. 55, no. 3, pp. 652–658, 1992, doi: 10.1093/ajcn/55.3.652.
- [53] G. P. M. R. L. S. J. M. H Yang, “Risk factors for gallstone formation during rapid loss of weight,” *Dig Dis Sci*, vol. 37, no. 6, pp. 912–918, Jun. 1992, doi: 10.1007/bf01300390.
- [54] D. E. Magouliotis *et al.*, “Ursodeoxycholic Acid in the Prevention of Gallstone Formation After Bariatric Surgery: an Updated Systematic Review and Meta-analysis,” *Obes. Surg.*, vol. 27, no. 11, pp. 3021–3030, 2017, doi: 10.1007/s11695-017-2924-y.
- [55] L. G. M. C. F. L. CS Stokes, “Ursodeoxycholic acid and diets higher in fat prevent gallbladder stones during weight loss: a meta-analysis of randomized controlled trials,” *Clin Gastroenterol Hepatol Off Clin Pr. J Am Gastroenterol Assoc*, vol. 12, no. 7, pp. 1090-1100.e2, 2014.
- [56] G. Friedman, “Natural history of asymptomatic and symptomatic gallstones,” *Am J Surg*, vol. 165, no. 4, pp. 399–404, 1993, doi: 10.1016/s0002-9610(05)80930-4.
- [57] E. Lo Menzo, M. Hinojosa, A. Carbonell, D. Krpata, J. Carter, and A. M. Rogers, “American Society for Metabolic and Bariatric Surgery and American Hernia Society consensus guideline on bariatric surgery and hernia surgery,” *Surg. Obes. Relat. Dis.*, vol. 14, no. 9, pp. 1221–1232, 2018, doi: 10.1016/j.soard.2018.07.005.
- [58] B. L. C. C. K. F. P. G. P. H. J. J. V. S. H Quan, “Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries,” *Am J Epidemiol*, vol. 173, no. 6, pp. 676–682, Mar. 2011, doi: 10.1093/aje/kwq433.
- [59] G. Rey, S. Rican, and E. Jouglu, “Mesure des inégalités de mortalité par cause de décès . Approche écologique à l ’ aide d ’ un indice de désavantage social,” *Distribution*, pp. 87–90, 2011.
- [60] S. S. STURROCK, “Time Warps, String Edits, and Macromolecules – The Theory and Practice of Sequence Comparison. David Sankoff and Joseph Kruskal. ISBN 1-57586-217-4. Price £13.95 (US\$22.95).,” *Genet. Res. (Camb).*, vol. 76, no. 3, pp. 327–329, Dec. 2000, doi: 10.1017/S0016672300219320.
- [61] A. Abbott and A. Hrycak, “Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians’ Careers,” <https://doi.org/10.1086/229495>, vol. 96, no. 1, pp. 144–185, Oct. 2015, doi: 10.1086/229495.
- [62] A. ABBOTT and A. TSAY, “Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect,” <http://dx.doi.org/10.1177/0049124100029001001>, vol. 29, no. 1, pp. 3–33, Jun. 2016, doi: 10.1177/0049124100029001001.
- [63] K. Dlouhy and T. Biemann, “Optimal matching analysis in career research: A review and some best-practice recommendations,” *J. Vocat. Behav.*, vol. 90, pp. 163–173, Oct. 2015, doi: 10.1016/J.JVB.2015.04.005.
- [64] N. SB and W. CD, “A general method applicable to the search for similarities in the amino acid sequence of two proteins,” *J. Mol. Biol.*, vol. 48, no. 3, pp. 443–453, Mar. 1970, doi: 10.1016/0022-2836(70)90057-4.
- [65] A. Lazzati *et al.*, “Natural history of adjustable gastric banding: Lifespan and revisional rate,” *Ann. Surg.*, vol. 265, no. 3, 2017, doi: 10.1097/SLA.0000000000001879.
- [66] B. Clapp, M. Wynn, C. Martyn, C. Foster, M. O’Dell, and A. Tyroch, “Long term (7 or more years) outcomes of the sleeve gastrectomy: a meta-analysis,” *Surg. Obes. Relat. Dis.*, vol. 14, no. 6, pp. 741–747, 2018, doi: 10.1016/j.soard.2018.02.027.
- [67] A. Lazzati, S. Bechet, S. Jouma, L. Paolino, and C. Jung, “Revision surgery after sleeve gastrectomy: a

- nationwide study with 10 years of follow-up," *Surg. Obes. Relat. Dis.*, vol. 16, no. 10, pp. 1497–1504, 2020, doi: 10.1016/j.soard.2020.05.021.
- [68] A.-M. Bouvier *et al.*, "Incidence of gastrointestinal cancers in France," *Gastroenterol. Clin. Biol.*, vol. 28, no. 10 Pt 1, pp. 877–81, Oct. 2004, doi: 10.1016/s0399-8320(04)95152-4.
- [69] "Chirurgie des cholangiocarcinomes intrahépatiques, hilaires et vésiculaires - Rapport présenté au 111ème Congrès Français de Chirurgie 2009 - J.-M. REGIMBEAU, F.-R. PRUVOT, O. FARGES, collaboration de D. FUKS - 9782718412115 - Livre - Unitheque.com." <https://www.unitheque.com/chirurgie-des-cholangiocarcinomes-intrahepatiques-hilaires-vesiculaires/monographies-de-l-association-francaise-de-chirurgie/arnette/Livre/32846> (accessed Oct. 20, 2021).
- [70] A. Bannay *et al.*, "The Best Use of the Charlson Comorbidity Index With Electronic Health Care Database to Predict Mortality.," *Med. Care*, vol. 54, no. 2, pp. 188–94, 2016, doi: 10.1097/MLR.0000000000000471.
- [71] INCa, "Algorithme de sélection des hospitalisations liées à la prise en charge du cancer dans les bases nationales d'activité hospitalière de court séjour «algorithme cancer», 2013. Accessed: Oct. 20, 2021. [Online]. Available: <https://lesdonnees.e-cancer.fr/Informations/Methodes/Methodologie-algorithme-cancer2>.
- [72] A. Gabadinho, G. Ritschard, N. S. Müller, and M. Studer, "Analyzing and Visualizing State Sequences in R with TraMineR," *J. Stat. Softw.*, vol. 40, no. 4, pp. 1–37, Apr. 2011, doi: 10.18637/JSS.V040.I04.
- [73] C. Lepage *et al.*, "Trends in the incidence and management of biliary tract cancer: A French population-based study," *J. Hepatol.*, vol. 54, no. 2, pp. 306–310, Feb. 2011, doi: 10.1016/j.jhep.2010.06.039.
- [74] C. Neuzillet, C. Emery, C. Tessier, S. Bouée, and A. Lièvre, "Épidémiologie et parcours de soins des cholangiocarcinomes intrahépatiques (iCCA) en France : données de vie réelle issues du Programme de médicalisation des systèmes d'information," *Rev. Epidemiol. Sante Publique*, vol. 68, p. S77, Sep. 2020, doi: 10.1016/j.respe.2020.04.029.
- [75] L. Blendis and Z. Halpern, "An increasing incidence of cholangiocarcinoma: why?," *Gastroenterology*, vol. 127, no. 3, pp. 1008–9, Sep. 2004, doi: 10.1053/j.gastro.2004.07.035.
- [76] J. N. Primrose *et al.*, "Capecitabine compared with observation in resected biliary tract cancer (BILCAP): a randomised, controlled, multicentre, phase 3 study," *Lancet Oncol.*, vol. 20, no. 5, pp. 663–673, May 2019, doi: 10.1016/S1470-2045(18)30915-X.
- [77] J. M. Robins and M. A. Hernán, "Causal Inference - what if," *Found. Agnostic Stat.*, pp. 235–281, 2020.
- [78] R. J. Glynn, E. L. Knight, R. Levin, and J. Avorn, "Paradoxical relations of drug treatment with mortality in older persons," *Epidemiology*, vol. 12, no. 6, pp. 682–689, 2001, doi: 10.1097/00001648-200111000-00017.
- [79] S. Schneeweiss *et al.*, "Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results," *Med. Care*, vol. 45, no. 10 SUPPL. 2, 2007, doi: 10.1097/mlr.0b013e318070c08e.
- [80] B. F. Arnold and A. Ercumen, "Negative control outcomes: A tool to detect bias in randomized trials," *JAMA - J. Am. Med. Assoc.*, vol. 316, no. 24, pp. 2597–2598, 2016, doi: 10.1001/jama.2016.17700.
- [81] B. Lauby-Secretan, C. Scoccianti, D. Loomis, Y. Grosse, F. Bianchini, and K. Straif, "Body Fatness and Cancer — Viewpoint of the IARC Working Group," *N. Engl. J. Med.*, vol. 375, no. 8, pp. 794–798, 2016, doi: 10.1056/nejmsr1606602.
- [82] M. L. Petersen, K. E. Porter, S. Gruber, Y. Wang, and M. J. Van Der Laan, "Diagnosing and responding to violations in the positivity assumption," *Stat. Methods Med. Res.*, vol. 21, no. 1, pp. 31–54, Feb. 2012, doi: 10.1177/0962280210386207.

- [83] S. Greenland, "An introduction to instrumental variables for epidemiologists," *Int. J. Epidemiol.*, vol. 29, pp. 722–729, 2000.
- [84] K. L. Sainani, "Instrumental Variables: Uses and Limitations," *PM R*, vol. 10, no. 3, pp. 303–308, 2018, doi: 10.1016/j.pmrj.2018.02.002.
- [85] S. L. Au Yeung *et al.*, "Moderate Alcohol Use and Cardiovascular Disease from Mendelian Randomization," *PLoS One*, vol. 8, no. 7, Jul. 2013, doi: 10.1371/JOURNAL.PONE.0068054.
- [86] Y. Xian *et al.*, "Association between stroke center hospitalization for acute ischemic stroke and mortality," *JAMA - J. Am. Med. Assoc.*, vol. 305, no. 4, pp. 373–380, Jan. 2011, doi: 10.1001/JAMA.2011.22.
- [87] R. Momosaki, H. Yasunaga, H. Matsui, H. Horiguchi, K. Fushimi, and M. Abo, "Effect of early rehabilitation by physical therapists on in-hospital mortality after aspiration pneumonia in the elderly," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 2, pp. 205–209, Feb. 2015, doi: 10.1016/J.APMR.2014.09.014.
- [88] K. D. Shetty, W. B. Vogt, and J. Bhattacharya, "Hormone replacement therapy and cardiovascular health in the United States," *Med. Care*, vol. 47, no. 5, pp. 600–606, May 2009, doi: 10.1097/MLR.0b013e31818bfe9b.
- [89] K. W. Saunders *et al.*, "Relationship of opioid use and dosage levels to fractures in older chronic pain patients," *J. Gen. Intern. Med.*, vol. 25, no. 4, pp. 310–315, Apr. 2010, doi: 10.1007/s11606-009-1218-z.
- [90] L. A. G. Rodríguez and A. Ruigómez, "Case validation in research using large databases," *British Journal of General Practice*, vol. 60, no. 572, pp. 160–161, Mar. 2010, doi: 10.3399/bjgp10X483472.
- [91] D. G. Ives *et al.*, "Surveillance and ascertainment of cardiovascular events. The Cardiovascular Health Study," *Ann. Epidemiol.*, vol. 5, no. 4, pp. 278–285, 1995, doi: 10.1016/1047-2797(94)00093-9.
- [92] A. Pierron *et al.*, "Évaluation de la qualité métrologique des données du programme de médicalisation du système d'information (PMSI) en périnatalité : étude pilote réalisée dans 3 CHU," *Rev. Epidemiol. Sante Publique*, vol. 63, no. 4, pp. 237–246, 2015, doi: 10.1016/j.respe.2015.05.001.
- [93] J. Guerra *et al.*, "Validité de la description d'un groupe spécifique de patients par le PMSI-SSR," *Rev. Epidemiol. Sante Publique*, vol. 63, no. 4, pp. 247–252, 2015, doi: 10.1016/j.respe.2015.04.013.
- [94] C. Quantin *et al.*, "ÉVALUATION DU PMSI COMME MOYEN D'IDENTIFICATION DES CAS INCIDENTS DE CANCER COLORECTAL," vol. 1, no. 26, pp. 55–63, 2014, [Online]. Available: <http://www.cairn.info/revue-sante-publique-2014-1-page-55.htm>.
- [95] A. A. Chantry *et al.*, "Le programme de médicalisation du système d'information (PMSI) - processus de production des données, validité et sources d'erreurs dans le domaine de la morbidité maternelle sévère," *Rev. Epidemiol. Sante Publique*, 2012, doi: 10.1016/j.respe.2011.11.003.
- [96] D. B. Preen, C. D. A. J. Holman, K. Spilsbury, J. B. Semmens, and K. J. Brameld, "Length of comorbidity lookback period affected regression model performance of administrative health data," *J. Clin. Epidemiol.*, vol. 59, no. 9, pp. 940–946, 2006, doi: 10.1016/j.jclinepi.2005.12.013.
- [97] M. T. A. Sharabiani, P. Aylin, and A. Bottle, "Systematic Review of Comorbidity Indices for Administrative Data," *Med. Care*, 2012, doi: 10.1097/MLR.0b013e31825f64d0.
- [98] D. P. Schauer *et al.*, "Bariatric Surgery and the Risk of Cancer in a Large Multisite Cohort," *Ann. Surg.*, vol. 269, no. 1, pp. 95–101, Jan. 2019, doi: 10.1097/SLA.0000000000002525.
- [99] M. Derogar, M. A. Hull, Y. Lu, and J. Lagergren, "Increased Risk of Colorectal Cancer After Obesity Surgery," *Ann. Surg.*, vol. 258, no. 6, pp. 983–988, 2013, doi: 10.1097/SLA.0b013e318288463a.
- [100] H. Hussan *et al.*, "Prior Bariatric Surgery Is Linked to Improved Colorectal Cancer Surgery Outcomes and Costs: A Propensity-Matched Analysis," *Obes. Surg.*, vol. 27, no. 4, pp. 1047–1055, 2017, doi:

10.1007/s11695-016-2421-8.

- [101] M. Di Cesare *et al.*, “Trends in adult body-mass index in 200 countries from 1975 to 2014: A pooled analysis of 1698 population-based measurement studies with 19.2 million participants,” *Lancet*, vol. 387, no. 10026, pp. 1377–1396, 2016, doi: 10.1016/S0140-6736(16)30054-X.
- [102] T. Pischon *et al.*, “General and Abdominal Adiposity and Risk of Death in Europe,” *N. Engl. J. Med.*, vol. 359, no. 20, pp. 2105–2120, 2008, doi: 10.1056/nejmoa0801891.
- [103] J. L. Schlienger, F. Luca, S. Vinzio, and A. Pradignac, “Obésité et cancer,” *Rev. Med. Interne*, vol. 30, no. 9, pp. 776–782, 2009, doi: 10.1016/j.revmed.2009.04.007.
- [104] M. L. Maciejewski *et al.*, “Bariatric Surgery and Long-term Durability of Weight Loss,” *JAMA Surg*, vol. 151, no. 11, pp. 1046–1055, 2016, doi: 10.1001/jamasurg.2016.2317.
- [105] T. D. Adams *et al.*, “Cancer incidence and mortality after gastric bypass surgery,” *Obesity (Silver Spring)*, vol. 17, no. 4, pp. 796–802, 2009, doi: 10.1038/oby.2008.610.
- [106] L. Sjöström *et al.*, “Effects of bariatric surgery on cancer incidence in obese patients in Sweden (Swedish Obese Subjects Study): a prospective, controlled intervention trial,” *Lancet Oncol.*, vol. 10, no. 7, pp. 653–662, Jul. 2009, doi: 10.1016/S1470-2045(09)70159-7.
- [107] M. P. Östlund, Y. Lu, and J. Lagergren, “Risk of obesity-related cancer after obesity surgery in a population-based cohort study,” *Ann. Surg.*, vol. 252, no. 6, pp. 972–976, 2010, doi: 10.1097/SLA.0b013e3181e33778.
- [108] D. P. Schauer *et al.*, “Bariatric Surgery and the Risk of Cancer in a Large Multisite Cohort,” *Ann. Surg.*, vol. 269, no. 1, pp. 95–101, Jan. 2019, doi: 10.1097/SLA.0000000000002525.
- [109] M. A. Hernan and J. M. Robins, “Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available,” *Am. J. Epidemiol.*, vol. 183, no. 8, pp. 758–764, 2016, doi: 10.1093/aje/kwv254.
- [110] G. Moulis, M. Lapeyre-Mestre, A. Palmaro, G. Pugnet, J. L. Montastruc, and L. Sailler, “French health insurance databases: What interest for medical research?,” *Rev. Med. Interne*, vol. 36, no. 6, pp. 411–417, 2015, doi: 10.1016/j.revmed.2014.11.009.
- [111] M. Lipsitch, E. T. Tchetgen, and T. Cohen, “Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies,” *Epidemiology*, vol. 21, no. 3, pp. 383–388, 2010, doi: 10.1097/EDE.0b013e3181d61eeb.Negative.
- [112] A. I. Phipps *et al.*, “Defining menopausal status in epidemiologic studies: A comparison of multiple approaches and their effects on breast cancer rates,” *Maturitas*, vol. 67, no. 1, pp. 60–66, Sep. 2010, doi: 10.1016/j.maturitas.2010.04.015.
- [113] P. C. Austin, “The performance of different propensity score methods for estimating marginal hazard ratios,” *Stat. Med.*, vol. 32, no. 16, pp. 2837–49, Jul. 2013, doi: 10.1002/sim.5705.
- [114] P. C. Austin and E. A. Stuart, “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.*, vol. 34, no. 28, pp. 3661–3679, 2015, doi: 10.1002/sim.6607.
- [115] R Core, “R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.” 2017.
- [116] A. Aravani, A. Downing, J. D. Thomas, J. Lagergren, E. J. A. Morris, and M. A. Hull, “Obesity surgery and risk of colorectal and other obesity-related cancers: An English population-based cohort study,” *Cancer Epidemiol.*, vol. 53, pp. 99–104, Apr. 2018, doi: 10.1016/j.canep.2018.01.002.
- [117] H. S. Feigelson *et al.*, “Bariatric Surgery is Associated With Reduced Risk of Breast Cancer in Both Premenopausal and Postmenopausal Women,” *Ann. Surg.*, vol. Publish Ah, Apr. 2019, doi:

10.1097/sla.0000000000003331.

- [118] L. Bailly, R. Fabre, C. Pradier, and A. Iannelli, "Colorectal cancer risk following bariatric surgery in a nationwide study of French individuals with obesity," *JAMA Surg.*, vol. 155, no. 5, pp. 395–402, 2020, doi: 10.1001/jamasurg.2020.0089.
- [119] Å. Anveden *et al.*, "Long-term incidence of female-specific cancer after bariatric surgery or usual care in the Swedish Obese Subjects Study," *Gynecol. Oncol.*, vol. 145, no. 2, pp. 224–229, 2017, doi: 10.1016/j.ygyno.2017.02.036.
- [120] W. Tao *et al.*, "Colon and rectal cancer risk after bariatric surgery in a multicountry Nordic cohort study," *Int. J. Cancer*, vol. 147, no. 3, pp. 728–735, Aug. 2020, doi: 10.1002/ijc.32770.
- [121] H. Mackenzie *et al.*, "Obesity surgery and risk of cancer," *Br. J. Surg.*, vol. 105, no. 12, pp. 1650–1657, Nov. 2018, doi: 10.1002/bjs.10914.
- [122] J. Berkson, "Limitations of the application of fourfold table analysis to hospital data," *Int. J. Epidemiol.*, vol. 43, no. 2, pp. 511–515, 2014, doi: 10.1093/ije/dyu022.
- [123] A. Pasquereau, R. Andler, P. Arwidson, R. Guignard, and V. Nguyen-Thanh, "Consommation de tabac parmi les adultes: bilan de cinq années de Programme National contre le Tabagisme, 2014-2019," *Bull. épidémiologique Hebd.*, vol. 14, no. Tabagisme, pp. 274–281, 2020, [Online]. Available: <https://www.santepubliquefrance.fr/determinants-de-sante/tabac/documents/magazines-revues/bulletin-epidemiologique-hebdomadaire-26-mai-2020-n-14-journee-mondiale-sans-tabac-2020>.
- [124] A. Lazzati *et al.*, "Natural History of Adjustable Gastric Banding," *Ann. Surg.*, vol. 265, no. 3, pp. 439–445, 2017, doi: 10.1097/SLA.0000000000001879.
- [125] A.-S. Schneck *et al.*, "One or two steps for laparoscopic conversion of failed adjustable gastric banding to sleeve gastrectomy: A nationwide French study on 3357 morbidly obese patients," *Surg. Obes. Relat. Dis.*, vol. 12, no. 4, 2016, doi: 10.1016/j.soard.2015.10.070.
- [126] K. Spaniolas *et al.*, "Conversion of Adjustable Gastric Banding to Stapling Bariatric Procedures," *Ann. Surg.*, vol. Publish Ah, p. 1, Apr. 2019, doi: 10.1097/sla.0000000000003332.
- [127] S. G. Barreto, J. Chisholm, A. Schloithe, J. Collins, and L. Kow, "Outcomes of Two-Step Revisional Bariatric Surgery: Reasons for the Gastric Banding Explantation Matter," *Obes. Surg.*, vol. 28, no. 2, pp. 520–525, Feb. 2018, doi: 10.1007/s11695-017-2889-x.
- [128] K. Samakar *et al.*, "Safety and efficacy of single-stage conversion of failed adjustable gastric band to laparoscopic Roux-en-Y gastric bypass: a case–control study," *Surg. Endosc.*, vol. 30, no. 12, pp. 5453–5458, 2016, doi: 10.1007/s00464-016-4905-y.
- [129] I. Debergh *et al.*, "A one-step conversion from gastric banding to laparoscopic Roux-en-Y gastric bypass is as safe as a two-step conversion: A comparative analysis of 885 patients," *Acta Chir. Belg.*, vol. 116, no. 5, pp. 271–277, Oct. 2016, doi: 10.1080/00015458.2016.1255005.