Étude des maladies inflammatoires chroniques de l'intestin à partir d'un registre en population générale : contributions à l'épidémiologie et à l'analyse statistique de données cliniques et omiques

Hélène Sarter

UNIVERSITE DE LILLE
Ecole doctorale Biologie-Santé

Thèse de doctorat

Discipline : Santé Publique

Spécialité : Epidémiologie

Présentée par Hélène Sarter

En vue de l'obtention du grade de docteur de l'Université de Lille

Titre : Étude des maladies inflammatoires chroniques de l'intestin à partir d'un registre en population générale: contributions à l'épidémiologie et à l'analyse statistique de données cliniques et omiques.

Présentée et soutenue publiquement le 9 octobre 2024 devant le jury composé de :

Docteur Anne-Marie BOUVIER	Rapporteure
Docteur Julien KIRCHGESNER	Rapporteur
Professeur Grégoire FICHEUR	Examinateur
Professeur Jean-Pierre HUGOT	Examinateur, président du jury
Professeur Alain DUHAMEL	Invité
Docteur Corinne GOWER-ROUSSEAU	Directrice de thèse
Professeur Guillemette MAROT	Directrice de thèse

UNITE D'ACCUEIL

Laboratoire Infinite - Institute for Translational Research in Inflammation | Infinite U1286 | Inserm | Université de Lille | CHU de Lille

Faculté de Médecine Pôle Recherche, 4ème étage centre, Place de Verdun, 59045 Lille Cedex

http://lille-inflammation-research.org

REMERCIEMENTS

Je tiens tout d'abord à remercier très chaleureusement mes deux directrices de thèse, Corinne Gower-Rousseau et Guillemette Marot, pour m'avoir fait confiance et m'avoir accompagnée tout au (très !) long de ce travail. Chacune dans son domaine m'a beaucoup apporté! Corinne, merci de m'avoir accueillie dans l'équipe du registre et pour toutes ces années passées à travailler ensemble. Merci pour tout!

Merci à Anne-Marie Bouvier et Julien Kirchgesner, vous me faites l'honneur d'être rapporteurs de cette thèse.

Merci à Grégoire Ficheur et Jean-Pierre Hugot d'avoir accepté d'être examinateurs de cette thèse, c'est un plaisir pour moi de vous avoir dans ce jury.

Merci à Alain Duhamel d'avoir accepté d'évaluer ce travail. Tu m'as accueillie dans ton service à mon arrivée au CHU et j'ai beaucoup appris à tes côtés, toujours dans la bonne humeur.

Merci à l'unité Inserm Infinite qui m'a accueillie pour cette thèse.

Merci à toute l'équipe du registre à Lille, Amiens et Rouen, sans qui ce travail n'aurait pas été possible. Merci à Ariane pour tout ce tu fais pour le registre et pour les joyeux lundis. Merci aux enquêtrices, Anne, Elise, Hélène, Lucie, Nathalie, Stéphanie, pour ce beau travail d'équipe! Un grand merci à Dominique Turck qui m'a agréablement accompagnée dans la coordination du registre au départ de Corinne. Un autre grand merci à Delphine Ley qui a pris sa suite et avec qui je partage, toujours aussi agréablement, la coordination du registre. Merci, bien sûr, à Mathurin Fumery et Guillaume Savoye, indispensables au registre, pour leur foisonnement d'idées brillantes. Merci à Pauline Wils et Nicolas Richard. Vous tous participez à la qualité scientifique du registre Epimad mais aussi à l'ambiance agréable, respectueuse et détendue des conseils scientifiques du registre.

Merci à Maryline Roy pour son aide à la préparation du génotypage et les quelques heures passées ensemble à remplir des plaques d'ADN. Merci à Antoine Lamer et Niels Martignène pour avoir développé le site predict-epimad.

Merci à toute l'équipe de service de Santé Publique du CHU de Lille pour les bons moments partagés au quotidien.

Merci à toute l'équipe de l'étude Mikinautes avec laquelle c'est un plaisir de travailler.

Merci à toutes les personnes avec lesquelles j'ai travaillé de près ou de loin, à Paris puis à Lille.

Merci à l'ensemble des gastroentérologues qui participent au registre et ont permis l'ensemble de ces travaux.

Merci à mes parents (sans qui rien n'aurait été possible), ma famille, ma belle-famille, mes amis, à Maïna pour avoir vécu avec moi cette fin de thèse en parallèle de la sienne.

Je remercie enfin Michael, pour tout, et mes enfants, Malo, Maël et Lison qui m'ont permis, chaque jour, de me souvenir qu'il y avait plus important que cette thèse. Merci Michael pour ta proposition de titre « Crohn'ique d'une thèse tant attendue » qui collait bien mais n'a pas été retenue finalement…

Titre : Étude des maladies inflammatoires chroniques de l'intestin à partir d'un registre en population générale: contributions à l'épidémiologie et à l'analyse statistique de données cliniques et omiques

Mots clés : Maladies inflammatoires chroniques de l'intestin, Crohn, Rectocolite Hémorragique, Epidémiologie, Biostatistique, Données omiques

Résumé : Les maladies inflammatoires chroniques de l'intestin (MICI), incluant la maladie de Crohn (MC) et la rectocolite hémorragique (RCH), sont des maladies chroniques évoluant par poussées entrecoupées de rémissions, pouvant mener à des complications (sténoses, fistules, cancer..) et des lésions irréversibles de l'intestin. Les MICI se manifestent souvent à un jeune âge et impactent fortement la qualité de vie. Leur fréquence élevée en Europe du Nord en fait un problème de santé publique, justifiant la mise en place du registre épidémiologique Epimad dans le Nord-Ouest de la France en 1988. Cette thèse vise à apporter des connaissances sur les MICI à partir des données de ce registre en population générale selon deux axes : épidémiologie et analyse de données cliniques et omiques.

Dans une première partie épidémiologique, nous avons étudié les incidences de MICI à partir de 22 879 cas incidents sur une période de 30 ans (1988-2017). L'incidence de la MC a augmenté de 5,1 à 7,9 (Annual percent change (APC) : +1,9 % [1,6 ; 2,2]), et celle de la RCH de 4,5 à 6,1 (APC : +1,3 % [0,9 ; 1,7]). Cette augmentation était particulièrement marquée chez les enfants et les jeunes adultes. Pour la RCH, l'incidence augmentait plus chez les femmes, atteignant celle des hommes en fin de période d'étude. En 2030, près de 0,6 % de la population du nord de la France pourrait être atteinte de MICI, avec un vieillissement de cette population.
Dans une deuxième étude épidémiologique menée sur 361 patients ayant débuté leur MICI dans l'enfance, nous avons montré que le taux de chômage était significativement plus bas parmi les patients (9 % contre 15 % dans la population générale de même âge et même sexe), et que le niveau d'études était plus élevé (57 % avaient obtenu un diplôme de l'enseignement supérieur contre 41 % dans la population générale). Cela souligne les capacités d'adaptation des patients ayant déclaré leur maladie dans l'enfance.

Dans la seconde partie, nous avons analysé des données cliniques et omiques. Un premier chapitre a comparé plusieurs méthodes d'apprentissage statistique sur des simulations de données. Nous avons conclu que l'approche en deux étapes, avec sélection des variables cliniques par analyses univariées et sélection des variables génétiques par Lasso avec stability selection au seuil de 0,7, était un choix raisonnable. Nous avons appliqué ces méthodes à des données de 156 patients du registre Epimad pour construire un modèle, PREDICT-EPIMAD, basé sur la combinaison de 8 facteurs cliniques, sérologiques et génétiques pour prédire le risque individuel à cinq ans de résection intestinale et/ou de complications sténosantes ou pénétrantes de la MC à début pédiatrique. La validation interne a montré une bonne discrimination et calibration du modèle, ainsi que son utilité clinique.

En conclusion, les MICI demeurent une question de santé publique importante avec une incidence croissante, en particulier chez les jeunes. Le coût des nouvelles thérapeutiques accentue leur poids sur le système de santé. Nous devons préparer notre système de santé et la recherche sur des méthodes de stratification des patients selon leur risque de complication et leur réponse au traitement doit se poursuivre. Ces travaux ont également montré l'importance des registres épidémiologiques, permettant de produire des données exhaustives et détaillées en population générale et de soutenir des études observationnelles et analytiques pour améliorer la connaissance sur les MICI.

Title: Study of Inflammatory Bowel Diseases from a Population-based Registry: Contributions to Epidemiology and Statistical Analysis of Clinical and Omics Data.

Keywords: Inflammatory Bowel Diseases, Crohn's Disease, Ulcerative Colitis, Epidemiology, Biostatistics, Omics Data.

Abstract: Inflammatory bowel diseases (IBD), including Crohn's disease (CD) and ulcerative colitis (UC), are chronic conditions characterized by an alternation of remission and flares, which can lead to complications (strictures, fistulas, cancer) and irreversible intestinal damage. IBD often manifest at a young age and significantly impact quality of life. Their high prevalence in Northern Europe constitutes a public health issue, justifying the establishment of the Epimad epidemiological registry in North of France in 1988. This work aims to provide insights into IBD based on data from this population-based registry through two main axes: epidemiology and the analysis of clinical and omics data.

In the first epidemiological part, we studied the incidences of IBD using 22,879 incident cases over a 30-year period (1988-2017). The incidence of CD increased from 5.1 to 7.9 (Annual percent change (APC): +1.9% [1.6; 2.2]), and that of UC from 4.5 to 6.1 (APC: +1.3% [0.9; 1.7]). This increase was particularly marked among children and young adults. For UC, the incidence rose more among women, reaching that of men by the end of the study period. By 2030, nearly 0.6% of the population in Northern France could be affected by IBD, with this population aging.
In a second epidemiological study conducted on 361 patients who developed IBD during childhood, we found that the unemployment rate was significantly lower among patients (9% compared to 15% in the general population of the same age and sex), and that the level of education was higher (57% had obtained a higher education degree compared to 41% in the general population). This highlights the capacities to cope of patients who developed their disease during childhood.

In the second part, we analyzed clinical and omics data. The first chapter compared several statistical learning methods on data simulations. We concluded that the two-step approach, with the selection of clinical variables by univariate analyses and selection of genetic variables by Lasso with stability selection at a 0.7 threshold, was a reasonable choice. We applied these methods to data from 156 patients in the Epimad registry to construct a model, called PREDICT-EPIMAD, based on the combination of 8 clinical, serological, and genetic factors to predict the individual five-year risk of intestinal resection and/or stenosing or penetrating complications of pediatric-onset CD. Internal validation showed good discrimination and calibration of the model, as well as its clinical utility.

In conclusion, IBD remain a significant public health issue with increasing incidence, particularly among young people. The cost of new therapeutics accentuates their burden on the healthcare system. We must prepare our healthcare system, and research on methods for stratifying patients according to their risk of complications and treatment response must continue. This work also highlighted the importance of epidemiological registries, enabling the production of exhaustive and detailed data in the general population and supporting observational and analytical studies to improve knowledge about IBD.

VALORISATIONS SCIENTIFIQUES

Publications associées à la thèse

Sarter H, Savoye G, Marot G, Ley D, Turck D, Hugot JP, Vasseur F, Duhamel A, Wils P, Princen F, Colombel JF, Gower-Rousseau C, Fumery M; EPIMAD study group. A Novel 8-Predictors Signature to Predict Complicated Disease Course in Pediatric-onset Crohn's Disease: A Population-based Study. Inflamm Bowel Dis. 2023;29(11):1793-1804.

Sarter H, Cretin T, Savoye G, Fumery F, Leroyer A, Dauchet L, Paupard T, Coevoet H, Wils P, Richard N, Turck D, Ley D, Gower-Rousseau C. Increasing incidence of inflammatory bowel diseases in children and young adults in Northern France: a 30-year study. Soumis à The Lancet Regional Health Europe.

Sarter H, Le Coniac M, Leroyer A, Savoye G, Fumery M, Guillon N, Gower-Rousseau C, Ley D, Turck D. Young adult patients with paediatric-onset inflammatory bowel disease have a higher educational level and a higher employment rate than the general population. Soumis à United European Gastroenterology Journal (UEGJ).

Communications orales et affichées associées à la thèse

Sarter H, Savoye G, Turck D, Vasseur F, Marot G, Pariente B, Singh S, Colombel JF, Gower-Rousseau C, Fumery M. Une combinaison de facteurs cliniques, sérologiques et génétiques prédit l'évolution vers une forme compliquée dans la maladie de Crohn à début pédiatrique : résultats d'une étude en population générale**. Communication orale** aux Journées Françaises d'Hépato-gastroentérologie et d'Oncologie Digestive, Paris, 21-24 Mars 2019

Sarter H, Savoye G, Turck D, Vasseur F, Marot G, Pariente B, Singh S, Colombel JF, Gower-Rousseau C, Fumery M. A combination of clinical, serological and genetic factors predicts complicated disease course in paediatric-onset Crohn's disease: Results from a population-based study. C**ommunication affichée** à l'United European of Gastroenterological Week, Vienna, 21-24 octobre 2018.

Sarter H, Savoye G, Turck D, Vasseur F, Marot G, Pariente B, Singh S, Colombel JF, Gower-Rousseau C, Fumery M. A combination of clinical, serological and genetic factors predicts complicated disease course in paediatric-onset Crohn's disease: Results from a population-based study. **Communication affichée** à la Digestive Disease Week, Washington D, USA, 2-5 juin 2018.

Sarter H, Savoye G, Turck D, Vasseur F, Marot G, Pariente B, Singh S, Colombel JF, Gower-Rousseau C, Fumery M. A combination of clinical, serological and genetic factors predicts complicated disease course in paediatric-onset Crohn's disease: Results from a population-based study. **Communication orale en séance plénière** à l'European Crohn's and Colitis, Vienna, 15-17 février 2018

Sarter H, Gower-Rousseau C, Marot G. Influence of SNP coding on the analysis of disease risk. **Communication affichée** aux Journées Ouvertes en Biologie, Informatique et Mathématiques, Lille, 3-6 juillet 2017.

Autres publications en rapport avec le sujet de thèse

Ghione S, **Sarter H**, Fumery M, Armengol-Debeir L, Savoye G, Ley D, Spyckerelle C, Pariente B, Peyrin-Biroulet L, Turck D, Gower-Rousseau C; Epimad Group. Dramatic Increase in Incidence of Ulcerative Colitis and Crohn's Disease (1988-2011): A Population-Based Study of French Adolescents. Am J Gastroenterol. 2018;113(2):265-272.

Gower-Rousseau C, **Sarter H**, Savoye G, Tavernier N, Fumery M, Sandborn WJ, Feagan BG, Duhamel A, Guillon-Dellac N, Colombel JF, Peyrin-Biroulet L; International Programme to Develop New Indexes for Crohn's Disease (IPNIC) group; Validation of the Inflammatory Bowel Disease Disability Index in a population-based cohort. Gut. 2017;66(4):588-596

Shafer LA, Walker JR, Chhibba T, Ivekovic M, Singh H, Targownik LE, Peyrin-Biroulet L, Gower-Rousseau C, **Sarter H**, Bernstein CN. Independent Validation of a Self-Report Version of the IBD Disability Index (IBDDI) in a Population-Based Cohort of IBD Patients. Inflamm Bowel Dis. 2018;24(4):766-774.

Bequet E, **Sarter H**, Fumery M, Vasseur F, Armengol-Debeir L, Pariente B, Ley D, Spyckerelle C, Coevoet H, Laberenne JE, Peyrin-Biroulet L, Savoye G, Turck D, Gower-Rousseau C; EPIMAD Group. Incidence and Phenotype at Diagnosis of Very-early-onset Compared with Later-onset Paediatric Inflammatory Bowel Disease: A Population-based Study [1988-2011]. J Crohns Colitis. 2017;11(5):519-526.

Williet N, **Sarter H**, Gower-Rousseau C, Adrianjafy C, Olympie A, Buisson A, Beaugerie L, Peyrin-Biroulet L. Patient-reported Outcomes in a French Nationwide Survey of Inflammatory Bowel Disease Patients. J Crohns Colitis. 2017;11(2):165-174.

Autres publications dans le domaine des maladies inflammatoires chroniques de l'intestin (depuis 2017)

Ley D, Leroyer A, Dupont C, **Sarter H**, Bertrand V, Spyckerelle C, Guillon N, Wils P, Savoye G, Turck D, Gower-Rousseau C, Fumery M; Epimad Group. New Therapeutic Strategies Are Associated With a Significant Decrease in Colectomy Rate in Pediatric Ulcerative Colitis. Am J Gastroenterol. 2023;118(11):1997-2004.

Mortreux P, Leroyer A, Dupont C, Ley D, Bertrand V, Spyckerelle C, Guillon N, Wils P, Gower-Rousseau C, Savoye G, Fumery M, Turck D, Siproudhis L, **Sarter H**. Natural History of Anal Ulcerations in Pediatric-Onset Crohn's Disease: Long-Term Follow-Up of a Population-Based Study. Am J Gastroenterol. 2023;118(9):1671-1678.

Fumery M, Dupont C, Ley D, Savoye G, Bertrand V, Guillon N, Wils P, Gower-Rousseau C, **Sarter H**, Turck D, Leroyer A. Long-term effectiveness and safety of anti-TNF in pediatric-onset inflammatory bowel diseases: A population-based study. Dig Liver Dis. 2024;56(1):21-28.

Dupont-Lucas C, Leroyer A, Ley D, Spyckerelle C, Bertrand V, Turck D, Savoye G, Maunoury V, Guillon N, Fumery M, **Sarter H**, Gower-Rousseau C; EPIMAD Study Group. Increased Risk of Cancer and Mortality in a Large French Population-Based Paediatric-Onset Inflammatory Bowel Disease Retrospective Cohort. J Crohns Colitis. 2023;17(4):524-534.

Ley D, Leroyer A, Dupont C, **Sarter H**, Bertrand V, Spyckerelle C, Guillon N, Wils P, Savoye G, Turck D, Gower-Rousseau C, Fumery M; Epimad Group. New Therapeutic Strategies Have Changed the Natural History of Pediatric Crohn's Disease: A Two-Decade Population-Based Study. Clin Gastroenterol Hepatol. 2022;20(11):2588-2597.

Sendid B, Salvetat N, **Sarter H**, Loridant S, Cunisse C, François N, Aijjou R, Gelé P, Leroy J, Deplanque D, Jawhara S, Weissmann D, Desreumaux P, Gower-Rousseau C, Colombel JF, Poulain D. A Pilot Clinical Study on Post-Operative Recurrence Provides Biological Clues for a Role of Candida Yeasts and Fluconazole in Crohn's Disease. J Fungi (Basel). 2021;7(5):324.

Danielou M, **Sarter H**, Pariente B, Fumery M, Ley D, Mamona C, Barthoulot M, Charpentier C, Siproudhis L, Savoye G, Gower-Rousseau C; EPIMAD Group. Natural History of Perianal Fistulising Lesions in Patients With Elderly-onset Crohn's Disease: A Population-based Study. J Crohns Colitis. 2020;14(4):501-507.

Genin M, Fumery M, Occelli F, Savoye G, Pariente B, Dauchet L, Giovannelli J, Vignal C, Body-Malapel M, **Sarter H**, Gower-Rousseau C, Ficheur G. Fine-scale geographical distribution and ecological risk factors for Crohn's disease in France (2007-2014). Aliment Pharmacol Ther. 2020;51(1):139-148.

Loreau J, Duricova D, Gower-Rousseau C, Savoye G, Ganry O, Ben Khadhra H, **Sarter H**, Yzet C, Le Mouel JP, Kohut M, Brazier F, Chatelain D, Nguyen-Khac E, Dupas JL, Fumery M. Long-Term Natural History of Microscopic Colitis: A Population-Based Cohort. Clin Transl Gastroenterol. 2019;10(9):e00071.

Fumery M, Pariente B, **Sarter H**, Savoye G, Spyckerelle C, Djeddi D, Mouterde O, Bouguen G, Ley D, Peneau A, Dupas JL, Turck D, Gower-Rousseau C; Epimad Group. Long-term outcome of pediatric-onset Crohn's disease: A population-based cohort study. Dig Liver Dis. 2019;51(4):496-502.

Duricova D, **Sarter H**, Savoye G, Leroyer A, Pariente B, Armengol-Debeir L, Bouguen G, Ley D, Turck D, Templier C, Buche S, Peyrin-Biroulet L, Gower-Rousseau C, Fumery M; Epimad Group. Impact of Extra-Intestinal Manifestations at Diagnosis on Disease Outcome in Pediatric- and Elderly-Onset Crohn's Disease: A French Population-Based Study. Inflamm Bowel Dis. 2019;25(2):394-402.

Duricova D, Pariente B, **Sarter H**, Fumery M, Leroyer A, Charpentier C, Armengol-Debeir L, Peyrin-Biroulet L, Savoye G, Gower-Rousseau C; Epimad Group. Impact of age at diagnosis on natural history of patients with elderly-onset ulcerative colitis: A French population-based study. Dig Liver Dis. 2018;50(9):903-909.

Sacleux SC, **Sarter H**, Fumery M, Charpentier C, Guillon-Dellac N, Coevoet H, Pariente B, Peyrin-Biroulet L, Gower-Rousseau C, Savoye G; EPIMAD Group. Post-operative complications in elderly onset inflammatory bowel disease: a population-based study. Aliment Pharmacol Ther. 2018;47(12):1652-1660.

Sendid B, Jawhara S, **Sarter H**, Maboudou P, Thierny C, Gower-Rousseau C, Colombel JF, Poulain D. Uric acid levels are independent of anti-Saccharomyces cerevisiae antibodies (ASCA) in Crohn's disease: A reappraisal of the role of S. cerevisiae in this setting. Virulence. 2018;9(1):1224-1229.

Lo B, Prosberg MV, Gluud LL, Chan W, Leong RW, van der List E, van der Have M, **Sarter H**, Gower-Rousseau C, Peyrin-Biroulet L, Vind I, Burisch J. Systematic review and meta-analysis: assessment of factors affecting disability in inflammatory bowel disease and the reliability of the inflammatory bowel disease disability index. Aliment Pharmacol Ther. 2018 ;47(1):6-15.

Chau A, Prodeau M, **Sarter H,** Gower C, Rogosnitzky M, Panis Y, Zerbib P. Persistent perineal sinus after abdominoperineal resection. Langenbecks Arch Surg. 2017;402(7):1063-1069.

Hochart A, Gower-Rousseau C, **Sarter H**, Fumery M, Ley D, Spyckerelle C, Peyrin-Biroulet L, Laberenne JE, Vasseur F, Savoye G, Turck D; Epimad Group. Ulcerative proctitis is a frequent location of paediatric-onset UC and not a minor disease: a population-based study. Gut. 2017;66(11):1912-1917.

Duricova D, Leroyer A, Savoye G, **Sarter H**, Pariente B, Aoucheta D, Armengol-Debeir L, Ley D, Turck D, Peyrin-Biroulet L, Gower-Rousseau C, Fumery M; EPIMAD Group. Extra-intestinal Manifestations at Diagnosis in Paediatric- and Elderly-onset Ulcerative Colitis are Associated With a More Severe Disease Outcome: A Population-based Study. J Crohns Colitis. 2017;11(11):1326-1334.

Collins M, **Sarter H**, Gower-Rousseau C, Koriche D, Libier L, Nachury M, Cortot A, Zerbib P, Blanc P, Desreumaux P, Colombel JF, Peyrin-Biroulet L, Pineton de Chambrun G. Previous Exposure to Multiple Anti-TNF Is Associated with Decreased Efficiency in Preventing Postoperative Crohn's Disease Recurrence. J Crohns Colitis. 2017;11(3):281-288.

Table des matières

Liste des abréviations :

5-ASA	5-Aminosalicylates
afa	Association François Aupetit
AIEC	Escherichia coli adhérents et invasifs
ANCA	Anticorps anti-cytoplasme des polynucléaires neutrophiles
APC	Annual Percent Change (Pourcentage de variation annuel)
ASCA	Anticorps anti-Saccharomyces cerevisiae
AUC	Area under the ROC curve (Aire sous la courbe ROC)
CI	Colite chronique inclassable
CNIL	Commission nationale de l'informatique et des libertés
CPP	Comité de Protection des Personnes
CRP	C-reactive protein
CSP	Cholangite sclérosante primitive
DIM	Département d'Information Médicale
FDep	French Deprivation index
GWAS	Genome Wide Association Study
HBI	Harvey-Bradshaw Index
HR	Hazard Ratio
IBD-DI	Inflammatory Bowel Disease – Disability Index
IBDQ	Inflammatory Bowel Disease Questionnaire
IC	Intervalle de confiance
IgA	Immunoglobulines A
IgG	Immunoglobulines B
INSEE	Institut National de la Statistique et des Etudes Economiques
IQR	Interquartile Range (Intervalle interquartile)
IRM	Imagerie par résonance magnétique
IRR	Incidence Rate Ratio (Rapport de taux d'incidence)
LASSO	Least Absolute Shrinkage and Selection Operator
MC	Maladie de Crohn
MICI	Maladies inflammatoires chroniques de l'intestin

SCCAI	Simple Clinical Colitis Activity Index
SGL	Sparse Group Lasso
SNDS	Système National des Données de Santé
SNP	Single Nucleotide Polymorphism
SPIRIT	Selecting End PoInts foR Disease-ModIfication Trials
STRIDE	Selecting Therapeutic Targets in Inflammatory Bowel Disease
OLS	Ordinary Least Squares (Moindres carrés ordinaires)
OR	Odds ratio
PA	Personnes-années
PCS	Professions et catégories socioprofessionnelles
PLS	Partial Least Squares
PMSI	Programme de Médicalisation des Systèmes d'Information
PRO	Patient Reported Outcome
RCH	Rectocolite Hémorragique
ROC	Receiver Operating Characteristic
RR	Risque Relatif
SGCCA	Sparse generalized canonical correlation analysis
SIBDQ	Short Inflammatory Bowel Disease Questionnaire
VPN	Valeur Prédictive Négative
VPP	Valeur Prédictive Positive

LISTE DES TABLEAUX

LISTE DES FIGURES

INTRODUCTION GENERALE

Les maladies inflammatoires chroniques de l'intestin

1. Définition

Les maladies inflammatoires chroniques de l'intestin (MICI) regroupent deux entités distinctes : la maladie de Crohn (MC) et la rectocolite hémorragique (RCH) (1,2). Une troisième catégorie, les maladies inflammatoires chroniques de l'intestin non classées (CI), concerne un petit groupe de patients pour lesquels la différenciation entre la MC et la RCH est difficile malgré un bilan complet. L'inflammation de l'intestin dans les CI est limitée au côlon et peut présenter des caractéristiques à la fois de la MC et de la RCH.

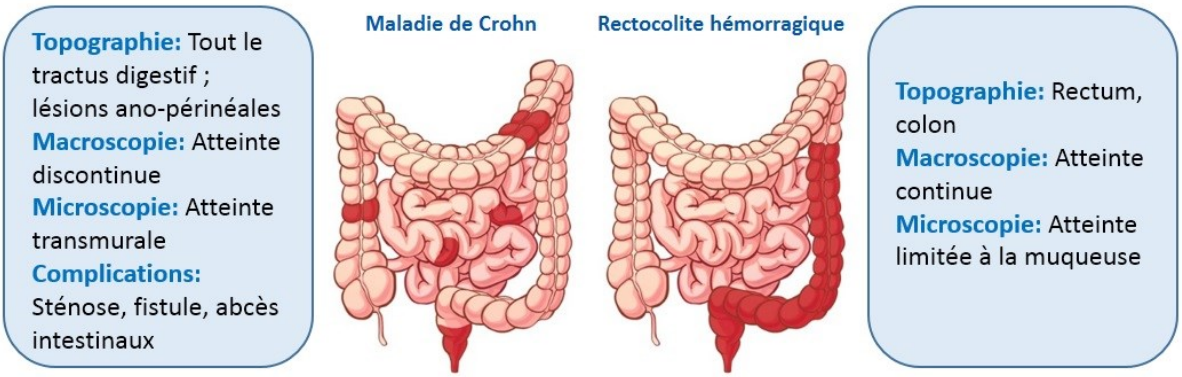
Les MICI sont des maladies chroniques, évoluant par phases de poussée entrecoupées de rémissions et dont l'évolution est progressive, pouvant mener à des complications et des lésions irréversibles de l'intestin. Les MICI se manifestent souvent à un âge jeune, majoritairement entre 20 et 30 pour la MC et entre 30 et 40 ans pour la RCH. Ces maladies, évoluant tout au long de la vie du patient, ont un impact fort sur la qualité de vie et peuvent entraîner un handicap fonctionnel important.

La MC se caractérise par une atteinte discontinue des voies digestives, pouvant atteindre l'ensemble des segments du tube digestif, de la bouche à l'anus. La partie la plus fréquemment touchée est l'iléon terminal. La RCH, quant à elle, consiste en une atteinte continue, sans espace de muqueuse saine, limitée au rectum et au colon. Une autre distinction importante entre les deux pathologies réside dans le fait qu'alors que, dans la RCH, l'atteinte de la muqueuse intestinale reste superficielle, la MC se caractérise par une atteinte plus profonde, transmurale, pouvant mener à des complications telles que des sténoses, des fistules et des abcès (Figure 1).

Le diagnostic des MICI est basé sur des critères cliniques, endoscopiques, radiologiques et histologiques. Les deux pathologies présentent des symptômes communs. Cependant, la MC se caractérise davantage par des douleurs abdominales, des diarrhées, une perte de poids, de la fatigue alors que les symptômes prédominants dans la RCH sont des douleurs abdominales, un syndrome rectal (fausses envies, selles glairo-sanglantes) et la présence de sang dans les

selles. Les MICI peuvent également s'accompagner de manifestations extradigestives (dermatologiques, oculaires, hépatiques, articulaires), plus fréquentes dans la MC.

Figure 1. Différences de présentation entre la maladie de Crohn et la Rectocolite Hémorragique.



Depuis 2005, la classification de Montréal est utilisée pour décrire les MICI (3). Pour la MC on distingue les localisations suivantes (Table 1) : L1 pour l'iléon seul, L2 pour le côlon seul, L3 pour l'atteinte iléo-colique ; L4 est ajouté en cas d'atteinte des voies digestives supérieures (œsophage, estomac, duodénum, jejunum). Trois phénotypes sont également décrits : B1 pour la forme purement inflammatoire, B2 pour la forme sténosante et B3 pour la forme pénétrante (fistules, abcès). La lettre "p" est ajoutée en cas d'atteinte ano-périnéale (abcès ou fistule). Les maladies sténosantes et/ou pénétrantes (B2 et/ou B3) sont regroupées sous le terme de « phénotype compliqué ».

Dans la RCH, on décrit la localisation selon l'extension des lésions au niveau du colon et du rectum (Table 2). E1 caractérise l'atteinte confinée au rectum, aussi appelée proctite ou rectite, E2 regroupe les maladies dont les lésions ne dépassent pas l'angle gauche colique, et enfin E3 désigne les atteintes les plus étendues, dépassant l'angle gauche colique. Le terme de pancolite désigne l'inflammation généralisée du colon. A noter, pour la RCH pédiatrique, la classification de Paris distingue E3 (atteinte colique dépassant l'angle gauche mais pas l'angle droit) et E4 (lésions coliques dépassant l'angle droit ou pancolite) (4).

Table 1. Classification de Montréal de la maladie de Crohn

Age au diagnostic	A1	< 17 ans
	A2	≥17 et <40 ans
	A3	Plus de 40 ans
Localisation	L1	Iléon seul
	L2	Colon seul
	L3	Iléon et colon
	L4	Atteinte digestive haute
Phénotype	B1	Inflammatoire
	B2	Sténosant
	B3	Pénétrant
	p	Atteinte ano-périnéale

Table 2. Classification de Montréal de la RCH

Localisation	E1	Rectum
	E2	En aval de l'angle gauche
	E3	En amont de l'angle gauche

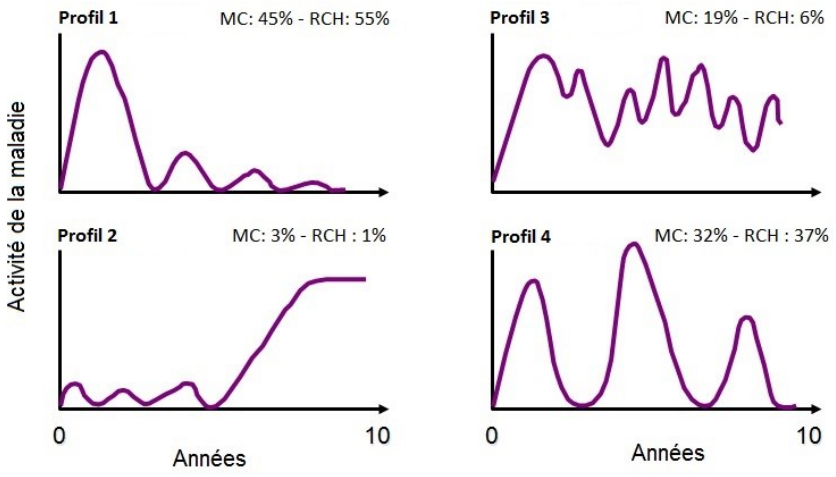
2. Histoire naturelle des MICI

Les MICI évoluent par une alternance de phases de rémissions et de poussées, caractérisées par des symptômes cliniques associés à des signes d'inflammation biologiques, endoscopiques et histologiques.

Quatre profils évolutifs ont été décrits à partir du ressenti des patients de la cohorte norvégienne IBSEN (Figure 2) (5,6):

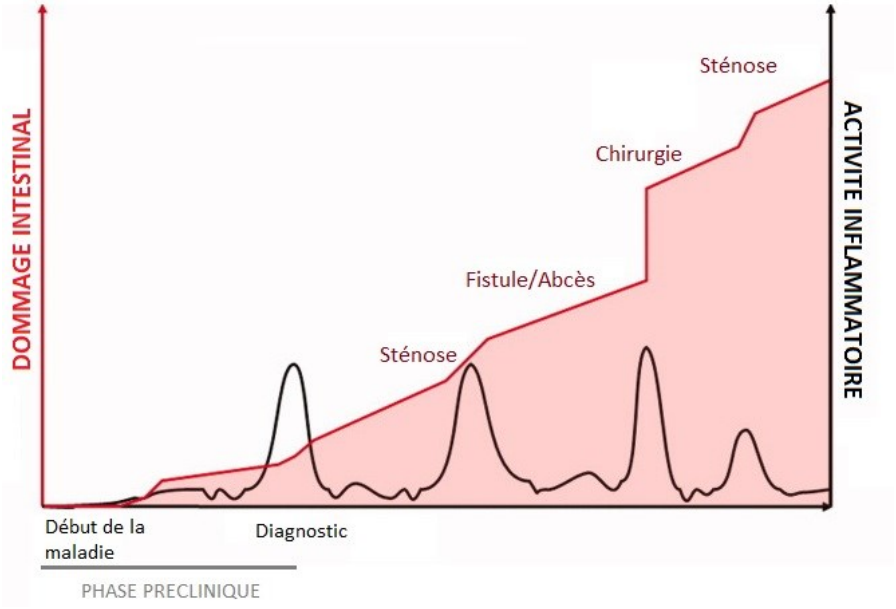
- Profil 1 : forte poussée inaugurale suivie d'une rémission ou d'une amélioration avec des poussées peu intenses (45 % des MC – 55 % des RCH)
- Profil 2 : aggravation tardive de la maladie (3 % des MC – 1 % des RCH)
- Profil 3 : maladie chronique active (19 % des MC – 6 % des RCH)
- Profil 4 : maladie chronique intermittente (32 % des MC – 37 % des RCH)

Figure 2. Classification des MICI en 4 profils évolutifs, issue de la cohorte danoise IBSEN, *d'après Solberg, CGH, 2007 et Solberg, Scand J Gastroenterol, 2009.*



Le caractère progressif de la MC est bien connu (1,7). Même pendant les périodes de rémission clinique, une inflammation subclinique peut persister et évoluer vers une sténose ou des lésions fistulisantes de l'intestin, reflétant une évolution progressive de la maladie pouvant mener à des dommages irréversibles de l'intestin et éventuellement à une résection chirurgicale. Après la chirurgie, ce cycle peut se répéter, entraînant une perte progressive de la fonction intestinale et un handicap fonctionnel important (Figure 3).

Figure 3. Illustration de la progression de la destruction intestinale en parallèle de l'activité inflammatoire, *tiré de Pariente, IBD, 2011.*



Le score de Lémann a été construit pour mesurer le degré de destruction intestinale (8,9). Sur la base de ce score, près de deux tiers des patients atteints de MC présentent des lésions muqueuses substantielles 2 à 10 ans après le diagnostic (10).

Alors qu'une majeure partie des patients atteints de MC présente un phénotype purement inflammatoire au moment du diagnostic (entre 60 et 80 % selon les études), un grand nombre évolue vers des lésions sténosantes et/ou pénétrantes à long terme (11–14). Au final, en population générale, 50 % des patients présenteront une complication sténosante ou pénétrante dans les 20 ans suivant le diagnostic et 50 % auront une résection intestinale dans les 10 ans suivant le diagnostic avec un risque de récurrence de 50 % dans les 10 années suivant la première résection (13).

La RCH, atteignant uniquement la muqueuse et n'étant généralement pas accompagnée des complications (sténoses, fistules) qui peuvent être observées dans la MC, est souvent considérée comme une maladie moins péjorative. Cependant, le caractère progressif de la RCH est de plus en plus reconnu et peut mener à un fonctionnement colique altéré et à une inflammation persistante (15–17). Au final, il est bien admis que les deux maladies représentent un fardeau similaire pour les patients notamment en termes d'impact sur la qualité de vie (18). La probabilité cumulée d'extension colique est de 31 % à 10 ans (19). Le risque cumulé de colectomie est de 15 % à 10 ans (16). Les complications de la RCH peuvent également prendre la forme de sténoses bien que plus rares que dans la MC (4 % à 10 ans), de pseudo polypose, de dysfonction anorectale et d'altération de la perméabilité intestinale (15,20).

Des études en population suggèrent un léger excès de mortalité dans la MC mais pas dans la RCH (21,22). Un sur-risque de cancer colorectal chez les patients atteints de RCH et de MC colique a été mis en évidence : le risque de cancer est quasiment doublé pour ces patients et le risque cumulé à 20 ans est de 5 % (23–25). Ce sur-risque n'apparaît qu'après 8-10 ans d'évolution et augmente ensuite de 0,5 à 1 % par an (26). Ce risque est également augmenté lorsque la maladie est associée à une cholangite sclérosante primitive (CSP) et ce, dès le diagnostic de CSP et quelle que soit l'ancienneté de la MICI. Par ailleurs, le risque de cholangiocarcinome est également augmenté dans la population atteinte de MICI, en particulier en cas de RCH, en lien avec l'association à une CSP (27).

3. Epidémiologie descriptive

HISTOIRE DE L'EMERGENCE

Les MICI ont d'abord émergé dans les pays occidentaux suite à l'industrialisation. Bien que des cas cliniques présentant des symptômes compatibles avec des MICI aient été décrits dès la fin du 18ème siècle puis au cours du 19ème siècle, c'est en 1888 et en 1932, respectivement, que les termes de RCH et de maladie de Crohn furent introduits par White et Crohn, respectivement (28,29). Au début du 20ème siècle, la prise de conscience autour de la RCH s'est accélérée en parallèle de l'augmentation de son incidence. Dès 1909, la RCH était décrite par W.H. Allchin comme une affection « of no great rarity » (30). La première description de la MC a permis une distinction claire entre les deux entités, la MC ayant été décrite comme une atteinte uniquement iléale. La description de l'atteinte colique de la MC n'a été faite que dans les années 1960 (31). L'existence de colites indéterminées pour lesquelles la maladie ne peut être catégorisée de manière spécifique en MC ou en en RCH a été décrite en 1978 (32). Dans les pays occidentaux, l'incidence des MICI a fortement augmenté dans la deuxième moitié du 20ème siècle. La maladie n'a émergé dans les autres pays que plus tardivement - en parallèle de l'industrialisation et de l'adoption d'un régime alimentaire et d'un mode de vie occidentaux - dans la seconde moitié du 20ème siècle pour les nouveaux pays industrialisés d'Asie et d'Amérique du Sud et plus tardivement encore, au début du 21ème siècle dans les pays en développement. Il est important de noter que, dans ces pays, les données sont peu nombreuses, notamment en Afrique. Par ailleurs, dans ces pays, le diagnostic est compliqué en l'absence d'examens endoscopiques et du diagnostic différentiel avec certaines parasitoses ou la tuberculose digestive.

STADES EPIDEMIQUES

En épidémiologie, l'incidence mesure le nombre de nouveaux cas survenant dans une population définie sur une période de temps spécifiée (au cours d'une année par exemple), indiquant le risque de développer la maladie. Elle est généralement exprimée en termes de nombre de nouveaux cas pour 100 000 personnes-années. La prévalence mesure le nombre total de cas (nouveaux et anciens) à un moment donné, reflétant l'ampleur globale de la maladie dans la population.

Kaplan et Windsor ont récemment décrit 4 stades épidémiologiques des MICI (Figure 4) (33) :

- Stade 1 : émergence,
- Stade 2 : accélération de l'incidence,
- Stade 3 : aggravation de la prévalence (« compounding prevalence »)
- Stade 4 : équilibre de la prévalence (« prevalence equilibrium »).

Ils ont suggéré que les pays occidentaux se trouvent actuellement au stade de l'aggravation de la prévalence, se caractérisant par une stabilisation des incidences associée à une augmentation rapide de la prévalence. L'incidence des MICI est supérieure à la mortalité des personnes atteintes et l'espérance de vie augmente. L'hypothèse avancée est, qu'avec i) le plafonnement de l'incidence des MICI et ii) l'augmentation de la mortalité dans la population vieillissante atteinte de MICI, la prévalence des MICI commencera à se stabiliser et conduira au quatrième stade : l'équilibre de la prévalence.

Par contre, les nouveaux pays industrialisés sont dans la phase d'accélération de l'incidence et les pays en développement dans la phase d'émergence (Figures 4 et 5).

Au final, les MICI ont désormais un poids mondial important, notamment pour les systèmes de santé. Elles touchent plus de 6,8 millions d'individus dans le monde (34), avec des incidences et des prévalences différentes selon les régions du monde.

INCIDENCES DANS LES PAYS OCCIDENTAUX (EUROPE, AMERIQUE DU NORD, AUSTRALIE)

La perception actuelle est que les taux d'incidence dans les pays occidentaux, qui ont connu auparavant une phase d'augmentation importante, semblent maintenant se stabiliser, voire diminuer. Depuis les années 1990, les études épidémiologiques menées dans les pays occidentaux montrent en effet que les tendances d'incidence ont évolué : dans une revue de la littérature publiée en 2018, 16 (73 %) des 22 études sur la MC et 15 (83 %) des 18 études sur la RCH ont rapporté une incidence stable ou en baisse en Amérique du Nord et en Europe (35). Ces tendances à la baisse ont été confirmées dans des études plus récentes au Canada et en Suède (36,37).

Figure 4. Définition de 4 stades épidémiques.
D'après Kaplan et al, Nat Rev Gastroenterol Hepatol, 2021.

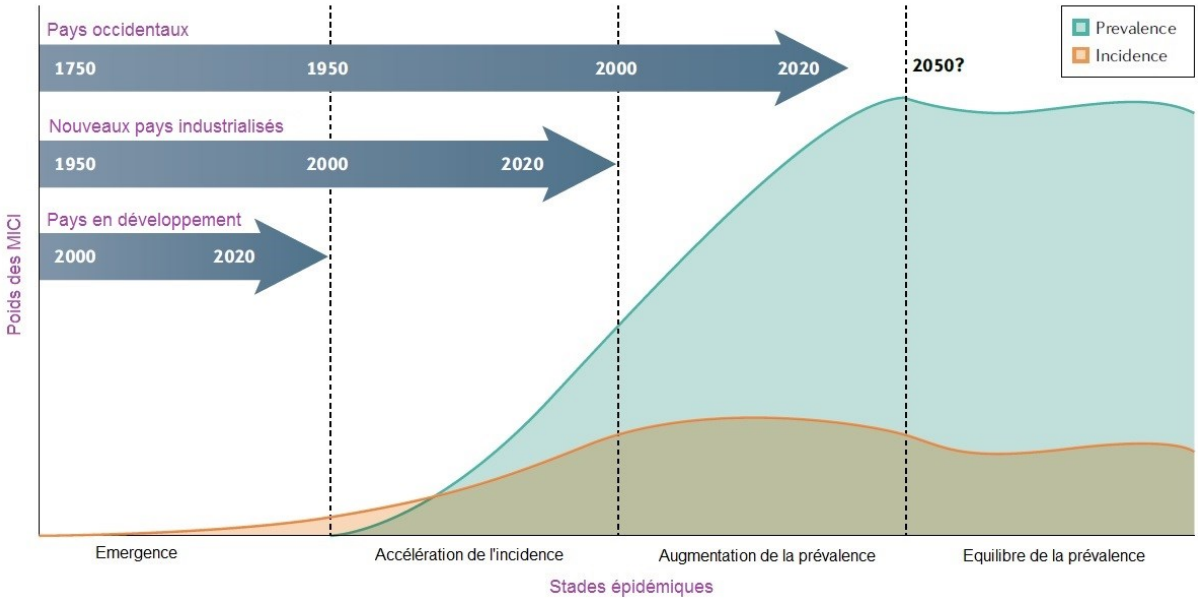
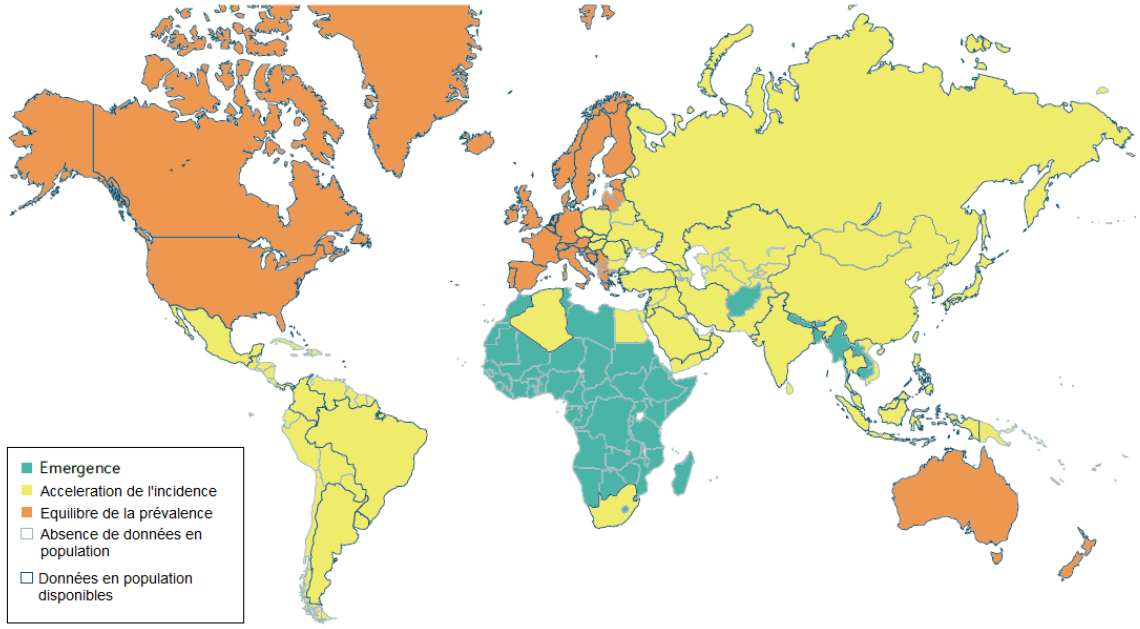


Figure 5. Répartition des pays selon les 4 stades épidémiques. *D'après Kaplan et al, NRGH, 2021.*



Amérique du Nord :

Aux Etats-Unis les données les plus fiables et les plus anciennes proviennent d'un registre en population dans le comté d'Olmsted (Minnesota) : les dernières estimations publiées montrent que l'incidence de la MC est passée de 6,9 à 10,7 pour 100 000 personnes-années

entre 1970 et 2010 ; l'incidence de la RCH a augmenté de 9,2 à 12,2/10^5 sur la même période (38).

Au Canada, l'incidence globale des MICI est stable sur la période 2007-2014 et est estimée à 30/10^5 en 2023 (MC : 12,2 ; RCH : 17,5). Les taux d'incidence et leurs tendances temporelles varient cependant considérablement d'une région à l'autre (39). L'incidence augmente de manière significative en Colombie-Britannique et au Québec, reste stable en Alberta et en Ontario, et diminue de manière significative au Manitoba et en Saskatchewan.

Australie et Nouvelle-Zélande :

Les incidences sont également élevées en Australie et en Nouvelle-Zélande. Une méta-analyse récente fait état d'une incidence de 23,5/10^5 en Australie (MC : 7,7 ; RCH : 9,4) et de 18,3/10^5 en Nouvelle-Zélande (MC : 9,2 ; RCH : 6,9) (40)[1].

Europe :

En Europe, les taux d'incidence varient fortement selon les pays. Les données les plus récentes indiquent que l'incidence de la MC varie de 0,4 à 22,8 pour 100 000 personnes-années, alors que l'incidence de la RCH est généralement plus élevée, se situant entre 2,4 et 44,0 pour 100 000 personnes-années (41). La plus forte incidence de la MC a été rapportée aux Pays-Bas (22,8/10^5), et l'incidence la plus basse en Moldavie (0,4/10^5) (42,43). Pour la RCH, l'incidence la plus élevée a été rapportée aux îles Féroé (44,0/10^5), tandis que l'incidence la plus basse a été observée en Roumanie (2,5/10^5)(42,44). Toutes MICI confondues, c'est aux îles Féroé que l'incidence est la plus forte atteignant, dans les estimations les plus récentes, 70/10^5 sur la période 2010-2020 et continue d'augmenter (45).

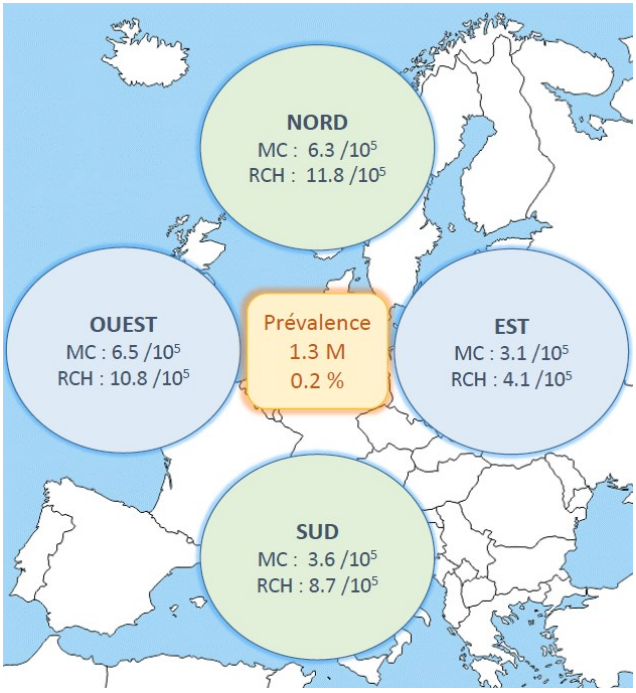
Ces données reflètent un gradient Ouest-Est également observé dans la cohorte européenne ECCO-EpiCom. Cette cohorte est une étude prospective et populationnelle de patients diagnostiqués en 2010 couvrant une population de base d'environ 10 millions de personnes dans 31 pays européens. Les taux d'incidence ont été comparés entre 14 pays d'Europe occidentale (6,3/10^5 pour la MC, 9,8/10^5 pour la RCH) et 8 pays d'Europe de l'Est (3,3/10^5 pour la MC, 4,6/10^5 pour la RCH)(46).

[1] La somme des incidences de MC et de RCH n'est pas égale au total pour l'ensemble des MICI car les estimations pour le total des MICI, la MC et la RCH ne sont pas basées sur le même nombre d'études.

Un gradient Nord-Sud a également été rapporté dans la cohorte EC-IBD (European Collaborative Study on Inflammatory Bowel Disease) des années 1990 (47). Cette cohorte incluait les patients diagnostiqués entre le 1er octobre 1991 et le 30 septembre 1993 dans 20 pays européens. Ce gradient semble persister malgré la stabilisation des incidences dans certains pays du Nord et la poursuite de l'augmentation de l'incidence dans certains pays du Sud comme en Espagne par exemple (48).

L'incidence continue également d'augmenter dans des pays ayant des incidences élevées comme aux iles Féroé et au Danemark. Au Danemark, entre 1995 et 2016, le taux d'incidence est passé de 9,1 à 17,8/10^5 pour la MC et de 21,0 à 28,4/10^5 pour la RCH (49). En Suède, par contre, l'incidence montrait une décroissance sur la période 2002-2014 (37). Des différences d'exposition à des facteurs environnementaux (tabac, alimentation…) pourraient expliquer ces différences d'évolutions entre des pays nordiques. Cependant, ces différences pourraient également être d'origine méthodologique selon l'origine des données (registres, données administratives) et la définition des cas.

Figure 6. Gradient Nord-Sud et Est-Ouest des incidences en Europe. *Données de Shivananda, Gut, 1996 pour le gradient Nord-Sud; Burisch, Gut, 2014 pour le gradient Ouest-Est et Zhao, JCC, 2021 pour l'incidence et la prévalence globales.*



France :

En France, les estimations d'incidence proviennent du registre Epimad, seul registre en population générale, couvrant 4 départements du Nord de la France (50–53). La dernière étude publiée portait sur l'évolution de l'incidence chez les enfants et les adolescents. De 1988-1990 à 2009-2011, une augmentation importante des incidences de la MC et de la RCH a été observée chez les adolescents (10 à 16 ans) : pour la MC, de 4,2 à 9,5/10^5 (+126 % ; P <0,001) et pour la RCH, de 1,6 à 4,1/10^5 (+156 % ; P <0,001)(54).

Tous âges confondus l'incidence était de 7,7/10^5 personnes-années pour la MC et de 4,4/10^5 personnes-années pour la RCH sur l'ensemble de la période 1988-2014 (55). Les incidences actualisées et détaillées par âge et sexe ainsi que leur évolution temporelle sont détaillées dans le premier chapitre de la thèse.

Il est à noter que la MC est plus fréquente que la RCH dans le Nord de la France alors que l'inverse est généralement observé, suggérant une émergence plus tardive de la RCH en France. La dernière méta-analyse en Nouvelle-Zélande faisait également état d'une plus forte incidence de la RCH que la MC (40). Ce ratio MC/RCH supérieur à 1 est également observé en Belgique (56).

Un gradient Nord-Sud de l'incidence de la MC, a été mis en évidence par Nerich et al. à partir de données de l'assurance maladie sur la prise en charge à 100 % entre 2000 et 2002 (données d'ALD) (57). La RCH présentait quant à elle une homogénéité spatiale sur l'ensemble du territoire. Plus récemment une hétérogénéité spatiale plus fine, à l'échelle du canton, de la prévalence de MC a également été mise en évidence à partir des données du PMSI : il n'a pas été mis en évidence de gradient Nord-Sud mais un total de 16 clusters de sur-incidence significatifs ont été identifiés, incluant un très large cluster au Nord et au Nord-Est de la France (Figure 7)(58).

L'hétérogénéité spatiale de la MC a également été étudiée à partir des données du registre Epimad et a permis de mettre en évidence 8 clusters de sur- et de sous-incidence sans la zone du registre (Figure 8) (55,59).

Figure 7. Hétérogénéité spatiale de la MC à l'échelle de la France à partir des données du PMSI. *Tiré de Génin et al., APT, 2020.*

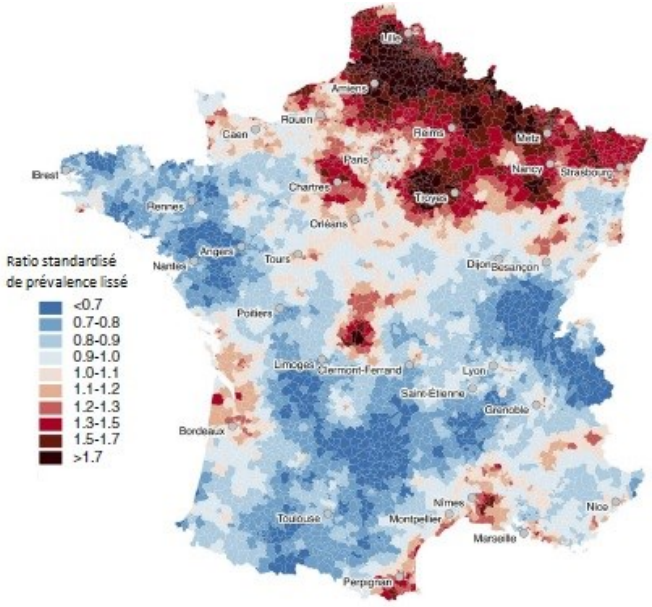
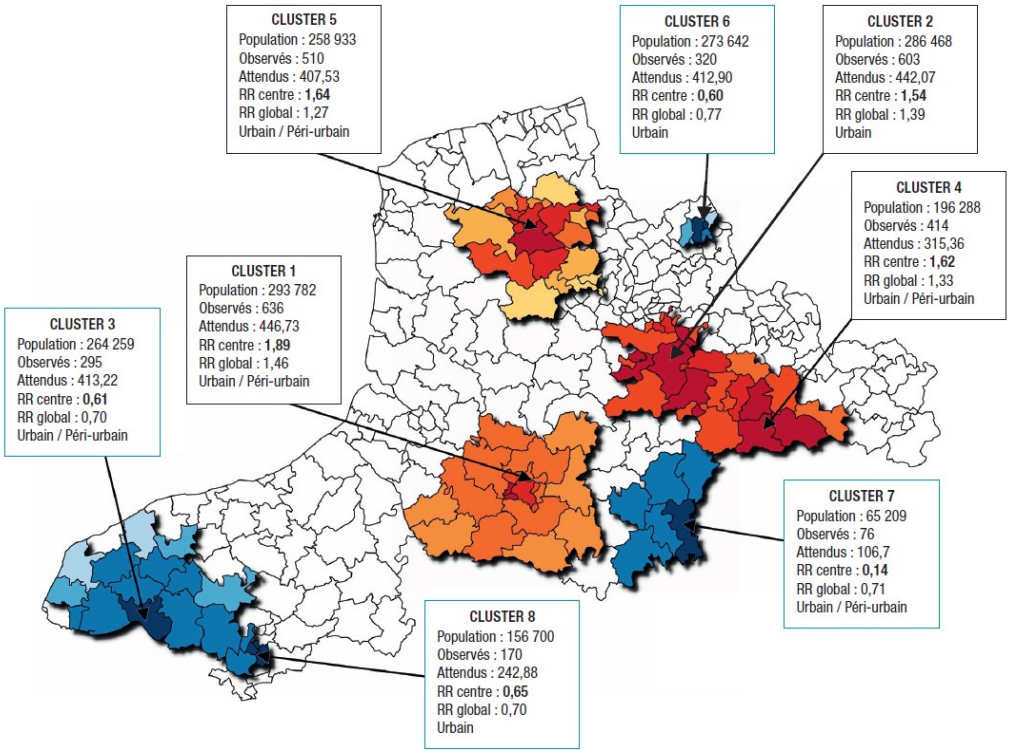


Figure 8. Hétérogénéité spatiale de la MC dans 4 départements du Nord-Ouest de la France à partir des données du registre Epimad. Tiré de Gower et al., BEH, 2019.



INCIDENCES DANS LES NOUVEAUX PAYS INDUSTRIALISES

Au cours du 20ème siècle, les MICI étaient rares en Asie, en Afrique et en Amérique latine, et les rares données disponibles montraient des valeurs nettement plus faibles par rapport aux pays occidentaux.

Au 21ème siècle, de nombreuses études épidémiologiques de meilleure qualité provenant de nouveaux pays industrialisés ont décrit une incidence des MICI en nette croissance (60,61). Par exemple, l'incidence de la MC et de la RCH ont augmenté respectivement de +11,1 % et +14,9 % par an au Brésil de 1988 à 2012 et de 4,0 % et 4,8 % par an à Taiwan de 1998 à 2008 (62,63).

PREVALENCE

Aux Etats-Unis, la prévalence a été estimée à 0,7 % de la population correspondant à un total d'environ 2,5 millions de patients vivant avec une MICI (64,65).

En 2023, plus de 320 000 Canadiens (0,8 %) seraient atteints d'une MICI. D'ici à 2035, la prévalence des MICI au Canada dépassera 1 % de la population, correspondant à environ 470 000 individus atteints (39,66).

En Europe, les estimations les plus récentes font état d'une prévalence des MICI de 0,2 % de la population soit 1,3 millions de cas prévalents avec de fortes variations en lien avec les disparités d'incidence relevées en Europe (41). Les prévalences les plus élevées concernent les pays du Nord de l'Europe avec, par exemple, des prévalences estimées à 0,8 % en Ecosse et en Norvège, 0,9 % au Danemark (67–69).

L'augmentation de la prévalence dans les pays occidentaux est associée au vieillissement de la population atteinte. En Ecosse par exemple, il a été estimé que la majorité des patients auraient plus de 50 ans en 2028 (67).

4. Spécificités des MICI à début pédiatrique

Les MICI peuvent apparaître à tout moment de la vie, de l'enfance (<17 ans) au sénior (>60 ans). Il existe des différences cliniques et phénotypiques importantes entre les MICI à début pédiatrique et les MICI de l'adulte (70).

INCIDENCE

Alors que l'incidence globale des MICI se stabilise dans la plupart des pays occidentaux, elle augmente chez l'enfant de manière importante dans la plupart des études (54,71,72). Dans une récente revue systématique de la littérature, 31 sur 37 (84 %) études ont rapporté des augmentations significatives de l'incidence, et toutes (7 sur 7) ont rapporté des augmentations significatives de la prévalence des MICI chez l'enfant (71). Comme chez l'adulte, les incidences les plus élevées sont observées au Canada, en Europe du Nord et en Nouvelle-Zélande.

RETARD DE CROISSANCE

Le retard de croissance est une complication de la MC spécifique des cas à début pédiatrique. Le retard de croissance peut même dans certains cas constituer le premier signe de la maladie. Dans une étude portant sur 50 patients atteints de MC, la réduction de la vitesse de croissance précédait le diagnostic pour 44 (88 %). Pour 21 cas (42 %), la réduction de la vitesse de croissance précédait les symptômes intestinaux (73). Dans une étude de patients issus du registre Epimad, 8 % des MC à début pédiatrique présentaient un retard de croissance au diagnostic et 5 % en développaient un au cours du suivi (délai médian : 4,9 ans) (74). Au final 29 % des patients présentaient une taille finale inférieure de 4 cm à leur taille cible. La vitesse de croissance était inversement corrélée à la CRP, mettant en évidence le rôle majeur de l'inflammation chronique dans le retard de croissance. Le retard de croissance est moins fréquent dans la RCH. La puberté est également souvent retardée quand la MC est diagnostiquée avant l'âge pubertaire. Ainsi, la prise en charge des MICI pédiatriques a pour spécificité d'avoir pour objectif majeur, en plus du contrôle de la maladie, de mener l'enfant à une taille finale normale. Contrairement à l'adulte, la nutrition entérale constitue un élément important du traitement permettant de limiter le retentissement de la maladie sur la croissance.

PRESENTATION DE LA MALADIE ET HISTOIRE NATURELLE

Les MICI à début pédiatrique présentent un tableau plus sévère dès le diagnostic ainsi qu'une évolution plus sévère (70,75,76). Chez l'enfant, la localisation iléocolique est la plus fréquente dans la MC au moment du diagnostic. De plus, une proportion importante, de l'ordre de 40 %, d'enfants présente une atteinte du tractus digestif haut (70). Une extension de la localisation de la maladie est rapportée chez jusqu'à un tiers des enfants au cours du suivi et un tiers des

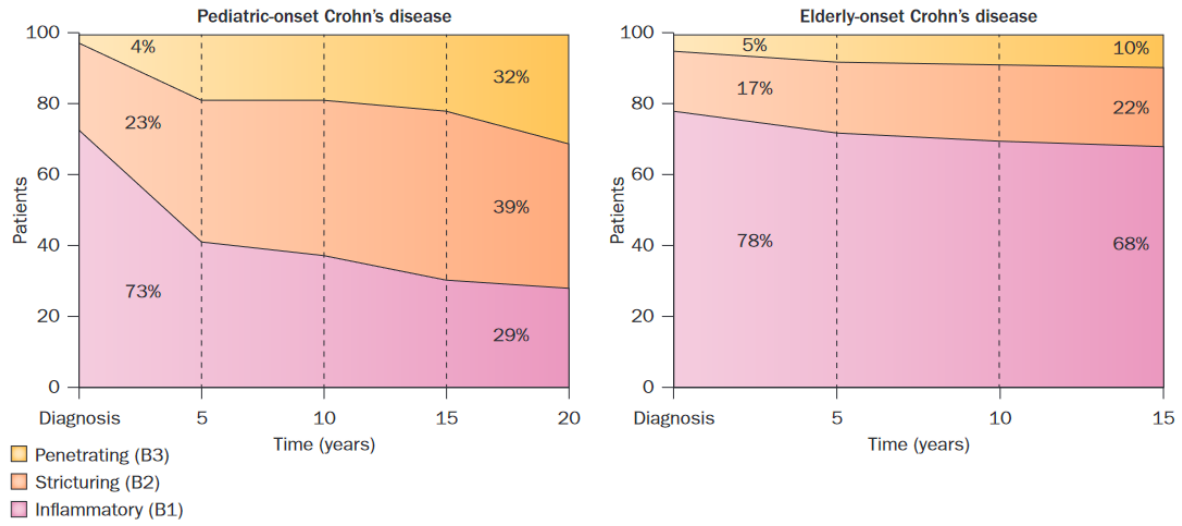
enfants a une MC évoluant vers un phénotype compliqué (B2 et/ou B3) dans les 10 ans suivant le diagnostic (76)(Figure 9).

Le risque d'extension de la RCH est également plus élevé dans les cas à début pédiatrique. La RCH s'étendait 10 ans après le diagnostic chez 30 % des enfants versus 20 % des adultes dans une méta-analyse (19). Le risque cumulé de chirurgie dans la RCH à début pédiatrique était de 15 % et 20 %, à respectivement 5 et 10 ans (75).

QUALITE DE VIE

La qualité de vie est significativement réduite chez les patients atteints de MICI, chez l'adulte comme chez l'enfant, notamment dans la MC (77). La qualité de vie est significativement associée à l'activité de la maladie (78). Les MICI à début pédiatrique ayant une morbidité plus élevée, la qualité de vie de ces jeunes patients est un sujet primordial. De plus, la maladie débutant tôt dans la vie, la perte de qualité de vie s'étend sur un plus long terme. La maladie affecte par ailleurs la qualité de vie à une période de changements et de construction de l'adolescent, ce qui pourrait avoir des conséquences sur sa vie à long terme.

Figure 9. Evolution du phénotype selon l'âge de début de maladie. Tiré de Ruel et al., Nat Rev Gastroenterol Hepatol, 2014.



5. Traitements

Leur cause étant inconnue, il n'existe pas de traitements curatifs des MICI. Le traitement des MICI implique des thérapies médicamenteuses ainsi que des interventions chirurgicales. Les

thérapies ont largement évolué ces dernières décennies, tant en nombre avec l'apparition de nouvelles molécules, qu'en termes de stratégies de traitement. Les classes de traitements utilisés sont : les 5-ASA (5-aminosalicylates) principalement utilisés dans le traitement de la RCH, les corticoïdes, les immunosuppresseurs utilisés comme traitement de fond dans les MICI à partir des années 1990, les biothérapies apparues au début des années 2000 (incluant les anti-TNF, les anti-interleukines et les anti-intégrines). De nouvelles molécules dites petites molécules, notamment les anti-JAK dont le tofacitinib déjà autorisé dans le traitement de la RCH, sont récemment apparues sur la scène du traitement des MICI et présentent l'avantage d'une administration par voie orale, d'une absence d'immunogénicité et d'une efficacité chez les patients en échec des anti-TNF.

Historiquement, l'objectif principal du traitement des MICI était le contrôle des symptômes, de préférence sans recourir aux corticoïdes, ainsi que la préservation de la croissance chez les enfants. Cette gestion conventionnelle de la MC consiste en l'utilisation séquentielle de corticostéroïdes, d'immunosuppresseurs et d'anti-TNF. Au fil du temps, il est apparu que se concentrer uniquement sur le contrôle des symptômes ne garantissait ni la guérison de l'inflammation intestinale ni la prévention de l'aggravation des lésions intestinales. Dans une analyse issue des données de l'essai SONIC, la moitié des patients atteints de MC traités par azathioprine et/ou infliximab en rémission clinique présentait une activité endoscopique de la maladie et/ou une CRP anormale (79). Dans la RCH, l'association entre les symptômes (notamment les rectorragies et le nombre de selles) et la cicatrisation muqueuse est plus élevée que dans la MC. Cependant, il est difficile d'atteindre une totale absence de symptômes (notamment le nombre de selles), même en cas de cicatrisation endoscopique (80). Chez l'enfant principalement, une nutrition entérale peut également être proposée au patient atteint de MC, en cas de dénutrition ou afin d'obtenir une rémission en cas de poussée sévère.

Par ailleurs, cette stratégie conventionnelle retarde l'utilisation des traitements efficaces pour les patients à haut risque de complication et risque de maintenir une exposition élevée aux corticoïdes. Il a été montré qu'une cicatrisation endoscopique précoce était associée à une évolution clinique favorable de la maladie, à la fois dans la MC et dans la RCH (81,82).

Pour répondre à cette problématique, les essais TOP-DOWN et SONIC ont montré la supériorité d'une combothérapie (utilisation combinée d'un immunosuppresseur et d'un anti-

TNF) précoce par rapport à la gestion conventionnelle chez les patients naïfs de traitement (83,84). Il a ensuite été montré qu'une intervention par biothérapie dans les 2 ans suivant le diagnostic de MC était associée à un taux plus faible de complications de la maladie et à des taux plus élevés de rémission clinique et endoscopique (85,86).

Enfin, désormais, l'approche "traitement ciblé" (treat-to-target, T2T) dont l'objectif est de définir des cibles de traitement à court, moyen et long termes en association à une évaluation régulière applique les principes de l'intervention précoce par biothérapie et du contrôle rigoureux de la maladie pour optimiser les résultats à long terme dans la MC mais aussi dans la RCH (87). Le consensus STRIDE-II a étendu les cibles de traitement, en plus de la rémission clinique et endoscopique, à la normalisation de la CRP et de la calprotectine fécale mais aussi à la restauration de la qualité de vie et la prévention du handicap fonctionnel à la fois dans la MC et la RCH (88).

En parallèle de cette augmentation de l'arsenal thérapeutique, le recours à la résection intestinale a fortement chuté ces dernières décennies (89,90). Une étude observationnelle récente menée à partir des bases de données danoises, a comparé le recours à la résection iléocæcale versus la prise d'anti-TNF précoces (dans la première année) (91). Le critère principal était le critère composite hospitalisation ou recours aux corticoïdes ou résection intestinale ou lésions anopérinéales. Le risque était plus faible dans le groupe de patients traités par résection (HR : 0,67 [0,54 ; 0,83]).

Avec d'avantage d'options thérapeutiques disponibles, les décisions de traitement deviennent plus complexes. De nombreux patients continuent à subir une évolution de la maladie invalidante et à perdre leur réponse au traitement au fil du temps. Par ailleurs, l'apparition de nouveaux traitements va de pair avec une augmentation des coûts et des effets secondaires, notamment infectieux. Il devient donc nécessaire de mieux connaitre les mécanismes des MICI et de développer des stratégies de traitement personnalisées. Bien que des études aient été réalisées dans ce sens, il est encore difficile d'appliquer ces traitements personnalisés dans les MICI (92,93).

6. Etiologie des MICI

L'étiologie des MICI reste à l'heure actuelle largement inconnue. L'hypothèse prédominante est que les MICI sont des maladies multifactorielles résultant de l'interaction complexe entre

des facteurs environnementaux, une susceptibilité génétique et un déséquilibre de la flore intestinale, résultant en une réponse immunitaire intestinale anormale (94).

6.1. Facteurs génétiques

La présence d'antécédents familiaux de MICI est le principal facteur de risque de MICI, suggérant l'existence d'une susceptibilité génétique des MICI. Des antécédents familiaux sont présents chez environ 8 à 12 % des patients atteints de MICI et le risque de développer une MICI chez les parents au premier degré d'un individu atteint est multiplié par 10 (95,96). La présence d'antécédents familiaux est plus marquée dans la MC que dans la RCH et la présence d'antécédents familiaux est associée à l'âge (ils sont plus fréquents chez les enfants). Dans une étude du registre Epimad menée dans la population pédiatrique, l'antécédent familial au 1er degré était le facteur de risque le plus important de développer une MICI notamment dans la fratrie (97).

Le domaine d'oligomérisation nucléotidique 2 (*NOD2*, également appelé *CARD15* et positionné sur le chromosome 16) a été le premier gène de susceptibilité à la MC identifié en 2001 (98,99). Les trois variants de *NOD2* les plus fréquents sont *R702W*, *G908R*, et *1007fs*. La moitié des patients atteints de MC sont porteurs d'au moins une mutation d'un variant à risque sur le gène NOD2 et 17 % sont porteurs d'une double mutation (100). Le portage d'un allèle de l'un des 3 variants principaux est associé à un risque de développer la MC multiplié par un facteur allant de 2 à 4. Ce risque est multiplié par 17 en cas de portage d'au moins deux allèles à risque (101). Depuis, plus de 200 variants génétiques associés aux MICI ont été identifiés, grâce à des études d'associations par « gènes candidats » puis des études GWAS (102–104). Les principaux variants identifiés incluent des gènes impliqués dans l'immunité innée (gènes *NOD2* et *CARD9*), dans le système immunitaire adaptatif (gène *IL23R*), dans l'autophagie (gènes *ATG16L1* et *IRMG*) ou dans le stress du réticulum endoplasmique (gène *XBP1*) (105–107). Les études ont révélé un degré important de chevauchement entre les variants génétiques de la RCH et de la MC, suggérant l'existence d'une voie similaire dans la pathogenèse des MICI (104).

Cependant la génétique est insuffisante pour expliquer le développement de la maladie. En effet, une discordance entre jumeaux monozygotes a été mise en évidence et montre que d'autres facteurs entrent en jeu dans le déclenchement des MICI (108). En effet, la

concordance entre les jumeaux monozygotes était de 6 % dans la RCH et de 58 % dans la MC. L'héritabilité serait donc plus faible dans la RCH que dans la MC.

6.2. Microbiote intestinal

Le microbiote intestinal est composé de 4 groupes bactériens majoritaires : les Firmicutes, les Bacteroidetes, les Actinobacteria et les Proteobacteria. Parmi ces bactéries, certaines ont un effet bénéfique (anti-inflammatoire) et d'autres un effet délétère (inflammatoire) dans la physiologie de l'intestin. Des études ont montré : i) une diminution de la diversité des bactéries du microbiote intestinal ; et ii) un déséquilibre entre les bactéries bénéfiques et les bactéries délétères, appelé dysbiose. Dans la MC, une augmentation de la proportion des entérobactéries dont certains *E. coli entéro-adhérents et invasifs* (AIEC), dont la colonisation pourrait être facilitée par la dysbiose et pouvant aggraver l'inflammation intestinale, a été décrite (109–111). L'origine de la dysbiose dans les MICI reste mal connue. Elle pourrait résulter de modifications alimentaires, être induite suite à des antibiothérapies fréquentes pendant l'enfance ou résulter d'une exposition insuffisante à des agents infectieux dans l'enfance (hypothèse de l'hygiène). Son rôle dans le déclenchement de la maladie reste à préciser.

6.3. Facteurs environnementaux

L'évolution temporelle des incidences des MICI ne peut pas s'expliquer uniquement par des variations génétiques sur un laps de temps aussi court. Il a également été observé que les populations migrantes des pays en développement ont un risque accru de développer des MICI lorsqu'elles s'installent dans les pays industrialisés, surtout pendant l'enfance, ce qui renforce l'existence de facteurs environnementaux dans la pathogénèse de la maladie (112,113). A l'inverse, il a été montré que le risque de RCH diminuait au fil du temps chez les individus des îles Féroé qui émigraient au Danemark et, par la suite, chez leurs enfants. Les individus de deuxième et troisième générations présentaient un taux d'incidence de RCH similaire à celui de la population danoise. Cela souligne l'influence des facteurs de risque environnementaux dans le développement des MICI, en particulier pour la RCH dont les taux sont particulièrement élevés et ont fortement augmenté au Danemark et aux Iles Féroé (114). Nous parlerons ici des facteurs environnementaux au sens large liés au mode de vie, à l'exposition à des traitements et à l'environnement physique et chimique.

FACTEURS ASSOCIES AUX EVENEMENTS DE SANTE

Il existe une association négative significative entre l'appendicectomie et la RCH (OR= 0,31 ; IC à 95 % = [0,25 ; 0,38]), correspondant à une réduction du risque de développer une RCH de 69 % pour les patients ayant subi une appendicectomie (115). Il a été montré que cette protection concernait uniquement les appendicectomies réalisées pour réelle appendicite et qu'elle ne concernait que les patients ayant eu une appendicectomie avant l'âge de 20 ans (116). Cet effet protecteur a donné lieu à un intérêt accru pour explorer le bénéfice thérapeutique d'une appendicectomie dans la RCH. Des études ont mis en évidence des taux de rechute et de colectomie plus faibles chez les patients ayant eu une appendicectomie (117). Un essai clinique est en cours (ACCURE) dont l'objectif principal est d'évaluer l'efficacité à court et moyen termes de l'appendicectomie dans le maintien de la rémission (118). Cependant, une élévation du taux de cancers colorectaux est également suspectée en lien avec l'appendicectomie (119,120). Une méta-analyse récente a identifié une réduction du taux de colectomie lorsque l'appendicectomie est pratiquée avant le diagnostic de RCH (OR : 0,76 [0,65 ; 0,89]), mais cette réduction n'est pas significative lorsque l'appendicectomie est réalisée après le diagnostic de la maladie. En revanche cette étude a identifié un intérêt thérapeutique potentiel chez les patients atteints de RCH réfractaire (120).

Des facteurs infectieux ont également été suspectés. En particulier, l'exposition à des espèces d'Helicobacter (non Helicobacter-pylori) favoriserait les MICI alors que l'exposition à Helicobacter pylori protégerait de la maladie (121).

Dans deux grandes cohortes prospectives de femmes américaines, l'utilisation de contraceptifs oraux était associée à un sur-risque de MC (HR : 2,82 [1,65 ; 4,82] pour la prise actuelle de contraceptifs et 1,39 [1,05 ; 1,85] pour l'utilisation passée de contraceptifs (122). Une étude récente a montré que les pilules purement progestatives n'étaient pas associées à la MC (123). Ces résultats corroborent l'hypothèse selon laquelle la composante œstrogénique de la contraception pourrait favoriser la pathogénie des MICI. L'association entre l'utilisation de contraceptifs oraux et la RCH était limitée aux femmes ayant des antécédents de tabagisme (122). Les auteurs attribuent ce résultat à un effet synergique du tabagisme et des contraceptifs oraux sur le risque de thrombose subaiguë augmentant le risque de développer une RCH.

Enfin, l'exposition à des antibiotiques a également été associée à un risque accru de MICI, en particulier lors de la prise de plus de 3 antibiotiques dans les 2 à 5 ans précédant le diagnostic de MICI, avec un effet dose (124). Comme évoqué plus haut, la prise d'antibiotique pourrait être associée à une dysbiose entrainant une inflammation de l'intestin, en particulier dans l'enfance. Des études ont établi un risque associé à l'exposition aux antibiotiques dans l'enfance (125–127). Par ailleurs, la prise de plus de 3 antibiotiques pendant la grossesse exposait le nouveau-né à un risque ultérieur de RCH accru de 45 % (HR : 1,45 [1,06 ; 2,00]) (128).

MODE DE VIE

Le tabagisme est l'un des principaux facteurs environnementaux connus avec des effets opposés selon le type de MICI. En effet, une méta-analyse a établi une association entre le tabagisme actuel et la MC (OR : 1,76 [1,40 ; 2,22]) et entre le tabagisme passé et la RCH (OR : 1,79 [1,37 ; 2,34]. Par contre, le tabagisme actif était protecteur de RCH (OR : 0,58 [0,45 ; 0,75]) (129). Une étude prospective a montré que le risque de RCH était significativement augmenté dans les 2 à 5 ans suivant l'arrêt du tabac (HR : 3,06 [2,00 ; 4,67) et restait élevé pendant les 20 ans suivant l'arrêt (130).

L'allaitement maternel serait protecteur à la fois de la MC (OR : 0,71 [0,59 ; 0,85]) et de la RCH (OR : 0,78 [0,67 ; 0,91)] avec un effet dose selon la durée de l'allaitement (131).

Parmi les autres facteurs environnementaux, les facteurs alimentaires, pouvant affecter l'inflammation intestinale et modifier le microbiote, en particulier par l'adoption d'un régime alimentaire occidentalisé, ont été incriminés (132). Dans l'étude Européenne EPIC (European Prospective Investigation into Cancer and Nutrition), le risque de développer une MC était plus faible chez les personnes consommant des proportions élevées d'aliments non transformés/peu transformés (133). Une méta-analyse récente a établi un risque accru de développement de la MC en lien avec une consommation plus élevée de produits ultra-transformés (HR : 1,71 [1,37-2,14]) et un risque plus faible de MC lors d'une consommation plus élevée d'aliments non transformés/peu transformés (HR : 0,71 [0,53-0,94]) (134). Ces associations n'étaient pas observées dans la RCH. Une étude interventionnelle chez des sujets sains a montré que la consommation de carboxymethylcellulose (CMC), un émulsifiant synthétique, était associée à une réduction de la diversité du microbiote et des modifications du métabolome fécal (135).

L'adoption d'un régime dit « méditerranéen », caractérisé par une forte consommation de fruits, de légumes, de céréales complètes, de légumineuses, de noix, d'huile d'olive et une consommation modérée de produits animaux et ultra-transformés, a montré des effets protecteurs potentiels dans la MC mais pas dans la RCH (136). Une étude récente a montré que ce régime alimentaire était associé à la composition du microbiote avec une augmentation de l'abondance des bactéries dégradant les fibres, telles que *Ruminococcus*, et à une inflammation intestinale réduite, prouvée par une calprotectine fécale faible (137).

Dans l'étude E3N (Etude Epidémiologique auprès de femmes de la Mutuelle Générale de l'Education Nationale), une consommation élevée de protéines, en particulier de protéines animales, était associée à un risque significativement accru de MICI (RR : 3,03 [1,45-6,34] pour la comparaison des 3ème versus 1er tertiles de consommation de protéines animales). Parmi les sources de protéines animales, une consommation élevée de viande ou de poisson était associée à un sur-risque de MICI (138). Une méta-analyse récente a mis en évidence une relation dose-effet : une augmentation de 100 g/jour de la consommation totale de viande était associée à un risque accru de 38 % de MICI (139).

Enfin la consommation de fibres, en particulier de fruits, était associée à une réduction du risque de MC mais pas de RCH dans une étude de la Nurse's Health Study (HR : 0.50 [0,39 ; 0,90] pour la comparaison des 5ème versus 1er quintiles de consommation) (140). Cependant, ce résultat n'a pas été confirmé dans la cohorte EPIC (141).

La consommation d'alcool ne semble pas associée au risque de MICI (142). En ce qui concerne les micronutriments, un effet protecteur de la vitamine D dans la MC ainsi qu'un effet protecteur du Zinc dans la MC mais pas dans la RCH ont été mis en évidence (143,144). Dans la cohorte EPIC, une consommation alimentaire plus élevée de resvératrol et de flavones, des polyphénols aux propriétés anti-oxydantes, était associée à un risque réduit de MC (145).

D'autres facteurs comme le stress et le manque d'activité physique ont également été rapportés (146).

Dernièrement, les données provenant de la biobanque « UK Biobank » ont montré qu'au sein de la population présentant un risque génétique élevé de MICI à l'âge adulte, ceux qui adoptent un mode de vie « favorable » peuvent réduire leur risque de moitié (147). Le risque génétique était évalué à partir d'un score de risque polygénique établi à partir de 51 et 30

variants pour la MC et la RCH, respectivement. Le score du mode de vie a été catégorisé comme favorable, intermédiaire et défavorable à partir de 6 facteurs de risque potentiellement modifiables et associés aux MICI : le tabagisme, l'alimentation, l'activité physique, l'obésité et la durée de sommeil.

Dans une étude portant sur six cohortes en Europe et aux États-Unis, il a été estimé que plus de 40 % des cas de MICI auraient pu être évités grâce à des changements dans le mode de vie et les facteurs alimentaires (148).

ENVIRONNEMENT PHYSIQUE ET CHIMIQUE

Les MICI ayant émergé au moment de l'industrialisation et augmentant avec l'adoption d'un mode de vie occidental, l'implication de facteurs physiques et chimiques de l'environnement est suspectée. Cependant, les études sont rares. Une récente revue de la littérature a identifié seulement 39 études sur le sujet dans la MC (149).

Concernant la pollution de l'air, une étude a mis en évidence un risque de MICI à début pédiatrique augmenté en lien avec une exposition dans l'enfance aux polluants aux capacités oxydantes (combinaison de NO_2 et O_3) (150). La même équipe a mis en évidence un effet protecteur des espaces verts sur le développement de MICI dans l'enfance, avec un effet-dose (151). Cet effet était retrouvé dans une étude réalisée à partir des données de la biobanque « UK Biobank ». Cette dernière étude a également mis en évidence un effet protecteur de la proximité de plans d'eau (152). L'association était renforcée dans les zones plus précaires. En France une étude à partir des données du PMSI a identifié un risque plus élevé de MC en lien avec un degré plus élevé d'urbanisation (58). Dans une méta-analyse les rapports des taux d'incidences comparant l'environnement urbain à l'environnement rural étaient estimés à 1,17 [1,03 ; 1,32] et 1,42 [1,26 ; 1,60] pour la MC et la RCH respectivement (153). Enfin, l'exposition à des métaux lourds (Plomb, cuivre, zinc et chrome) pendant la grossesse ou le début de vie a été identifiée comme facteur de risque de MICI dans une étude pilote portant sur l'analyse de l'exposition mesurée dans les dents de 12 patients atteints de MICI versus 16 témoins (154). L'exposition aux per- et polyfluoroalkylées (PFAS), utilisés dans de nombreux domaines depuis les années 1950 et persistant dans l'environnement, ont été étudiés dans 2 études mais avec des résultats contrastés, possiblement liés à des niveaux d'exposition variant fortement d'une étude à l'autre (155–159).

Enfin l'exposition aux pesticides est suspectée. Une étude récente menée auprès d'agriculteurs et de leurs épouses a mis en évidence un risque accru de MICI (HR > 1,2 pour la comparaison des utilisateurs versus non utilisateurs de 5 pesticides) (160). L'exposition en début de vie à des zones agricoles était associé à un risque accru de MC dans une étude écologique danoise récente (161).

Les registres épidémiologiques

1. La surveillance épidémiologique

L'épidémiologie est une discipline scientifique dont l'objectif est d'étudier les questions de santé au niveau de la population : fréquence, variation dans le temps, étude des facteurs influençant la santé.

Selon l'Organisation Mondiale de la Santé (OMS) la surveillance épidémiologique « *s'entend de la collecte, de la compilation et de l'analyse systématiques et continues de données à des fins de santé publique et de la diffusion d'informations de santé publique en temps voulu à des fins d'évaluation et aux fins d'une action de santé publique, selon les besoins* » (162). La surveillance était, au départ, plutôt consacrée à des maladies transmissibles afin d'en contrôler la diffusion, puis son champ s'est étendu à tout évènement de santé, des infections nosocomiales en milieu hospitalier, aux maladies chroniques, aux maladies professionnelles etc… En France, le réseau « Sentinelles » surveille depuis 1984 en continu différents indicateurs de santé, notamment les épidémies hivernales de grippes et de gastroentérites. Le principal acteur de la surveillance épidémiologique est l'agence Santé Publique France, mais de nombreux autres acteurs de la santé publique interviennent dans le champ de la surveillance, notamment les registres.

2. Définition des registres

Un registre épidémiologique est défini comme : « *un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs événements de santé dans une population géographiquement définie, à des fins de recherche et de santé publique, par une équipe ayant les compétences appropriées* », selon l'arrêté du 6 novembre 1995 relatif au Comité National des Registres. Depuis 2013, un Comité d'Evaluation des Registres a été mis en place par Santé Publique France, l'Inserm et l'INCA. Ses missions sont : i) évaluer les registres en prenant en considération à la fois leurs missions de recherche et de santé publique ; ii) émettre des recommandations sur le fonctionnement et les activités de recherche et de surveillance du registre évalué ; iii) évaluer la mise en œuvre des recommandations préalables ; et iv) faire

des propositions au Comité stratégique des registres en matière de besoins de registres, notamment au regard de l'épidémiologie et de la politique de prévention et de prise en charge.

Ainsi, l'objectif d'un registre est de constituer une base de données exhaustive dans laquelle sont recueillies des informations détaillées sur les cas incidents d'une maladie spécifique au sein d'une population bien définie dans une zone géographique délimitée. La zone géographique est définie comme la zone d'habitation des patients au moment du diagnostic de la maladie. Les registres épidémiologiques sont principalement utilisés pour étudier l'incidence des maladies, ainsi que pour suivre leur évolution temporelle mais les registres sont également le support de nombreuses études analytiques ayant pour but l'étude des facteurs de risque, de l'histoire naturelle etc.

L'utilisation des registres épidémiologiques est cruciale pour la recherche en santé publique, car ils permettent aux épidémiologistes et aux chercheurs de surveiller l'incidence des maladies, d'identifier des tendances, de comprendre les facteurs de risque et, le cas échéant, d'évaluer l'efficacité des interventions de santé publique. Selon le type de pathologies suivies, les données recueillies dans ces registres peuvent être utilisées pour informer les politiques de santé et élaborer des stratégies de prévention et de contrôle des maladies. Ils permettent par l'estimation des incidences et des prévalences, d'évaluer le poids des maladies dans le système de santé et d'anticiper les évolutions futures afin d'adapter le système de soins aux évolutions du nombre de patients et des coûts associés.

Les registres se distinguent notamment des autres sources de données (cas hospitaliers, données médico-administratives) par le fait qu'ils couvrent l'ensemble des cas (et non uniquement des cas hospitaliers par exemple) avec une définition stricte de la maladie et une expertise médicale de chaque cas.

3. Le registre Epimad

3.1. Contexte et objectifs

Comme cela a été décrit dans la première partie de cette introduction, les MICI sont des maladies inflammatoires intestinales chroniques, touchant des sujets jeunes et ayant une morbidité élevée. L'étiologie est à ce jour toujours largement méconnue. Leur fréquence

élevée dans le Nord de l'Europe en fait un réel problème de santé publique ayant justifié la mise en place d'un registre dans le Nord-Ouest de la France en 1988, à l'initiative des gastroentérologues et épidémiologistes des CHU d'Amiens, Lille et Rouen.

Le registre Epimad a été reconnu par la Commission Nationale des Registres en 1992. Cette qualification a été renouvelée en 1996, 2000, 2004, 2008, 2012, 2016 et 2021 avec un soutien officiel de l'Inserm et de l'Institut National de Veille Sanitaire, puis de Santé Publique France. Le dernier renouvellement quinquennal du Registre a été obtenu le 1er décembre 2021 lors de l'assemblée du Comité d'Evaluation des Registres avec l'obtention de la note maximale « A » pour les 3 items évalués et un renouvellement pour 5 ans. Le registre bénéficie des autorisations du CCTIRS (Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé) et de la CNIL (Commission Nationale de l'Informatique et des Libertés) (autorisation n°917089).

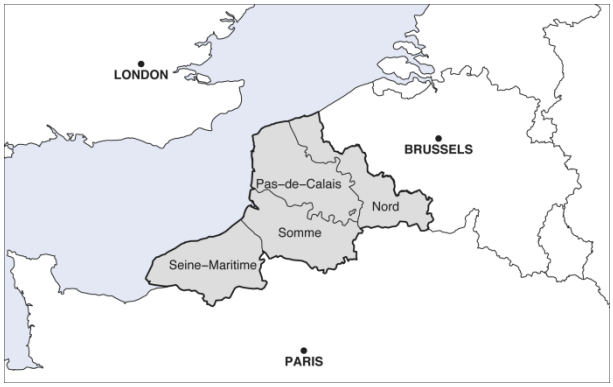
L'objectif principal du registre Epimad est de calculer l'incidence de MICI, c'est à-dire le nombre de nouveaux cas pour une période de temps donnée rapporté à la population totale dans la zone concernée, et d'en étudier les évolutions spatio-temporelles.

Les objectifs secondaires sont de mener des études d'épidémiologie analytique en population générale (sur les facteurs de risque de MICI, les traitements, l'histoire naturelle etc). Pour ces objectifs secondaires, le caractère populationnel de l'étude est fondamental, permettant de se distinguer des études faites uniquement à l'hôpital dans des centres de référence et ne reflétant pas l'entièreté de la réalité des MICI puisque près de 80 % des cas sont diagnostiqués par les gastroentérologues en ville.

3.2. Population

Le registre Epimad couvre les départements du Nord, du Pas-de-Calais, de la Somme et de la Seine-Maritime, correspondant à une population totale de 5 887 559 habitants soit 9 % de la population française (source : INSEE, estimation au 1er janvier 2024). Tous les cas demeurant dans cette zone au moment du diagnostic de la maladie sont inclus.

Figure 10. Zone géographique couverte par le registre Epimad.



3.3. Recueil des données

3.3.1. Méthodologie

Le recueil des données repose sur la participation active de l'ensemble des gastroentérologues adultes et pédiatriques de la zone (soit environ 250 praticiens, source principale des données), quel que soit leur mode d'exercice en secteur libéral, public ou mixte. Ces gastroentérologues déclarent régulièrement au registre les nouveaux cas vus en consultation ayant des symptômes évocateurs d'une MICI. Des enquêteurs du registre se rendent régulièrement (en moyenne 3 fois par an) dans les cabinets des gastroentérologues et dans les services hospitaliers afin de recueillir les informations dans les dossiers médicaux des patients.

Afin de s'assurer de l'exhaustivité du recueil, le registre utilise également deux sources secondaires de données :

- Depuis 2014, les données du PMSI (Programme de Médicalisation des Systèmes d'Information) des départements d'information médicale (DIM) des centres hospitaliers sont croisées avec les données du registre afin de récupérer certains cas qui n'auraient pas été déclarés. Ce recueil a été initié avec les données du DIM de Lille puis a été étendu petit-à-petit aux autres DIM des hôpitaux publics. L'inclusion des DIM de certaines cliniques privées a débuté et sera étendue dans les années qui viennent.
- Dans le département de la Somme, les 3 laboratoires d'anatomie pathologique transmettent leurs données aux enquêteurs du registre depuis 1988.

3.3.2. Données recueillies

Les enquêtrices du registre recueillent les données sur un questionnaire standardisé contenant environ 500 items.

Les données recueillies concernent : l'âge, le sexe, la date du diagnostic, les données cliniques, radiologiques, endoscopiques et histologiques ayant permis de faire le diagnostic.

Les variables cliniques recueillies incluent :

- Le délai entre les premiers symptômes et le diagnostic évoqué (délai diagnostique).
- Le décès éventuel à la première poussée.
- Les antécédents personnels (abcès/fistule/fissure anales, tuberculose, appendicectomie, tabac, prise de traitements notamment d'antibiotiques et d'anti-inflammatoires non-stéroïdiens).
- Les antécédents familiaux de MICI.
- Les symptômes cliniques dans les 6 semaines précédant la date du diagnostic (transit, sang et glaires dans les selles, douleurs abdominales, syndrome rectal, hyperthermie, signes extra-intestinaux (œil, peau, bouche, articulations, foie et voies biliaires, pancréas), masse abdominale, pathologie anale au moment de l'examen, autres symptômes d'entrée dans la maladie (syndrome occlusif, colectasie, …), morbidité associée (grossesse, complications thromboemboliques, maladies auto-immunes, etc).

Lorsqu'ils sont disponibles dans les 6 semaines avant la prise en charge et dans les 3 mois suivants, les résultats biologiques sont également recueillis. Ils incluent :

- Les ASCA (anticorps anti-*Saccharomyces cerevisiae*) et les ANCA (anticorps anti-cytoplasme des polynucléaires), marqueurs sérologiques des MICI.
- La CRP, la calprotectine fécale.
- L'hémoglobine, les plaquettes, l'albumine.
- Les résultats de coproculture, la recherche de toxines.

Les explorations digestives réalisées dans les 6 semaines précédant la prise en charge et dans les 3 mois suivants sont recueillies :

- Type d'examens : iléo/coloscopie, gastroscopie, échographie, scanner/enteroscanner, IRM pélvienne, entero-IRM, videocapsule, enteroscopie.

- Le résultat (normal/anormal) est noté pour chaque segment du tube digestif (Anus/périnée, Rectum, Sigmoïde, colon gauche, colon transverse, Colon droit, Caecum, Iléon, Jéjunum, Duodénum, Estomac, Œsophage).

Les résultats des biopsies réalisées dans les 6 semaines précédant la prise en charge et dans les 3 mois suivants sont recueillis et résumés selon chaque segment du tube digestif :

- Histologie selon chaque segment du tube digestif (muqueuse anormale, muqueuse normale, non réalisé, inconnu)
- Caractéristiques histologiques (Granulome épithélioïde et giganto-cellulaire, muqueuse inflammatoire, diminution de la muco-sécrétion, infiltrat inflammatoire spécifique, infiltrat inflammatoire non spécifique, perte de substance/décollement, abcès cryptique)

Les données issues de pièces d'exérèse sont également recueillies, lorsqu'elles existent (Granulome épithélioïde et giganto-cellulaire, fissure, fistule, nodule lymphoïde, diminution de la muco-secrétion, sclérose).

Pour chaque segment, la présence de lésions macroscopiques et ou microscopiques est décrite. Le phénotype de la MC (inflammatoire, sténosant, pénétrant), la présence de lésion(s) segmentaire(s) suspendue(s) sur 1 même organe, de lésions bi- ou pluri-focales sur un même organe, de lésions de plus de 15 cm sur l'intestin grêle et sur le colon sont notés.

Les traitements prescrits dans les 3 mois suivant le diagnostic sont notés (5-ASA, corticoïdes, antibiotiques, immunosuppresseurs, biothérapies, nutrition artificielle).

3.3.3. Validation des données et critères diagnostiques

Les questionnaires sont relus et validés en double aveugle par deux experts gastroentérologues dans chacun des 3 centres (Lille, Amiens, Rouen) qui posent ensuite un diagnostic parmi 14 diagnostics possibles (Table 3).

Table 3. Liste des diagnostics posés après expertise des dossiers Epimad.

1. RCH certaine	8. Proctite probable
2. RCH probable	9. Proctite possible
3. RCH possible	10. Colite chronique inclassable
4. Crohn certain	11. Colite aigüe non spécifique
5. Crohn probable	12. Autre diagnostic
6. Crohn possible	13. Colite aigüe
7. Proctite certaine	14. Cas inclassé

Un diagnostic final de MC, de RCH ou de proctite est posé et rapporté soit sous forme certaine, probable ou possible. Certains dossiers sont classés en colite aiguë si le délai entre le début des symptômes et le diagnostic est inférieur à 6 semaines et en l'absence de granulome avec cellules épithélioïdes et giganto-cellulaires aux biopsies ou sur la pièce opératoire et en l'absence de lésion de l'intestin grêle.

Les dossiers pour lesquels le diagnostic de MICI est probable, mais sans argument permettant de différencier une MC d'une RCH, sont classés « colite chronique inclassable (CI) ».

Seuls les cas certains et probables ainsi que les colites chroniques inclassables (CI) sont considérés comme d'authentiques MICI pour la suite des analyses.

Les cas possibles et les colites aigües font l'objet d'un suivi afin de permettre leur reclassement éventuel en MICI certaine ou probable. Ce suivi est réalisé tous les 2 ans pendant 10 ans. Ainsi, les incidences sont revues régulièrement et peuvent varier en fonction du classement final de ces cas en authentiques MICI ou non.

Les patients ayant une coproculture positive, ayant consommé des antibiotiques et/ou des anti-inflammatoires non stéroïdiens dans le mois précédant le début des symptômes ne sont considérés par le registre que s'il y a persistance des symptômes au moins 6 semaines après le traitement de l'infection et l'arrêt des antibiotiques et/ou des anti-inflammatoires non stéroïdiens.

Les critères suivants sont utilisés (53) :

MALADIE DE CROHN

Maladie de Crohn certaine :

Présence d'un granulome avec cellules épithélioïdes et giganto-cellulaires sur des biopsies ou des spécimens chirurgicaux.

Maladie de Crohn probable :

1) Lésions du côlon sans atteinte de l'intestin grêle et présence d'au moins 3 des 4 critères suivants :

a. Histoire clinique de diarrhée et/ou de douleurs abdominales depuis plus de 6 semaines.

b. Aspect radiologique et/ou endoscopique évocateur d'une MC, avec des lésions segmentaires et/ou une sténose inflammatoire colique.

c. Aspect histologique compatible avec une MC.

d. Existence de fistule et/ou d'abcès en relation avec la maladie inflammatoire digestive.

2) Lésions de l'intestin grêle avec ou sans atteinte colique, quelle que soit la durée des symptômes cliniques, et la présence d'au moins 2 des 4 critères ci-dessus.

Maladie de Crohn possible :

1) Lésions du côlon sans atteinte de l'intestin grêle et présence de 2 des 4 critères ci-dessus, incluant une histoire clinique évoluant depuis plus de 6 semaines.

2) Lésions de l'intestin grêle avec ou sans atteinte colique, quelle que soit la durée d'évolution des symptômes cliniques et la présence de 1 des 4 critères ci-dessus.

RECTOCOLITE HEMORRAGIQUE

Rectocolite hémorragique certaine :

1) Histoire clinique de diarrhée et/ou de rectorragies évoluant depuis plus de 6 semaines et au moins 2 des 3 critères suivants :

a- Aspect endoscopique typique incluant une muqueuse friable, granuleuse ou ulcérée, ou les deux, au niveau de la muqueuse en surface.

b- Aspect radiologique typique incluant des ulcérations et/ou des sténoses du côlon.

c- Aspect histologique compatible avec une RCH sur des biopsies ou des spécimens chirurgicaux nécrotiques.

2) Spécimens chirurgicaux ou nécrotiques macroscopiquement typiques de RCH avec une histoire clinique également typique.

Rectocolite hémorragique probable :

1) Histoire clinique de diarrhée et/ou de rectorragies évoluant depuis plus de 6 semaines et 1 des 3 critères ci-dessus.
2) Histoire clinique évoluant depuis plus de 6 semaines mais sans diarrhée ni rectorragies et 2 des 3 critères ci-dessus.
3) Spécimens chirurgicaux ou nécrotiques macroscopiquement typiques de RCH, mais sans aspect histologique typique.

Rectocolite hémorragique possible :

Histoire clinique typique évoluant depuis plus de 6 semaines, mais sans aspect morphologique ni histologique compatible avec le diagnostic.
Les proctites ulcérées (localisation uniquement rectale de la RCH) certaines, probables et possibles sont définies avec les mêmes critères que ceux de la RCH, mais avec à l'évidence un aspect macroscopiquement normal du sigmoïde sus-jacent.
La jonction recto-sigmoïdienne a été arbitrairement estimée à 20 cm de la marge anale.

COLITE CHRONIQUE INCLASSABLE (CI)

Les patients ayant un tableau clinique de colite chronique compatible avec l'un ou l'autre des diagnostics de MC ou de RCH sont classés en colite chronique inclassable.

Les patients ayant une histoire clinique évoluant depuis moins de 6 semaines (sans granulome, ni lésions de l'intestin grêle) sont classés en colite aiguë.

3.3.4. Saisie des données

Depuis août 2016, les données sont saisies via une application développée par la société Epiconcept, certifiée “hébergeur de données de santé à caractère personnel" (HDS) et membre de l'association française des hébergeurs de données de santé à caractère personnel (AFHADS). Elle respecte les règles de conformité induites par le règlement général sur la protection des données (RGPD).

Chaque enquêteur effectue la saisie à partir d'un ordinateur relié à internet, à partir duquel il s'authentifie sur l'application dédiée au Registre Epimad (authentification des accès par CPS (Carte de Professionnel de Santé) ou CPE (Carte de Professionnel d'Etablissement), ou par mot de passe généré automatiquement). La saisie des éléments d'identification d'une part et du questionnaire d'incidence d'autre part (reliés par le numéro de dossier) se fait avec chiffrement des données. Deux bases sont créées, l'une comportant les identifiants et numéros de dossiers (numéro d'ordre), l'autre comportant les données médicales et le numéro de dossier.

Le stockage des données est sécurisé (gestion par la société Epiconcept), sur un serveur répondant aux qualités requises pour l'hébergement de données de santé à caractère personnel.

3.4. Analyses statistiques

La gestion de la base de données et la réalisation des analyses statistiques sont effectuées par la biostatisticienne du registre Epimad et les MCU-PH en épidémiologie.

Les calculs d'incidence sont effectués sur des périodes de 3 années consécutives, afin d'améliorer la précision et la robustesse des estimations fournies. Ils ne prennent en compte que les diagnostics de MC, RCH certains ou probables, ainsi que les colites chroniques inclassables (CI). Afin de pouvoir comparer les données du Registre avec celles des autres études parues dans la littérature, les formes possibles et les formes aiguës ne sont pas incluses dans le calcul des taux d'incidence.

Les calculs d'incidence sont standardisés sur l'âge par méthode directe (par groupe d'âge quinquennal) sur la population type européenne (poids révisés en 2013 par Eurostat) (163). Les intervalles de confiance à 95 % sont estimés en utilisant une distribution gamma (164).

Les analyses statistiques diffèrent ensuite selon les études. Les méthodes statistiques seront spécifiquement détaillées dans chaque chapitre.

3.5. Etudes analytiques et cohortes nichées dans le registre

Le registre Epimad, outre les études sur l'incidence des MICI et leur phénotype au diagnostic, s'implique dans des recherches analytiques qui lui sont propres ainsi que dans des études collaboratives nationales et internationales.

Le registre a ainsi développé depuis de nombreuses années deux cohortes spécifiques : une cohorte de patients pédiatriques (diagnostic avant l'âge de 17 ans) et une cohorte de patients séniors (diagnostic après l'âge de 60 ans). Ces cohortes n'impliquent pas directement la participation du patient : les données de suivi sont collectées dans les dossiers médicaux des patients. La cohorte pédiatrique a été réactualisée (étude Inspired). L'objectif de cette étude, qui comprenait 1 344 patients (1 007 MC, 337 RCH) diagnostiqués entre 1988 et 2011 et suivis jusqu'en 2013, était d'étudier l'impact des stratégies thérapeutiques sur l'histoire naturelle de la maladie, notamment sur le risque de chirurgie intestinale et le développement de complications notamment cancéreuses (89,90,165–167).

Le registre mène également des études observationnelles et interventionnelles dans lesquelles les patients sont directement impliqués pour des prélèvements (sang, selles …) et/ou des questionnaires spécifiques.

Au niveau Européen, le registre est impliqué dans la cohorte européenne ECCO-EPICOM-2010 (168,169) : les cas incidents diagnostiqués en 2010 dans la Somme ont été inclus dans cette cohorte et font depuis 2014 l'objet d'un suivi régulier.

Objectifs et plan de la thèse

Le registre Epimad a été, depuis sa mise en œuvre en 1988, le support de nombreux travaux sur l'incidence et l'histoire naturelle des MICI, permettant une amélioration des connaissances par l'apport de ces données en population générale.

L'objectif de cette thèse est d'apporter des connaissances sur les MICI, selon 2 axes définis ci-dessous, à partir des données du registre Epimad (Figure 11).

- **Le premier axe** s'intéresse à l'épidémiologie des MICI.

Un premier chapitre présentera les incidences et prévalences des MICI, et plus particulièrement leurs évolutions temporelles. La connaissance de l'incidence et de la prévalence des MICI, et de leur évolution temporelle, permet d'en estimer le poids actuel et futur dans le système de soins et ainsi d'anticiper les coûts associés à leur prise en charge. Les incidences seront également étudiées selon l'âge et le sexe afin de proposer des pistes pour l'identification des facteurs associés à ses maladies.

Dans un deuxième chapitre nous nous intéresserons plus particulièrement aux impacts sur la vie au quotidien, en particulier à l'impact sur la vie professionnelle et le niveau d'études pour les patients débutant leur maladie dans l'enfance. La prise en compte de l'impact sur la vie socio-professionnelle des patients est également importante pour ces maladies chroniques car, au-delà du contrôle de l'inflammation et de la cicatrisation endoscopique recherchée par le médecin, le retour à une vie « normale » est quant à elle recherchée par le patient. L'incidence augmentant plus fortement dans l'enfance, ce sujet est crucial. Par ailleurs, lorsque les MICI débutent dans l'enfance, en raison de la chronicité de ces maladies, l'impact sur la vie sociale et sur la qualité de vie s'étend à long terme.

- **Le second axe** s'intéresse à l'analyse de données cliniques et omiques.

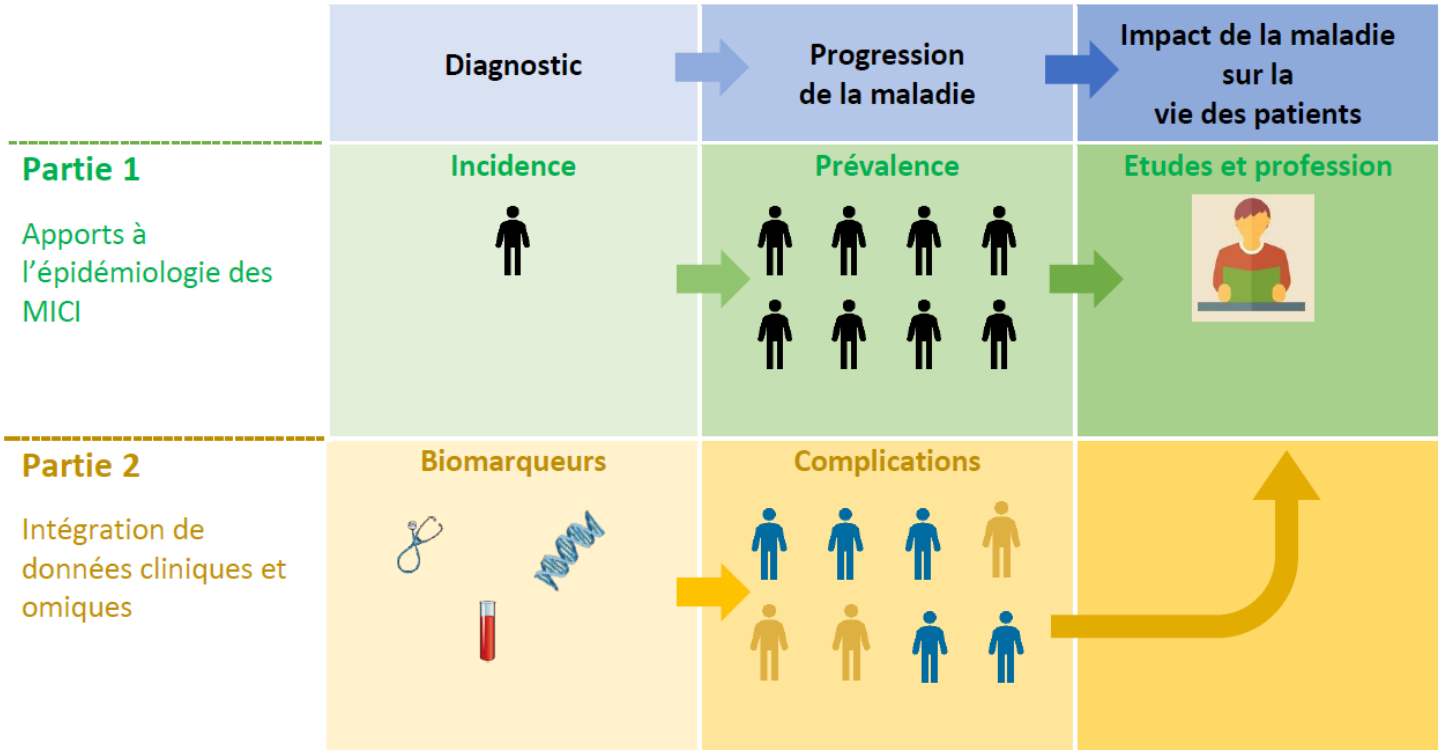
Cette problématique fait suite à l'augmentation, ces dernières décennies, du nombre de données qu'il est possible de recueillir, notamment omiques, et au besoin de développer des outils de médecine de précision. Les registres peuvent être le support d'études mobilisant des données variées récoltées auprès des patients (sang, selles etc) avec l'avantage de disposer d'un plus large éventail de patients qu'à l'hôpital. En particulier, le registre Epimad a recueilli des échantillons de sang auprès de patients ayant débuté leur maladie dans l'enfance dans le

but de développer un modèle de prédiction de la maladie de Crohn compliquée à partir de données cliniques, sérologiques et génétiques.

Avant d'analyser ces données, nous nous sommes interrogés sur les méthodes statistiques permettant d'analyser conjointement ces données cliniques et omiques. Dans un premier chapitre, nous décrirons les méthodes identifiées et réaliserons des simulations de données afin de comparer ces méthodes et d'analyser les données du registre Epimad dans un second chapitre.

Ce second chapitre a pour objectif l'identification des patients à risque de complication. En effet, mieux prendre en charge les patients signifie aussi traiter les patients à bon escient, c'est-à-dire ne pas exposer inutilement des patients à des traitements pouvant avoir des effets secondaires importants et au contraire traiter précocement les patients dont la maladie est à risque de complication. Dans ce chapitre, nous proposerons un score permettant d'identifier les patients à risque de complication dans un contexte de maladie à début pédiatrique.

Figure 11. Plan de la thèse.



Articles

Sarter H, et al. Increasing incidence of inflammatory bowel diseases in children and young adults in Northern France: a 30-year population-based study. Soumis.

Sarter H, et al. Young adult patients with paediatric-onset inflammatory bowel disease have a higher educational level and a higher employment rate than the general population. Soumis.

Sarter H, et al. A Novel 8-Predictors Signature to Predict Complicated Disease Course in Pediatric-onset Crohn's Disease: A Population-based Study. Inflamm Bowel Dis. 2023 ;29(11):1793-1804.

PARTIE 1 : Apports à l'épidémiologie des MICI

INTRODUCTION

1. L'incidence, objectif principal d'un registre

Comme cela a été décrit en introduction générale, l'objectif principal d'un registre est de mesurer l'incidence d'une pathologie à partir de données exhaustives sur un territoire donné.

Au cours des dernières décennies, les MICI sont devenues un problème de santé publique dans le monde entier (170), ces maladies ayant tout d'abord émergé dans les pays occidentaux puis dans les pays émergents. Bien qu'il soit actuellement admis que les incidences sont en phase de stabilisation, voire de décroissance, dans les pays occidentaux (35), elles continuent d'augmenter dans certains pays d'Europe occidentale et chez les enfants et adolescents (45,49,54,71,171). Par ailleurs, de nombreuses études réalisées dans différents pays sont basées sur des sources de données administratives ou hospitalières. Ces études manquent, d'une part, d'informations cliniques expertisées pour confirmer le diagnostic selon des critères validés et, d'autre part, d'informations sur les caractéristiques phénotypiques de la maladie. Or, il est important d'établir les incidences à partir de données exhaustives et détaillées en population générale.

Depuis 1988, le registre Epimad permet d'étudier les incidences des MICI, MC et RCH et leurs évolutions dans le temps et dans l'espace (50,51,53,54,59). La dernière étude portait spécifiquement sur les incidences de MICI chez les enfants et adolescents au cours de la période 1988-2011 (54). Cette étude a montré que les incidences de MC et de RCH ont augmenté de façon spectaculaire chez les adolescents au cours de cette période de 24 ans. Les dernières données publiées dans une revue internationale sur l'ensemble des classes d'âge concernaient la période 1988-2007 (50). Une réévaluation était donc nécessaire. Par ailleurs, les données de prévalence n'ont jamais été estimées à partir du registre Epimad. Les estimations de prévalences font l'objet d'un intérêt plus particulier dans la littérature ces dernières années en raison de la poursuite de leur augmentation - même en phase de stabilisation des incidences -, en lien avec l'espérance de vie de plus en plus élevée et le vieillissement de la population.

2. L'impact des MICI sur la vie des patients et sa mesure

Les MICI peuvent avoir un impact significatif sur la vie quotidienne des personnes qui en sont atteintes. Cet impact peut varier en fonction de la gravité de la maladie, de la manière dont elle est gérée et/ou traitée.

Le diagnostic de MICI constitue un changement important dans la vie des patients. Malgré l'accroissement des options thérapeutiques au cours des dernières décennies, les interventions médicales et/ou chirurgicales peuvent ne pas être suffisantes pour normaliser l'état de santé et prévenir l'invalidité à long terme.

Les PRO (Patients Reported Outcomes) mesurent la perception qu'une personne a de sa propre santé à l'aide de questionnaires. Ils permettent aux patients de rendre compte de leur qualité de vie, de leurs symptômes physiques et psychologiques non directement observables (fatigue, énergie, douleurs,…), leurs limitations fonctionnelles (sommeil, mobilité, anxiété…) et d'autres aspects de leur santé et de leur bien-être, sans interprétation par un médecin ou tout autre personne.

Ils apportent ainsi un complément aux données objectives mesurées par le clinicien : symptômes, hospitalisations, traitements etc… Un des premiers et plus connus des PRO est la qualité de vie, mesurée notamment à l'aide du SF-36 qui la mesure de manière générique ou, dans le cas particulier des MICI, à l'aide du questionnaire IBDQ ou SIBDQ. Les PRO ont pris de plus en plus d'ampleur ces dernières années et sont désormais un objectif à long terme à atteindre dans la stratégie de traitement (consensus STRIDE-II) et sont également considérés comme des critères de jugement à atteindre dans les essais cliniques (consensus SPIRIT) (88,172).

Dans le domaine des MICI, plusieurs questionnaires spécifiques à la maladie ou génériques ont été développés et/ou validés pour l'évaluation de la qualité de vie, la fatigue, la productivité au travail, le handicap fonctionnel et la dépression/l'anxiété. Le handicap fonctionnel, en particulier, fait partie des PRO inclus dans les consensus STRIDE II et SPIRIT. Dans le domaine des MICI, le questionnaire IBD-DI (IBD-Disability Index), mesurant le handicap fonctionnel (ou incapacité) a été développé par le groupe IPNIC (International Programme to Develop New Indexes for Crohn's Disease) (173). Ce questionnaire évalue le retentissement fonctionnel des MICI. Il permet d'objectiver le handicap sur différents aspects, à savoir,

physique, psychologique, familial et social. J'ai participé à la validation quantitative de cet outil (174) réalisée à partir de patients du registre Epimad. Il comprend 14 questions et varie de 0 à 100. L'IBD-DI présentait une validité factorielle, une cohérence interne, une reproductibilité inter-observateur et une consistance externe élevées, ainsi qu'une reproductibilité intra-observateur modérée. Le score moyen de l'IBD-DI dans les MICI était de 35,3 (écart-type : 20,5) et était associé au sexe, à l'activité clinique de la maladie et à la durée de la maladie. Un résultat intéressant de cette étude était que 16 % des patients présentaient une invalidité sévère telle que définie par le quatrième quartile de l'IBD-DI alors qu'ils étaient en rémission clinique, indiquant qu'il peut y avoir un décalage entre l'activité clinique de la maladie et l'invalidité ressentie par certains patients atteints de MICI et confirmant l'intérêt de tenir compte des PRO dans les essais cliniques et dans l'évaluation de la réponse au traitement.

Les PRO ont également été étudiés dans une large étude française à laquelle j'ai participé (analyses statistiques et interprétation) dans laquelle 1185 patients adultes atteints de MICI étaient interrogés via la plateforme du site de l'afa Crohn RCH France (Association François Aupetit, Association de patients atteints de MC ou RCH) (175). Cette étude montrait que le poids des MICI était important. Près de la moitié des patients avait une mauvaise qualité de vie, une fatigue élevée, et/ou était atteinte de dépression et environ un tiers des patients rapportait une anxiété et/ou un handicap élevés. Par ailleurs, un impact sur la productivité au travail ainsi que sur l'activité au travail était également mis en évidence.

3. Objectifs

L'objectif de cette première partie est de décrire et illustrer, à l'aide de deux études, les apports des données de registre dans l'épidémiologie des MICI :

- Incidences et prévalences des MICI,
- Impact sur le niveau d'études et l'insertion professionnelle des patients ayant eu une MICI à début pédiatrique.

Chapitre 1 : Incidences et prévalences des MICI

1. Contexte et objectifs

Les objectifs de cette étude étaient d'évaluer l'incidence et les caractéristiques cliniques des cas de MICI chez l'adulte et l'enfant en population générale sur une période de 30 ans (1988 à 2017), d'en évaluer les distributions selon l'âge et le sexe ainsi que les tendances temporelles. L'objectif était également d'estimer la prévalence des MICI dans la zone du registre Epimad ainsi que d'en évaluer la valeur attendue en 2030.

2. Méthodes

2.1. Population

La présente étude prenait en compte l'ensemble des patients du registre Epimad atteints de MICI (MC, RCH et CI) certaine ou probable, diagnostiquée entre 1988 et 2017, quel que soit l'âge diagnostique de la maladie.

2.2. Données

Les données phénotypiques au diagnostic incluaient l'âge au diagnostic de la maladie, la date diagnostique, le délai diagnostique, le sexe, la localisation de la maladie, le phénotype de la MC, la présence de lésions ano-périnéales pour la MC, la présence de symptômes extra-digestifs, les antécédents familiaux de MICI.

Les sites anatomiques affectés par la MC étaient définis selon la classification de Montréal (3): atteinte iléale pure (L1), atteinte colique pure (L2), atteinte iléo-colique (L3) et atteinte digestive haute (L4, qui peut être combinée avec L1, L2 ou L3) ; les patients présentant une atteinte iléo-cæcale étaient classés dans la catégorie L3. Le phénotype de la MC était classé comme B1 (purement inflammatoire), B2 (sténosant) ou B3 (pénétrant).

La localisation de la RCH était également définie selon la classification de Montréal, comme suit : proctite, avec une maladie limitée au rectum (E1) ; colite gauche, avec une maladie ne

dépassant pas l'angle gauche (E2) ; ou colite étendue, avec une atteinte dépassant l'angle gauche (E3).

2.3. Analyses statistiques

Les variables continues ont été décrites par la médiane et l'intervalle interquartile (IQR) et les variables catégorielles par la fréquence et le pourcentage. Les comparaisons intergroupes des variables continues ont été effectuées à l'aide du test de Wilcoxon-Mann-Whitney. Les comparaisons intergroupes de variables qualitatives ont été effectuées à l'aide du test du χ^2 ou, en fonction des effectifs attendus, du test exact de Fisher. Les changements dans le temps des variables catégorielles ont été évalués à l'aide d'un test du χ^2 de tendance lorsque cela était adapté.

INCIDENCES

Les taux d'incidence ont été calculés comme le nombre de cas incidents (nouveaux diagnostics) divisé par la population à risque, et présentés pour l'ensemble de la période d'étude et pour dix périodes de trois ans (de 1988-1990 à 2015-2017). Les taux d'incidence ont été standardisés sur l'âge en utilisant les pondérations pour la population européenne fournies par Eurostat (163) et déterminés pour la population dans son ensemble, pour chaque classe d'âge (<17 ans, 17-39 ans, 40-59 ans, et 60 ans et plus), et par sexe. Les taux d'incidence standardisés sont présentés avec leur intervalle de confiance (IC) à 95 %, basé sur une loi gamma (164). Pour chacun des quatre départements de la zone du registre, les données annuelles de population par âge et par sexe ont été obtenues à partir des données de recensement de l'INSEE.

Les différences d'incidence en fonction de l'âge, du sexe ou du temps (tendances temporelles) ont été testées à l'aide de modèles de régression de Poisson log-linéaires tenant compte du nombre de personnes-années à risque (introduit comme variable d'offset) et de la sur-dispersion, le cas échéant. Les différences selon l'âge et le sexe ont été présentées sous forme de rapports de taux d'incidence (IRR : Incidence Rate Ratio) obtenus par exponentiation des coefficients pour l'âge ou le sexe. Pour évaluer les tendances temporelles linéaires, l'année a été introduite comme variable explicative du modèle. Les tendances temporelles ont été présentées sous la forme de pourcentage annuel de variation (APC : annual percent change) obtenu par l'exponentiation du coefficient de la régression de Poisson pour la variable

« année ». Pour évaluer les différences de tendances temporelles en fonction du sexe ou de l'âge, une interaction entre le sexe et l'année ou entre la classe d'âge et l'année a été introduite dans le modèle.

PREVALENCES

Pour estimer la prévalence des MICI dans la population en 2010, 2020 et 2030 :

i) Nous avons considéré que l'augmentation de l'incidence a débuté en 1975 dans la zone Epimad (avis d'expert). Ainsi, une augmentation linéaire du nombre de cas de 1975 à 1987 a été extrapolée. Le nombre de cas incidents entre 1975 et 1987 a ainsi été estimé et les cas incidents répartis par âge, sexe et type de MICI selon la distribution observée dans le registre Epimad sur la période 1988-1990.

ii) Entre 1988 et 2017, les cas incidents étaient ceux observés dans le registre Epimad.

iii) Les taux d'incidence annuels de 2018 à 2030 ont été projetés à partir des APC observés de 1988 à 2017 pour chaque groupe d'âge, sexe et type de MICI et ont été appliqués aux projections de population fournies par l'INSEE (176). Les projections de population de l'Insee (projections Omphale) sont basées sur des hypothèses concernant l'évolution de trois composantes des variations de population : fécondité, mortalité et migrations. Pour ces projections de population, le scénario « central » de l'Insee a été utilisé (Table 4).

Un processus de vieillissement a ensuite été appliqué à chaque cas incident, chaque année, de l'année de diagnostic à 2030 en utilisant les taux de survie annuels par âge, sexe et cohorte de naissance fournis par l'INSEE (177).

Les prévalences ont été décrites selon l'âge en 2010, 2020 et 2030.

L'analyse statistique a été réalisée à l'aide des logiciels SAS (version 9.4, SAS Institute Inc., Cary, NC, USA) et R version 3.6.1 (R Foundation for Statistical Computing, Vienne, Autriche). Le seuil de signification statistique a été fixé à $p \leq 0,05$.

Table 4. Hypothèses utilisées pour les projections de population du scénario central des projections Omphale de l'Insee.

	Situation en 2020	Hypothèse centrale
Fécondité		
Indice conjoncturel de fécondité	1,83 enfant par femme	1,80 enfant à partir de 2022
Âge moyen à la maternité	30,8 ans	33,0 ans à partir de 2052
Espérance de vie		
Espérance de vie à la naissance des femmes	85,1 ans	90,0 ans en 2070
Espérance de vie à la naissance des hommes	79,1 ans	87,5 ans en 2070
Migrations		
Valeur du solde migratoire annuel	+ 70 000 par an	+ 70 000 par an

3. Résultats

Entre le 1er janvier 1988 et le 31 décembre 2017, 22 879 cas incidents de MICI ont été identifiés dans le registre Epimad, dont 13 445 cas de MC (59 %), 8 803 cas de RCH (38 %) et 631 cas de CI (3 %).

3.1. Présentation clinique de la maladie au diagnostic

Les caractéristiques cliniques des patients sont décrites dans la table 5. L'âge médian au moment du diagnostic était de 26 ans (IQR : [20-38]) pour la MC et de 35 ans ([25-48]) pour la RCH. Dix pour cent des patients (n=2 329) présentaient des antécédents familiaux de MICI. Le délai médian entre l'apparition des symptômes et le diagnostic (délai au diagnostic) était significativement plus élevé dans la MC : 3 mois ([1-9]) versus 2 mois ([1-6]) dans la RCH (p<0,0001). Des manifestations extra-intestinales étaient présentes chez 8 % (n=1 899) des patients, 12 % (n=1 565) des cas de MC et 3 % (n=299) des RCH.

Dans la MC, 50 % des patients (n=6 467) présentaient une atteinte iléo-colique (L3), 30 % (n=3 913) une atteinte colique pure (L2) et 20 % (n=2 666) une atteinte iléale pure (L1) ; 22 % (n=2 951) présentaient une atteinte du tractus digestif supérieur (L4). Cinq pour cent des patients atteints de MC (n=658) présentaient des abcès et/ou des fistules ano-périnéales au moment du diagnostic. En ce qui concerne le phénotype de la maladie au moment du diagnostic de MC, 74 % des patients présentaient un phénotype purement inflammatoire (n=3 840), 18 % un phénotype sténosant (n=928) et 8 % un phénotype pénétrant (n=433), soit 26% (n=1361) avec un phénotype d'emblée compliqué.

L'atteinte colorectale au moment du diagnostic de RCH était classée comme proctite (E1) pour 36 % (n=2 741) des patients, colite gauche (E2) pour 35 % (n=2 640) et colite étendue (E3) pour 29 % (n=2 257).

Table 5. Données sociodémographiques et cliniques des patients atteints de MICI issus du registre Epimad de 1988 à 2017 (n=22 879).

Caractéristiques au diagnostic	n (%) ou médiane [IQR]
Type de MICI	
MC	13 445 (58,8 %)
RCH	8 803 (38,5 %)
CI	631 (2,8 %)
Maladie de Crohn	
Age médian [IQR] (années)	26 [20-38]
Sexe féminin	7 535 (56,0 %)
Antécédents familiaux de MICI	1 700 (12,6 %)
Délai au diagnostic (mois)	3 [1; 9]
Localisation[1]	
L1	2 666 (20,4 %)
L2	3 913 (30,0 %)
L3	6 467 (49,6 %)
L4	2 951 (21,9 %)
Phénotype[1,2]	
B1	3 840 (73,8 %)
B2	928 (17,8 %)
B3	433 (8,3 %)
Lesions ano-périnéales	658 (4,9 %)
Symptômes extra-digestifs	1 565 (11,6 %)
RCH	
Age médian [IQR] (années)	35 [25-48]
Sexe féminin	4 104 (46,6 %)
Antécédents familiaux de MICI	594 (6,7 %)
Délai au diagnostic (mois)	2 [1; 6]
Localisation[1]	
E1	3 157 (36,3 %)
E2	3 210 (36,9 %)
E3	2 329 (26,8 %)
Symptômes extra-digestifs	299 (3,4 %)

[1] *Selon la classification de Montréal.*
[2] *Le phénotype n'est enregistré que depuis 2008.*

3.1.1. Caractéristiques selon le sexe

La proportion de femmes était significativement plus élevée dans la MC (sex-ratio femmes/hommes : 1,27) que dans la RCH (sex-ratio : 0,87) (p<0,0001). Les données sociodémographiques et cliniques sont résumées par sexe dans la table 6.

MALADIE DE CROHN :

Les femmes et les hommes ne différaient pas significativement concernant le phénotype (p=0,550) et la localisation de la maladie (0,476), à l'exception de l'atteinte L4 (observée chez 24 % des hommes versus 20 % des femmes ; p<0,0001) et de la présence de lésions ano-périnéales (6 % des hommes versus 4 % des femmes ; p<0,0001). Les manifestations extra-intestinales étaient significativement plus fréquentes chez les femmes que chez les hommes (12 % versus 11 %, respectivement ; p=0,027). La distribution de l'âge diagnostique ne différait pas significativement selon le sexe (p=0,128).

RECTOCOLITE HEMORRAGIQUE :

Les hommes présentaient une maladie plus étendue au moment du diagnostic que les femmes (39 % d'atteinte E2 et 28 % d'E3 chez les hommes, versus 35 % d'atteinte E2 et 25 % d'atteinte E3 chez les femmes ; p<0,0001). Les antécédents familiaux de MICI étaient plus fréquents chez les femmes (8 % versus 6 % chez les hommes ; p<0,0001). Les manifestations extra-intestinales étaient plus fréquentes chez les femmes (4 % versus 3 % chez les hommes ; p=0,011). L'âge médian au diagnostic était significativement plus élevé chez les hommes (38 ans [27 ; 51] versus 32 ans [24 ; 44] chez la femme, p<0,0001).

Chez les hommes comme chez les femmes, le délai médian entre l'apparition des symptômes et le diagnostic était de 3 mois pour la MC et de 2 mois pour la RCH.

Table 6. Données sociodémographiques et cliniques des patients atteints de MICI issus du registre Epimad pour la période 1988-2017, selon le sexe (n=22 879).

Variables at diagnosis	Hommes (n=10 925)	Femmes (n=11 954)	p-value
Type de MICI			
MC	5 910 (54,0 %)	7 535 (63,0 %)	
RCH	4 699 (43,0 %)	4 104 (34,3 %)	<0,0001
CI	316 (3,0 %)	315 (2,6 %)	
Maladie de Crohn			
Age médian [IQR] (années)	26 [19-39]	26 [20-38]	0,128
Antécédents familiaux de MICI	720 (12,2 %)	980 (13,0 %)	0,154
Délai au diagnostic (mois)	3 [1; 8]	3 [1; 9]	0,001
Localisation[1]			
L1	1 176 (20,6 %)	1 490 (20,3 %)	
L2	1 681 (29,4 %)	2 232 (30,4 %)	0,476
L3	2 853 (50,0 %)	3 614 (49,3 %)	
L4	1 418 (24,0 %)	1 533 (20,3 %)	<0,0001
Phénotype[1,2]			
B1	1 443 (73,7 %)	1 769 (74,9 %)	
B2	3 57 (18,2 %)	420 (17,8 %)	0,550
B3	158 (8,1 %)	172 (7,3 %)	
Lesions ano-périnéales	374 (6,3 %)	284 (3,8 %)	<0,0001
Symptômes extra-digestifs	647 (10,9 %)	918 (12,2 %)	0,027
RCH			
Age médian [IQR] en années	38 [27-51]	32 [24-44]	<0,0001
Antécédents familiaux de MICI	264 (5,6 %)	330 (8,0 %)	<0,0001
Délai au diagnostic (mois)	2 [1; 6]	2 [1; 6]	0,013
Localisation[1]			
E1	1 536 (33,1 %)	1 621 (40,0 %)	
E2	1 802 (38,8 %)	1 408 (34,7 %)	<0,0001
E3	1 304 (28,1 %)	1 025 (25,3 %)	
Symptômes extra-digestifs	138 (2,9 %)	161 (3,9 %)	0,011

[1] *Selon la classification de Montreal*
[2] *Le phénotype n'est enregistré que depuis 2008*

3.1.2. Caractéristiques selon l'âge

Les données sociodémographiques et cliniques des patients atteints de MICI sont décrites par groupe d'âge dans la table 7.

MALADIE DE CROHN :

La localisation de la maladie différait significativement en fonction de l'âge (p<0,0001) : l'atteinte iléale (L1 + L3) était plus fréquente dans les classes d'âge <17 ans et 17-39 ans (74 % dans les deux groupes) que dans les classes plus âgées (62 % et 41 % pour les 40-59 ans et pour les 60 ans et plus, respectivement). La proportion de patients présentant une atteinte digestive haute (L4) diminuait significativement avec l'âge : de 32 % chez les <17 ans à 12 % chez les 60 ans et plus (p<0,0001). Le phénotype était inflammatoire dans une plus forte proportion chez les cas pédiatriques (83 % des moins 17 ans, 73 % des 17-39 ans, 70 % des 40-59 ans et 67 % des 60 ans et plus, p<0,0001).

RECTOCOLITE HEMORRAGIQUE :

La proctite (E1) était la localisation la plus fréquente dans les classes d'âge 17-39 ans et 40-59 ans (40 % des patients), tandis que la colite gauche (E2) était la plus fréquente (55 %) chez les 60 ans et plus et la colite étendue (E3) plus fréquente chez les moins de 17 ans (47 %) (p<0,0001).

Dans la MC comme dans la RCH, les antécédents familiaux et la présence de manifestations extra-intestinales au moment du diagnostic étaient significativement plus fréquents chez les enfants. En effet, les antécédents familiaux étaient présents chez 18 % des moins de 17 ans dans la MC (versus 13 %, 9 % et 5 % des 17-39 ans, 40-59 ans et 60 ans et plus, respectivement ; p<0,0001) et 14% des moins de 17 ans dans la RCH (versus 8 %, 4 % et 3 % des 17-39 ans, 40-59 ans et 60 ans et plus, respectivement ; p<0 ,0001). Les manifestations extra-digestives concernaient 22 % des moins de 17 ans dans la MC (versus 10 %, 11 % et 7 % des 17-39 ans, 40-59 ans et 60 ans et plus, respectivement ; p<0,0001) et 6% des moins de 17 ans dans la RCH (versus 3 % des 17-39 ans, 40-59 ans et 60 ans et plus, p=0,003).

Table 7. Données sociodémographiques et cliniques des patients atteints de MICI issus du registre Epimad pour la période 1988-2017, selon la classe d'âge (n=22 879).

Caractéristiques au diagnostic	<17 (n=2 103)	17-39 (n=13 894)	40-59 (n=4 905)	≥ 60 (n=1 977)	p-value
Type de MICI					
MC	1 510 (71,8 %)	8 796 (63,3 %)	2 318 (47,3 %)	821 (41,5 %)	<0,0001
RCH	562 (26,7 %)	4 766 (34,3 %)	2 424 (49,4 %)	1 051 (53,2 %)	
CI	31 (1,5 %)	332 (2,4 %)	163 (3,3 %)	105 (5,3 %)	
Maladie de Crohn					
Antécédents familiaux de MICI	276 (18,3 %)	1 168 (13,3 %)	212 (9,1 %)	44 (5,4 %)	<0,0001
Femmes	672 (44,5 %)	5 150 (58,6 %)	1 229 (53,0 %)	484 (58,9 %)	<0,0001
Délai au diagnostic (mois)	3 [2 ; 7]	3 [1; 9]	3 [1; 9]	3 [1; 8]	0,101
Localisation[1]					
L1	222 (15,5 %)	1 785 (20,9 %)	531 (23,6 %)	125 (15,8 %)	<0,0001
L2	378 (26,0 %)	2 213 (25,9 %)	858 (38,0 %)	464 (58,5 %)	
L3	849 (58,5 %)	4 549 (53,2 %)	865 (38,4 %)	204 (25,7 %)	
L4	482 (31,9 %)	2 208 (23,1 %)	343 (14,8 %)	98 (11,9 %)	<0,0001
Phénotype[1,2]					
B1	665 (82,8 %)	2 386 (73,1 %)	594 (70,2 %)	195 (67,5 %)	<0,0001
B2	105 (13,1 %)	584 (17,9 %)	175 (20,7 %)	64 (22,1 %)	
B3	33 (4,1 %)	293 (9,0 %)	77 (9,1 %)	30 (10,4 %)	
Lesions ano-périnéales	92 (6,1 %)	398 (4,5 %)	118 (5,1 %)	50 (6,1 %)	0,019
Symptômes extra-digestifs	340 (22,5 %)	924 (10,5 %)	246 (10,6 %)	55 (6,7 %)	<0,0001
Rectocolite Hémorragique					
Antécédents familiaux de MICI	80 (14,2 %)	372 (7,8 %)	109 (4,5 %)	33 (3,1 %)	<0,0001
Femmes	317 (56,4 %)	2 472 (51,9 %)	890 (36,7 %)	425 (40,4 %)	<0,0001
Délai au diagnostic (mois)	2 [1 ; 5]	2 [1; 6]	2 [1; 6]	2 [1; 5]	0,636
Localisation[1]					
E1	122 (22,1 %)	1 895 (40,2 %)	925 (38,7 %)	215 (20,6 %)	<0,0001
E2	171 (30,9 %)	1 537 (32,7 %)	934 (39,0 %)	568 (54,6 %)	
E3	260 (47,0 %)	1 278 (27,1 %)	533 (22,3 %)	258 (24,8 %)	
Symptômes extra-digestifs	34 (6,0 %)	162 (3,4 %)	71 (2,9 %)	32 (3,0 %)	0,003

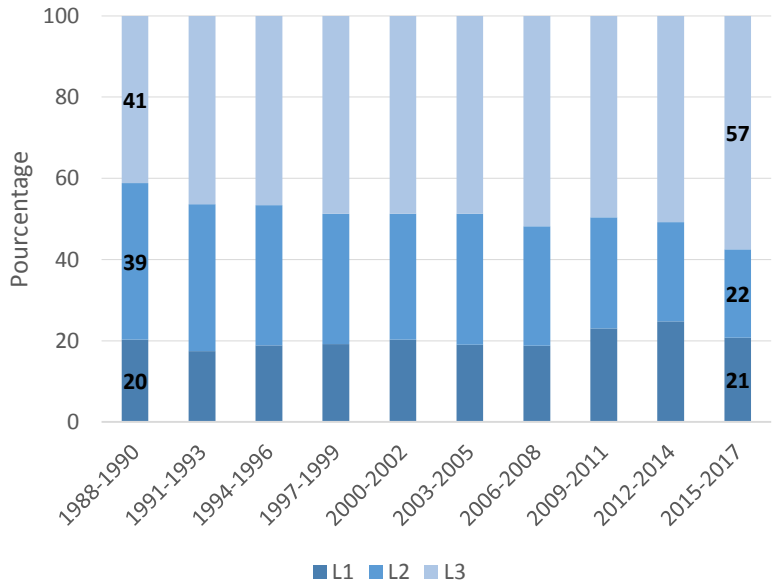
[1]Selon la classification de Montréal.

[1,2]Le phénotype n'est enregistré que depuis 2008.

3.1.3. Evolution des caractéristiques phénotypiques dans le temps

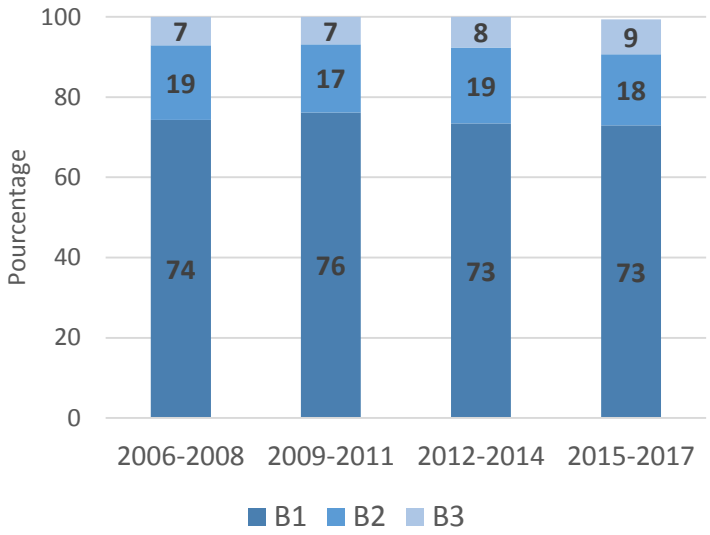
La proportion de patients ayant une atteinte iléo-colique (L3) a augmenté de manière significative entre 1988-1990 et 2015-2017 (passant de 41 % à 57 %), tandis que la proportion de patients présentant une localisation purement colique (L2) a diminué de manière significative, passant de 38 % à 22 % (test de tendance : p<0,0001) (Figure 12). On peut noter que le type d'exploration a également varié significativement dans le temps. En effet, la proportion de patients ayant eu une radiographie colique a diminué de manière significative, passant de 48 % en 1988-1990 à 0 % en 2015-2017 (p<0,0001), de même que la proportion de patients ayant eu une radiographie du grêle (de 72 % à 0 % sur la même période; p<0,0001). En revanche, la proportion de patients ayant eu une entérographie par scanner et/ou IRM a augmenté de manière significative, passant de 0 % en 1988-1990 à 58 % en 2015-2017 (p<0,0001). La proportion de patients ayant eu une iléoscopie est passée de 6 % à 69 % sur la même période (p<0,0001) et la proportion de patients ayant eu une coloscopie totale est passée de 50 % à 87 % toujours sur la même période (p<0,0001). La proportion de patients présentant une atteinte digestive haute L4 a significativement augmenté de 11 % en 1988-1990 à 25 % en 2015-2017 (p<0,0001), parallèlement à l'augmentation de la proportion de gastroscopies, passant de 34 % à 72 % sur la même période (p<0,0001).

Figure 12. Évolution des localisations de la maladie chez les patients atteints de MC (n=13 445) dans le registre Epimad de 1988 à 2017.



La proportion de lésions ano-périnéales (abcès et/ou fistules) était stable dans le temps (test de tendance : p=0,386). La répartition des phénotypes de la maladie, enregistrée depuis 2008, restait également stable au fil du temps (p=0,494) (Figure 13).

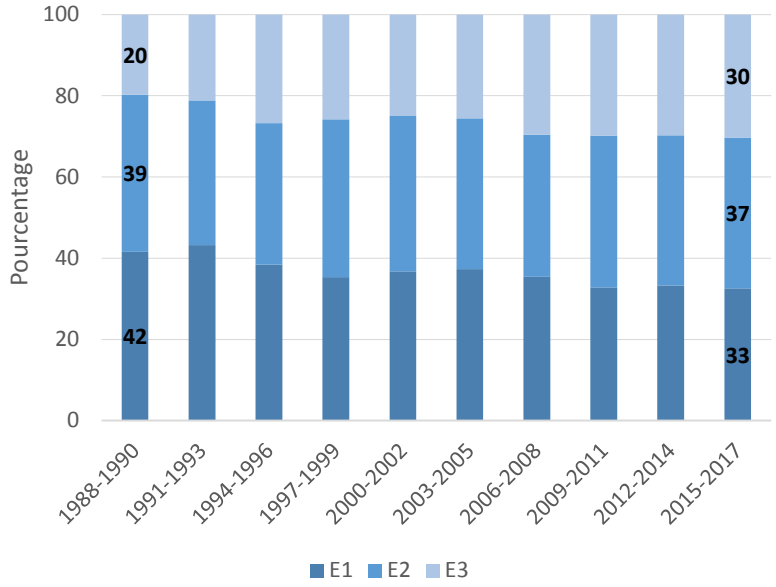
Figure 13. Évolution du phénotype de la maladie chez les patients atteints de MC (n=13 445) dans le registre Epimad de 2006-2008 à 2015-2017.



Chez les patients atteints de RCH, la proportion de proctites (E1) a significativement diminué de 42 % à 33 %, tandis que la proportion d'atteintes E3 a augmenté significativement de 20 % à 30 % (p<0,0001) (Figure 14). Le type d'exploration du côlon a également varié de manière significative dans la RCH, avec une augmentation de la coloscopie totale de 54 % en 1988-1990 à 77 % en 2015-2017 (p<0,0001) et une diminution de la radiographie colique de 31 % à 0 % sur la même période (p<0,0001).

La proportion de patients présentant des manifestations extra-intestinales ne montrait aucune tendance à la baisse ou à la hausse, que ce soit dans la MC (test de tendance : p=0,079) ou dans la RCH (p=0,121). La proportion de patients ayant des antécédents familiaux de MICI augmentait significativement dans les deux maladies, passant de 9 % à 15 % dans la MC (p<0,0001) et de 4 % à 10 % dans la RCH (p<0,0001), possiblement en lien avec le gastroentérologue prenant en charge le patient qui pose plus souvent la question avec l'évolution des connaissances scientifiques. Il n'y a eu aucun changement significatif au fil des années dans le délai diagnostique, ni dans l'âge diagnostique et ce dans les deux maladies, MC ou RCH.

Figure 14. Évolution des localisations de la maladie chez les patients atteints de de RCH (n=8 803) dans le registre Epimad de 1988 à 2017.



3.2. Estimation des incidences

Au cours de la période 1988-2017, l'incidence annuelle de MICI, standardisée sur l'âge, était de 12,7 pour 10^5 personnes-années (IC 95 % : [12,5 ; 12,8]). Les incidences de MC, de RCH et de CI étaient respectivement de 7,2 pour 10^5 [7,1 ; 7,3], 5,1 pour 10^5 [5,0 ; 5,2], et 0,37 pour 10^5 [0,34 ; 0,40].

3.2.1. Incidences selon le sexe et l'âge

Sur l'ensemble de la période 1988-2017, le taux d'incidence de MC était significativement plus élevé chez les femmes (8,0/10^5 [7,8 ; 8,2]) que chez les hommes (6,4 [6,3 ; 6,6]) (p<0,0001), correspondant à un rapport de taux de d'incidence (Incidence Rate Ratio : IRR) de 1,25 [1,19 ; 1,31] (Table 8). L'âge médian au moment du diagnostic était de 26 ans pour les femmes comme pour les hommes (p=0,128).

Le taux d'incidence de RCH était significativement plus faible chez les femmes (4,5 [4,4 ; 4,7]) que chez les hommes (5,7 [5,5 ; 5,9]) (p<0,0001), correspondant à un rapport de taux de d'incidence de 0,83 [0,78 ; 0,89]. L'âge médian au moment du diagnostic était significativement plus faible chez les femmes que chez les hommes (32 ans [24-44] versus 38 ans [27-51] ; p<0,0001).

Le taux d'incidence de MC était significativement plus élevé dans le groupe d'âge des 17-39 ans (15,4 [15,1 ; 15,8]) que dans les groupes d'âge des moins de 17 ans (3,7 [3,5 ; 3,9]), des 40-59 ans (5,3 [5,1 ; 5,6]) et des 60 ans et plus (2,4 [2,3 ; 2,6]) (p<0,0001). De même, dans la RCH, le taux d'incidence était significativement plus élevé dans le groupe d'âge des 17-39 ans (8,7 [8,4 ; 8,9]) que dans le groupe des moins de 17 ans (1,4 [1,3 ; 1,5], le groupe des 40-59 ans (5,6 [5,4 ; 5,8]) et le groupe des 60 ans et plus (3,1 [2,9 ; 3,3]) (p<0,0001). Les rapports de taux d'incidences correspondant sont présentés dans la table 8.

Table 8. Taux d'incidence /10^5 personnes-années sur la période d'étude 1988-2017 et ratio de taux d'incidence (IRR) selon le sexe et le groupe d'âge dans la MC et la RCH.

Type de MICI	Variable	Taux d'incidence 1988-2017 /10^5 personnes-années (IC à 95%)	Rapport de taux d'incidence (IRR) et IC à 95%	p
MC	Hommes	6,4 [6,3 ; 6,6]	Réf	
	Femmes	8,0 [7,8 ; 8,2]	1,25 [1,19 ; 1,31]	<0,0001
	<17 ans	3,7 [3,5 ; 3,9]	Réf	
	17-39 ans	15,4 [15,1 ; 15,8]	4,28 [3,94 ; 4,65]	<0,0001
	40-59 ans	5,3 [5,1 ; 5,6]	1,43 [1,29 ; 1,58]	<0,0001
	60 ans et plus	2,4 [2,3 ; 2,6]	0,63 [0,56 ; 0,72]	<0,0001
RCH	Hommes	5,7 [5,5 ; 5,9]	Réf	
	Femmes	4,5 [4,4 ; 4,7]	0,83 [0,78 ; 0,89]	<0,0001
	<17 ans	1,4 [1,3 ; 1,5]	Réf	
	17-39 ans	8,7 [8,4 ; 8,9]	6,26 [5,41 ; 7,25]	<0,0001
	40-59 ans	5,6 [5,4 ; 5,8]	4,07 [3,49 ; 4,74]	<0,0001
	60 ans et plus	3,1 [2,9 ; 3,3]	2,28 [1,92 ; 2,71]	<0,0001

Comme le montre la figure 16, il existe une interaction significative entre le sexe et l'âge dans la MC (p<0,0001) et dans la RCH (p<0,0001).

Dans la MC, l'incidence est significativement plus élevée chez les hommes que chez les femmes dans le groupe d'âge des moins de 17 ans (Ratio de taux d'incidence, IRR : 0,8 [0,7 ; 1,0], p=0,012), alors que l'incidence est significativement plus élevée chez les femmes que chez les hommes dans le groupe d'âge des 17-39 ans (IRR : 1,4 [1,3 ; 1,5], p<0,0001).

Dans la RCH, on observe une inversion du sex-ratio avec l'âge. L'incidence est significativement plus élevée chez les femmes dans les classes d'âge des moins de 17 ans et des 17-39 ans (IRR : 1,4 [1,1 ; 1,7], p=0,004 et 1,1 [1,1 ; 1,2], p=0,036, respectivement) et significativement plus élevée chez les hommes dans les groupes d'âge 40-59 ans et 60 ans et plus (IRR : 0,6 [0,5 ; 0,6], p<0,0001 et 0,5 [0,4 ; 0,6], p<0,00001, respectivement).

3.2.2. Evolution temporelle des incidences sur la période 1988-2017

INCIDENCES GLOBALES

Au cours de la période d'étude, l'incidence globale des MICI est passée de 10,4 [9,9 ; 10,9] pour 10^5 personnes-années en 1988-1990 à 14,1 [13,6 ; 14,7] pour 10^5 personnes-années en 2015-2017 (Pourcentage de variation annuel, APC : +1,5 % [1,2 ; 1,8] ; p<0,0001). L'incidence de MC a augmenté de 5,1 [4,8 ; 5,5] en 1988-1990 à 7,9 [7,4 ; 8,3] en 2015-2017 (APC : +1,9 % [1,6 ; 2,2] ; p<0. 0001), tandis que l'incidence de la RCH a augmenté de 4,5 [4,1 ; 4,9] en 1988-1990 à 6,1 [5,7 ; 6,5] en 2015-2017 (APC : +1,3 % [0,9 ; 1,7] ; p=<0,0001) (Figure 15).

Figure 15. Évolution dans le temps des taux d'incidence standardisés des MICI (n=22 879), de la maladie de Crohn (n=13 445) et de la RCH (n=8 803) dans le registre Epimad de 1988 à 2017. Chaque point correspond à la valeur pour une période de 3 ans. PA : personnes-années.

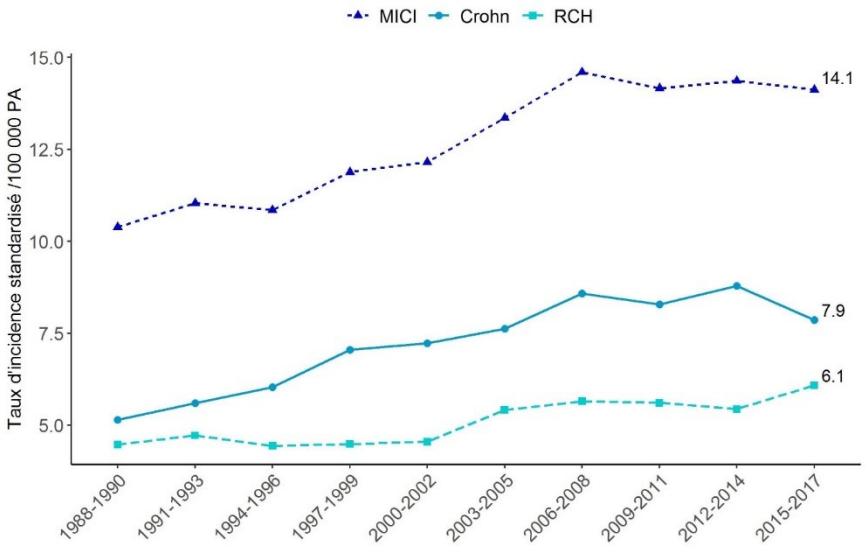
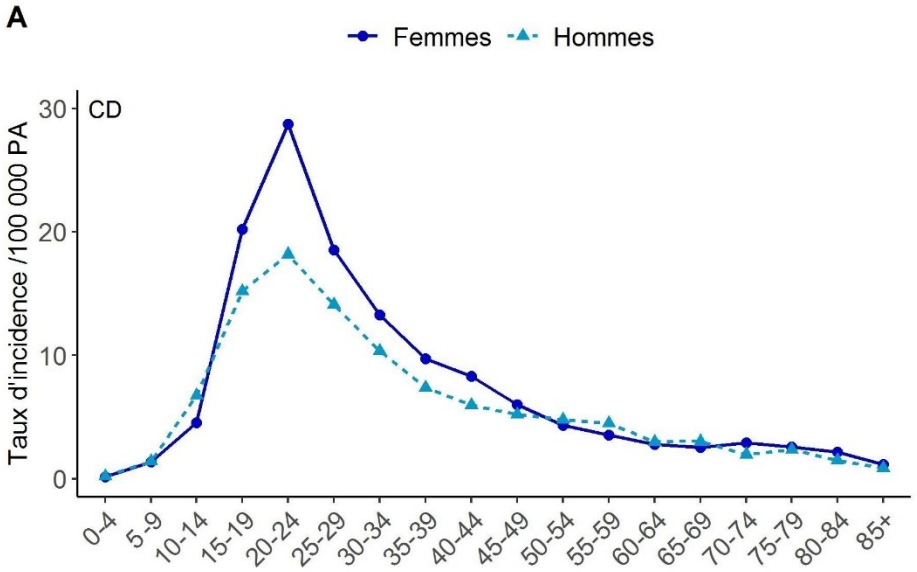
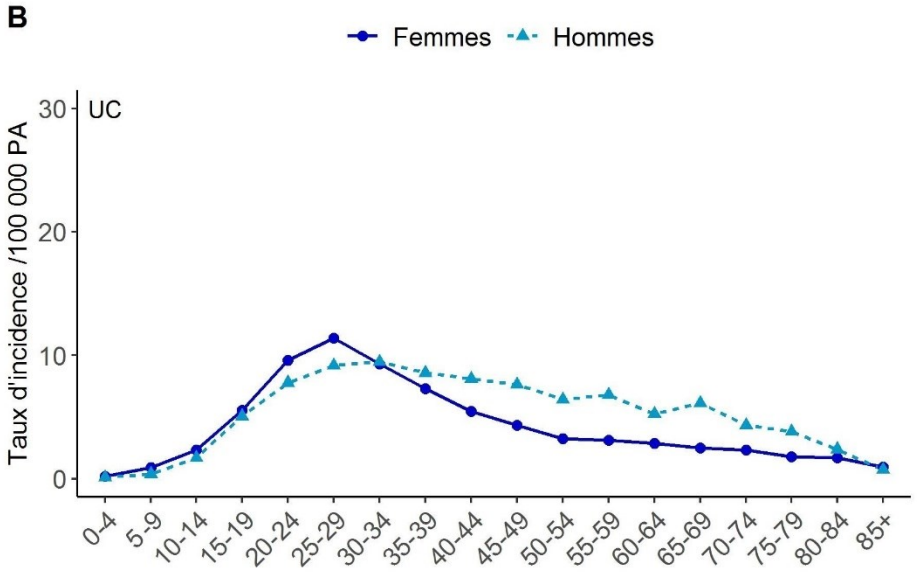


Figure 16. Taux d'incidence de maladie de Crohn (n=13 445, panel A) et de RCH (n=8 803, panel B) dans le registre Epimad sur la période d'étude (1988-2017), par sexe et par tranche d'âge de 5 ans ; Rapport des taux d'incidence femmes/hommes selon la tranche d'âge (<17 ans, 17-39 ans, 40-59 ans et 60 ans et plus) dans la maladie de Crohn (panel C) et la RCH (panel D) sur la période d'étude (1988-2017). PA : personnes-années.



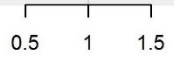


C

Classe d'âge	IRR Femmes/Hommes (IC 95%)	p-value
<17 ans	0.8 [0.7 ; 1.0]	0.012
17-39 ans	1.4 [1.3 ; 1.5]	<.0001
40-59 ans	1.1 [1.0 ; 1.2]	0.098
60 ans et plus	1.0 [0.8 ; 1.2]	0.974

D

Classe dâge	IRR Femmes/Hommes (IC 95%)	p-value
<17 ans	1.4 [1.1 ; 1.7]	0.004
17-39 ans	1.1 [1.0 ; 1.2]	0.036
40-59 ans	0.6 [0.5 ; 0.6]	<.0001
60 ans et plus	0.5 [0.4 ; 0.6]	<.0001



EVOLUTIONS SELON L'AGE ET LE SEXE

Ensemble des MICI

L'augmentation de l'incidence des MICI dans le temps n'était pas significativement différente selon le sexe (interaction année*sexe : p=0,229). En revanche, les tendances temporelles différaient significativement selon la classe d'âge (interaction temps*âge : p<0,0001), avec une stabilité dans le groupe des 60 ans et plus (APC : -0,5 % [-1,1 ; 0,1], p=0,131) et la plus forte augmentation observée dans le groupe des <17 ans (APC : +4,6 % [3,9 ; 52], p<0,0001) (Table 9).

Table 9. Evolutions temporelles des taux d'incidence des MICI à partir des données du registre Epimad de 1988 à 2017, par sexe et par groupe d'âge (n=22 879).

Sexe	Classe d'âge	Incidence standardisée /10^5		APC % / an	p
		1988-1990	2015-2017		
Tous	<17 ans	3,1 [2,6 ; 3,7]	8,6 [7,7 ; 9,5]	**+4,6 [3,9 ; 5,2]*** **	**<0,0001**
	17-39 ans	18,8 [17,7 ; 20,1]	27,5 [26,1 ; 28,9]	**+1,6 [1,4 ; 1,8]*** **	**<0,0001**
	40-59 ans	10,0 [8,9 ; 11,2]	11,4 [10,4 ; 12,4]	**+0,5 [0,1 ; 0,9]***	**0,012**
	≥60 ans	6,5 [5,5 ; 7,6]	6,0 [5,3 ; 6,8]	-0,5 [-1,1 ; 0,1]	0,131
	Tous	10,4 [9,9 ; 10,9]	14,1 [13,6 ; 14,7]	**+1,5 [1,2; 1,8] *** **	<0,0001
Femmes	<17 ans	3,2 [2,4 ; 4,1]	8,1 [6,8 ; 9,5]	**+4,3 [3,4 ; 5,2]*** **	**<0,0001**
	17-39 ans	20,7 [19,0 ; 22,5]	29,5 [27,5 ; 31,7]	**+1,8 [1,5 ; 2,1]*** **	**<0,0001**
	40-59 ans	7,8 [6,4 ; 9,3]	11,2 [9,9 ; 12,6]	**+1,4 [0,8 ; 2]*** **	**<0,0001**
	≥60 ans	5,6 [4,4 ; 7,0]	5 [4,1 ; 6,1]	-0,5 [-1,3 ; 0,4]	0,281
	Tous	10,1 [9,4 ; 10,9]	14,3 [13,6 ; 15,1]	**+1,6 [1,2 ; 2,1]*** **	**<0,0001**
Hommes	<17 ans	3,0 [2,3 ; 3,8]	9,1 [7,8 ; 10,5]	**+4,8 [3,9 ; 5,7]*** **	**<0,0001**
	17-39 ans	17,0 [15,5 ; 18,7]	25,5 [23,6 ; 27,5]	**+1,4 [1 ; 1,7]*** **	**<0,0001**
	40-59 ans	12,3 [10,6 ; 14,2]	11,5 [10,2 ; 13]	-0,1 [-0,7 ; 0,4]	0,609
	≥60 ans	8,0 [6,3 ; 10,2]	7,1 [5,9 ; 8,4]	-0,6 [-1,4 ; 0,3]	0,181
	Tous	10,9 [10,0 ; 11,8]	14,0 [13,2 ; 14,8]	**+1,3 [0,8 ; 1,7]*** **	**<0,0001**

* Tendance temporelle significative, 0.001≤p<0.05
** Tendance temporelle significative, 0.0001≤p<0.001
*** Tendance temporelle significative, p<0.0001
APC (annual percent change) : pourcentage de variation annuel estimé par modèle log-linéaire de Poisson

Maladie de Crohn

Les tendances temporelles de l'incidence de MC ne différaient pas selon le sexe (interaction temps*sexe : p=0,365) (Figure 17, panel A). Parmi les patients atteints de MC, le sex-ratio (F/H) était stable dans le temps (p=0,085), fluctuant entre 1,1 et 1,4. En revanche, les tendances temporelles différaient selon la classe d'âge (interaction année*âge : p<0,0001) (Figure 17, panel B et Table 10). La plus forte augmentation était observée dans le groupe d'âge des moins de 17 ans avec un pourcentage de variation annuel de +4,3% [3,5 ; 5,1] (p<0,0001), suivi par le groupe d'âge des 17-39 ans (APC : +1,9% [1,5 ; 2,2], p<0,0001), et le groupe d'âge des 50-59 ans (APC : +0,9 % [0,3 ; 1,5], p=0,006). Chez les 60 ans et plus, l'incidence de la MC était stable dans le temps (APC : +0,1 % [-1,0 ; 1,1], p=0,893). Dans chaque classe d'âge, les tendances temporelles n'étaient pas significativement différentes entre les hommes et les femmes (interaction année*sexe : p=0,387, 0,661, 0,071 et 0,101 dans les groupes d'âge <17 ans, 17-39 ans, 40-59 ans et 60 et plus, respectivement) (Table 10 et Figure 18).

Table 10. Evolutions temporelles des taux d'incidence de maladie de Crohn à partir des données du registre Epimad de 1988 à 2017, par sexe et par groupe d'âge (n=13 445).

Sexe	Classe d'âge	Incidence standardisée /10^5		APC % / an	p
		1988-1990	2015-2017		
Tous	<17 ans	2,2 [1,7 ; 2,7]	5,7 [5,0 ; 6,5]	**+4,3 [3,5 ; 5,1]*** **	**<0,0001**
	17-39 ans	10,6 [9,7 ; 11,5]	16,7 [15,6 ; 17,8]	**+1,9 [1,5 ; 2,2]*** **	**<0,0001**
	40-59 ans	3,8 [3,2 ; 4,6]	5,1 [4,5 ; 5,8]	**+0,9 [0,3 ; 1,5]****	**0,006**
	≥60 ans	2,6 [2,0 ; 3,4]	2,4 [2,0 ; 3,0]	+0,1 [-1,0 ; 1,1]	0,893
	Tous	5,1 [4,8 ; 5,5]	7,9 [7,4 ; 8,3]	**+1,9 [1,6 ; 2,2]*** **	**<0,0001**
Femmes	<17 ans	2,1 [1,5 ; 2,9]	4,9 [4,0 ; 6,0]	**+3,9 [2,8 ; 5,1]*** **	**<0,0001**
	17-39 ans	12,9 [11,5 ; 14,3]	18,2 [16,6 ; 20,0]	**+1,8 [1,4 ; 2,2]*** **	**<0,0001**
	40-59 ans	4,1 [3,1 ; 5,2]	5,8 [4,9 ; 6,9]	**+1,3 [0,5 ; 2,2]***	**0,002**
	≥60 ans	2,8 [2,0 ; 3,9]	2,3 [1,7 ; 3,0]	-0,5 [-1,7 ; 0,8]	0,477
	Tous	5,9 [5,4 ; 6,5]	8,3 [7,7 ; 8,9]	**+1,7 [1,3 ; 2,1]*** **	**<0,0001**
Hommes	<17 ans	2,2 [1,6 ; 3,0]	6,5 [5,5 ; 7,8]	**+4,6 [3,6 ; 5,5]*** **	**<0,0001**
	17-39 ans	8,3 [7,2 ; 9,4]	15,1 [13,6 ; 16,7]	**+2,0 [1,5 ; 2,4]*** **	**<0,0001**
	40-59 ans	3,6 [2,7 ; 4,7]	4,4 [3,5 ; 5,3]	+0,4 [-0,4 ; 1,2]	0,347
	≥60 ans	2,4 [1,4 ; 3,9]	2,5 [1,9 ; 3,4]	+0,8 [-0,6 ; 2,3]	0,244
	Tous	4,4 [3,9 ; 5,0]	7,4 [6,8 ; 8,0]	**+2,0 [1,6 ; 2,5]*** **	**<0,0001**

Rectocolite Hémorragique

En ce qui concerne la RCH, les tendances temporelles différaient significativement selon le sexe (interaction année*sexe : p=0,006). L'augmentation de l'incidence de la RCH était significativement plus importante chez les femmes au cours de la période d'étude (APC : +1,9 % [1,3 ; 2,6] chez la femme, p<0,0001 versus +0,8 % [0,2 ; 1,3] chez l'homme, p=0,006), l'incidence chez la femme rejoignant celle chez l'homme à la fin de la période d'étude (Figure 17, panel C). Le ratio femmes/hommes a augmenté au fil du temps, passant de 0,7 en 1988-1990 à 0,9 en 2015-2017 (test de tendance : p<0,0001).

Les tendances temporelles de la RCH différaient significativement selon la classe d'âge et le sexe. Chez les moins de 17 ans, les tendances temporelles n'étaient pas différentes selon le sexe (interaction année*sexe : p=0,591) avec un APC de +5,8 % [4,0 : 7,5] chez les hommes (p<0,0001) et +5,2 % [3,9 ; 6,6] chez les femmes (p<0,0001). Les tendances temporelles étaient significativement différentes selon le sexe dans les groupes d'âge des 17-39 ans (interaction année*sexe : p<0,001) et 40-59 ans (p=0,003). Les taux d'incidence augmentaient de manière significative chez les femmes dans les groupes d'âge 17-39 ans et 40-59 ans (de +2,1 % par an [1,7 ; 2,6], p<0,0001, et de 1,7 % par an [1,0 ; 2,5], p<0,0001, respectivement). Chez les hommes, les taux d'incidence n'augmentaient que dans le groupe des 17-39 ans, de +0,9 % [0,4 ; 1,4] (p=0,001), et restaient stables dans le groupe des 40-59 ans (+0,1 % [-0,6 ; 0,7], p=0,883). Dans le groupe des 60 ans et plus, les tendances temporelles étaient stables et ne différaient pas de manière significative en comparant les hommes et les femmes (interaction année*sexe : p=0,102 ; APC : -0,8 % [-1,7 ; 0,2] chez les hommes ; p=0,118 ; +0,4 % [-0,7 ; 1,5] chez les femmes ; p=0,504) (Table 11 et figure 18).

Table 11. Evolutions temporelles des taux d'incidence de RCH à partir des données du registre Epimad de 1988 à 2017, par sexe et par groupe d'âge (n=8 803).

Sexe	Classe d'âge	Incidence standardisée /10^5		APC % / an	p
		1988-1990	2015-2017		
Tous	<17 ans	0,8 [0,6 ; 1,2]	2,7 [2,2 ; 3,3]	**+5,4 [4,3 ; 6,6]*** **	**<0,0001**
	17-39 ans	7,4 [6,7 ; 8,2]	10,6 [9,7 ; 11,5]	**+1,5 [1,2 ; 1,9]*** **	**<0,0001**
	40-59 ans	4,9 [4,2 ; 5,8]	6,1 [5,4 ; 6,8]	+0,7 [0,1 ; 1,2]*	0,014
	≥60 ans	3,3 [2,6 ; 4,1]	3,4 [2,9 ; 4,1]	-0,2 [-1 ; 0,5]	0,53
	Tous	4,5 [4,1 ; 4,9]	6,1 [5,7 ; 6,5]	**+1,3 [0,9 ; 1,7]*** **	**<0,0001**
Femmes	<17 ans	1,0 [0,6 ; 1,5]	3,0 [2,3 ; 3,9]	**+5,2 [3,9 ; 6,6]*** **	**<0,0001**
	17-39 ans	6,9 [5,9 ; 8,0]	11,0 [9,8 ; 12,4]	**+2,1 [1,7 ; 2,6]*** **	**<0,0001**
	40-59 ans	3,1 [2,3 ; 4,1]	5,2 [4,3 ; 6,2]	**+1,7 [1 ; 2,5]*** **	**<0,0001**
	≥60 ans	2,1 [1,5 ; 3,1]	2,6 [2,0 ; 3,4]	+0,4 [-0,7 ; 1,5]	0,504
	Tous	3,6 [3,1 ; 4,0]	5,8 [5,3 ; 6,3]	**+1,9 [1,3 ; 2,6]*** **	**<0,0001**
Hommes	<17 ans	0,7 [0,4 ; 1,2]	2,4 [1,8 ; 3,2]	**+5,8 [4 ; 7,5]*** **	**<0,0001**
	17-39 ans	7,9 [6,8 ; 9,1]	10,2 [9 ; 11,5]	**+0,9 [0,4 ; 1,4]***	**0,001**
	40-59 ans	6,9 [5,6 ; 8,3]	7,0 [6,0 ; 8,2]	+0,1 [-0,6 ; 0,7]	0,883
	≥60 ans	4,9 [3,7 ; 6,7]	4,4 [3,5 ; 5,5]	-0,8 [-1,7 ; 0,2]	0,118
	Tous	5,5 [5,0 ; 6,2]	6,4 [5,9 ; 7,0]	**+0,8 [0,2 ; 1,3]***	**0,006**

Figure 17. Évolution dans le temps des taux d'incidence standardisés de la MC (n=13 445) et de la RCH (n=8 803) dans le registre Epimad de 1988 à 2017, par sexe et par classe d'âge. Chaque point correspond à la valeur moyenne pour une période de 3 ans. A) Taux d'incidence de MC selon le sexe. B) Taux d'incidence de MC selon l'âge. C) Taux d'incidence de RCH selon le sexe. D) Taux d'incidence de RCH selon l'âge.

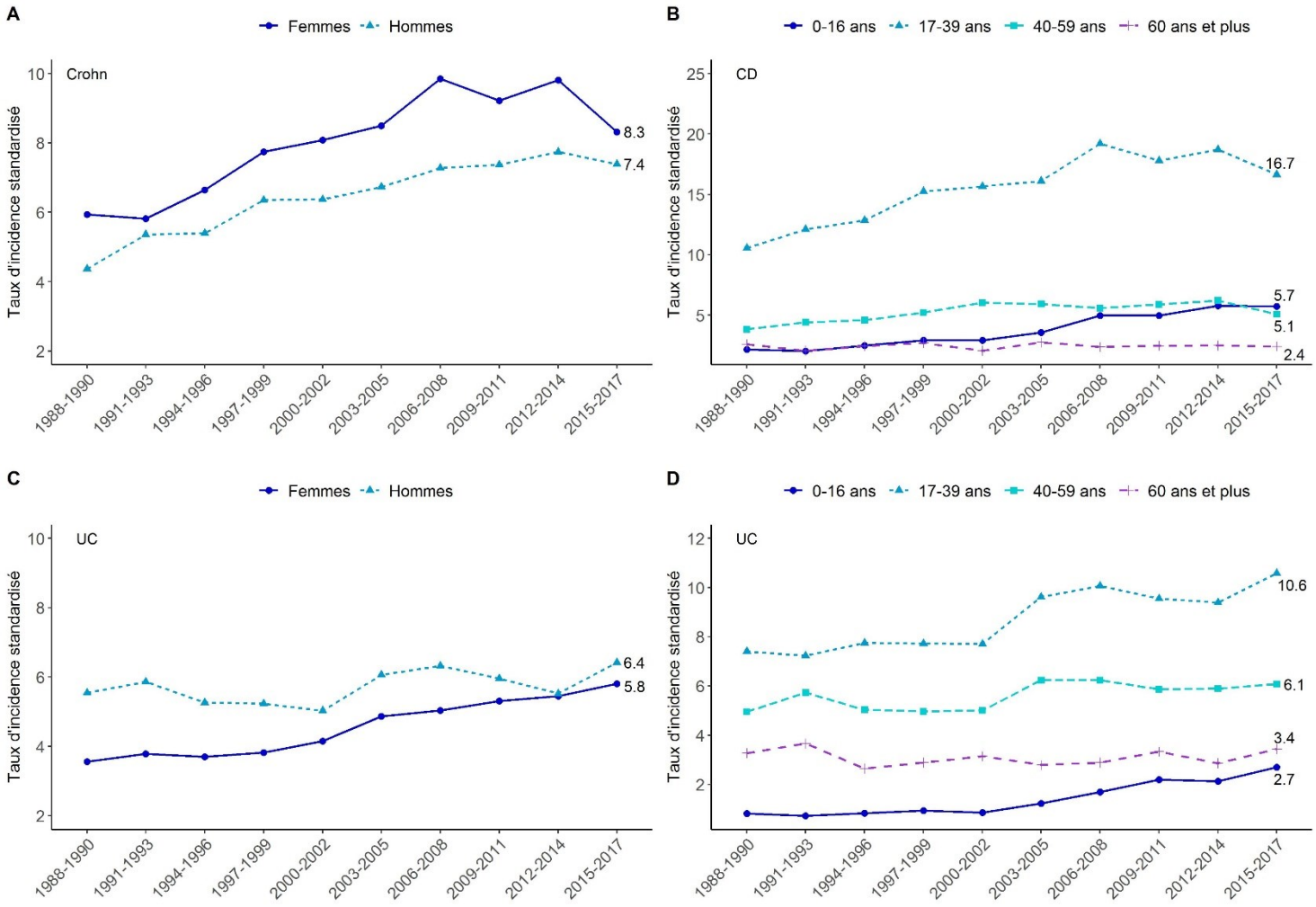
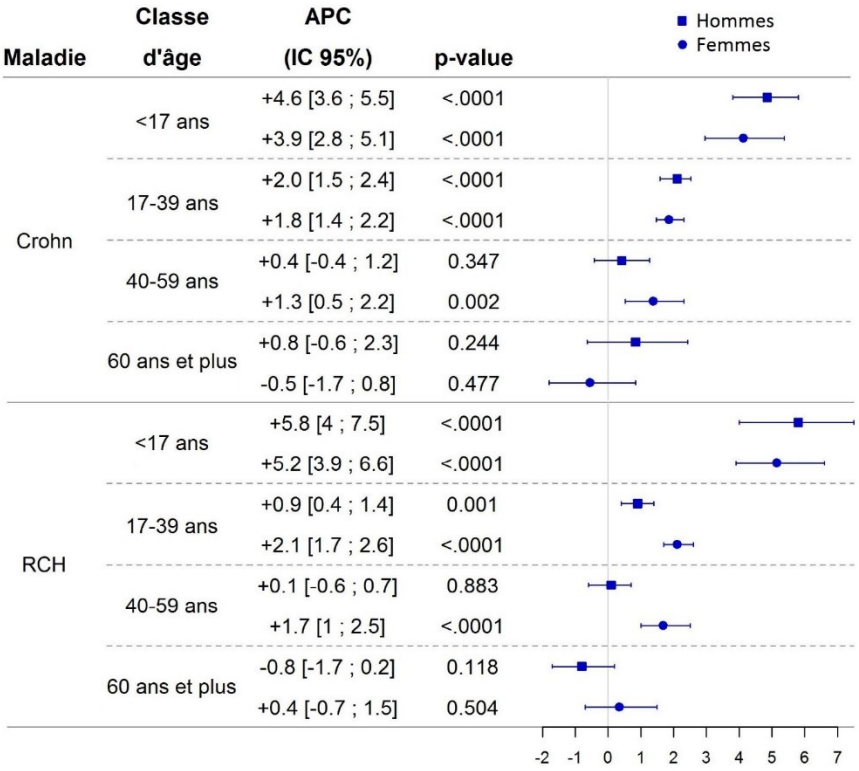


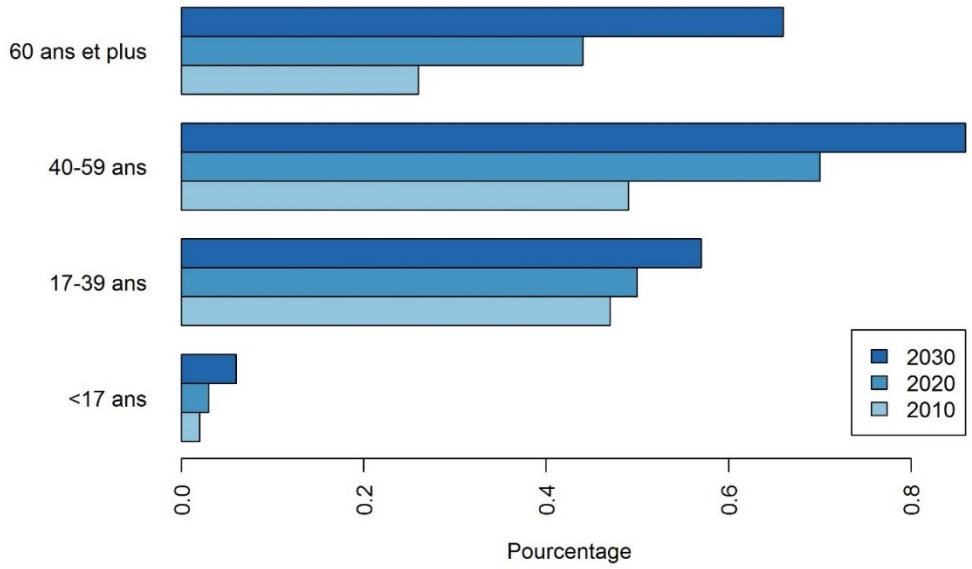
Figure 18. Pourcentage de variation annuel (APC en % par an) sur la période 1988-2017 chez les patients atteints de MC (n=13 445) et les patients atteints de RCH (n=8 803) dans le registre Epimad, par classe d'âge et par sexe.



3.3. Estimation de la Prévalence

La prévalence des MICI estimée était de 0,31 % de la population totale en 2010 et de 0,43 % en 2020. En projetant les résultats présentés ci-dessus, la prévalence devrait atteindre 0,57 % en 2030 (soit une augmentation d'environ 30 % des cas prévalents en 10 ans). À noter que l'augmentation de la prévalence s'accompagne également d'un vieillissement de la population atteinte de MICI (Figure 19).

Figure 19. Prévalence de MICI selon l'âge en 2010, 2020 et 2030 dans la zone du registre Epimad.

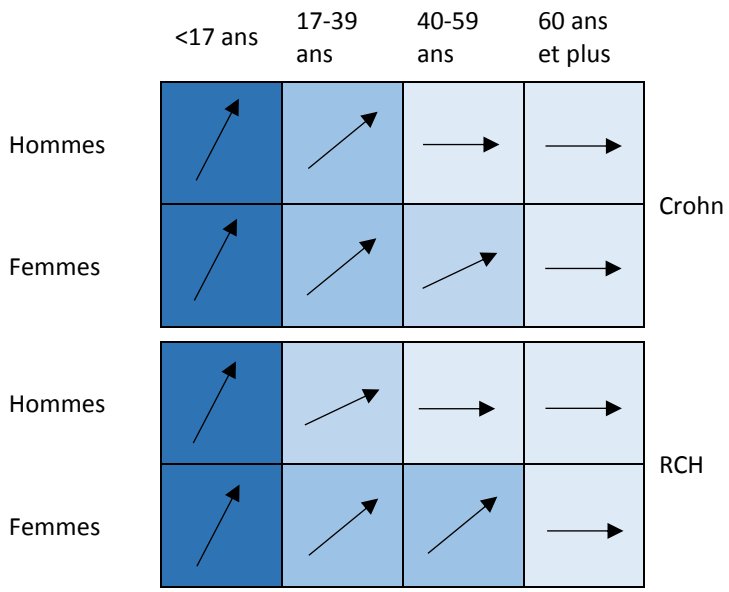


4. Discussion

RESUME DES PRINCIPAUX RESULTATS

Dans cette étude en population générale, incluant 22 879 cas incidents de MICI sur une période de 30 ans (1988-2017), les incidences annuelles de MICI, de MC et de RCH étaient respectivement de 12,7, 7,2 et 5,1 pour 10^5 personnes-années. L'incidence de la MC a augmenté régulièrement de 5,1 à 7,9 (APC : +1,9 % [1,6 ; 2,2]) et celle de la RCH est passée de 4,5 à 6,1 (APC : +1,3 % [0,9 ; 1,7]). L'augmentation de l'incidence était particulièrement marquée chez les enfants et les jeunes adultes, tant pour la MC que pour la RCH (Figure 20). Dans la RCH, l'incidence augmentait de manière plus prononcée chez les femmes que chez les hommes, les incidences chez les femmes rejoignant celles des hommes en fin de période d'étude. Sur la base de ces résultats, nous estimons qu'en 2030, près de 0,6 % de la population du nord de la France sera atteinte de MICI, avec un vieillissement de la population prévalente atteinte de MICI.

Figure 20. Résumé de tendances temporelles observées sur la période 1988-2017, selon l'âge, le sexe et le type de MICI.



INCIDENCES : COMPARAISON AUX DONNEES DE LA LITTERATURE

Dans la littérature, les taux d'incidence en Europe varient de 0,4 à 22,8 pour 10^5 pour la MC et de 2,4 à 44,0 pour 10^5 pour la RCH avec une incidence moyenne de 6.3 pour la MC et de 9.8 pour la RCH dans les 14 pays de l'Ouest de l'Europe (41,42). Les valeurs d'incidences observées dans notre étude sont donc en accord avec ce qui est décrit plus largement en Europe. L'incidence dans notre région est cependant légèrement plus élevée que la moyenne pour la MC et plus faible pour la RCH. En effet, dans la zone du registre Epimad, le ratio MC/RCH est supérieur à 1. Cette valeur contraste avec la plupart des autres études réalisées en Europe occidentale, où la RCH apparait plus fréquente que la MC (42,45,49,171,178). Cependant, avec l'augmentation de l'incidence de la RCH, notamment chez la femme, dans notre étude, l'incidence de la RCH pourraient prochainement rejoindre puis dépasser celle de la MC. Ce ratio MC/RCH supérieur à 1 est également observé dans d'autres pays notamment en Belgique et en Nouvelle-Zélande (40,56).

Depuis les années 1990, les études épidémiologiques menées dans les pays occidentaux montrent que les tendances temporelles ont changé : plus des deux tiers des études sur la MC ou la RCH ont rapporté des incidences stables ou en baisse (35). Selon Kaplan et al, les pays occidentaux entrent, depuis le début des années 2000, dans le stade épidémiologique appelé « compound prevalence », caractérisée par une stagnation du taux d'incidence et une

augmentation de la prévalence (33). Dans une étude canadienne récente, les incidences de la MC et de la RCH ont diminué de manière significative entre 1990 et 2012, et cela dans les deux genres (36). Cette tendance à la baisse a également été observée dans une étude récente en Suède (37). Ce résultat contraste avec notre observation actuelle d'une augmentation de l'incidence globale des MICI dans le nord de la France. Cependant, des études récentes ont également souligné une augmentation continue de l'incidence, notamment au Danemark et aux îles Féroé, pays dont les incidences sont pourtant parmi les plus élevées (45,49,171). Chez les enfants, une récente revue systématique d'études réalisées en population générale a indiqué que 84 % des études rapportaient des augmentations significatives d'incidence (71).

PREVALENCE : COMPARAISON AUX DONNEES DE LA LITTERATURE

Bien que les incidences de la MC et de la RCH semblent s'être stabilisées dans certains pays occidentaux, la prévalence globale des MICI continue d'augmenter en raison d'un faible taux de mortalité et du vieillissement de la population atteinte de MICI. Notre prévalence estimée à près de 0,6 % de personnes atteintes de MICI en 2030 est en accord avec les publications récentes. La prévalence mondiale des MICI en Europe est actuellement estimée à 0,2 % (41), mais varie selon les pays en relation avec le gradient Est-Ouest et Sud-Nord des incidences (46). Un modèle prédictif développé au Canada a estimé la prévalence des MICI dans ce pays à 0,6 % en 2015 et une prévalence prédite de 0,9 % d'ici à 2025 (66). Les résultats d'une étude écossaise basés sur une méthode de capture-recapture suggéraient que la prévalence des MICI en Écosse atteindrait 1 % dans les 10 prochaines années (67). D'après les estimations du « Global Disease Burden » en 2017, la prévalence des MICI serait d'environ 0,45 % aux États-Unis et au Royaume-Uni (34). Selon Kaplan et al., la prévalence est censée doubler en 20-25 ans (33). Nos estimations sont en accord avec cette hypothèse, avec un quasi-doublement de la prévalence entre 2010 et 2030. Notre étude décrit également un vieillissement de la population prévalente. Ce vieillissement a également été rapporté dans la littérature. Il doit être pris en considération compte tenu de la sous-représentation de cette population âgée dans les essais cliniques. En effet, les interactions médicamenteuses et effets secondaires liés aux multiples prescriptions, les comorbidités et la fragilité de certains de ces séniors, les coûts plus élevés liés à l'hospitalisation, sont des évènements bien connus et bien décrits dans cette population de séniors (67,179). Une revue de la littérature récente a mis en évidence une exclusion des patients âgés dans 58 % des essais cliniques ainsi qu'une sous-représentation

dans les études permettant leur inclusion dans lesquelles les 65 ans et plus ne représentaient que 5,4 % du total des inclusions (180).

FACTEURS EXPLICATIFS

Un résultat majeur de notre étude est la variation des tendances temporelles des incidences de la MC et de la RCH en fonction de l'âge et du sexe. Les taux d'augmentation les plus élevés étaient observés chez les enfants, suivis des jeunes adultes (17-39 ans) tant pour la MC que pour la RCH et ce, dans les deux sexes, tandis que les incidences restaient stables chez les personnes de 60 ans et plus. Dans la population d'âge moyen (40 à 59 ans), les taux d'incidence n'augmentaient que chez les femmes, surtout dans la RCH. D'une manière générale, les incidences de RCH augmentaient plus fortement chez la femme. Shah et al. ont décrit des différences selon le sexe dans l'incidence des MICI en fonction de l'âge au moment du diagnostic (181). Ces résultats suggéreraient un possible rôle des hormones féminines dans le développement des MICI. Dans la MC, un changement du sex-ratio femmes/hommes vers l'âge de la puberté et vers l'âge de la ménopause a également été observé dans notre étude. Une analyse basée sur la cohorte « Nurses' Health Study » a par ailleurs suggéré que l'utilisation de contraceptifs oraux étaient associés à un sur-risque de MC multiplié par un facteur 3 (122).

La relation entre le tabagisme et les MICI est également bien connue, avec des effets opposés observés pour la MC et la RCH. Le tabagisme actif protège de la RCH, tandis que le statut d'ancien fumeur est associé à un risque accru - principalement au cours des cinq premières années après l'arrêt du tabac (129,182). L'augmentation de la consommation régulière de tabac chez les femmes depuis le début des années 1980 pourrait contribuer à l'augmentation de l'incidence de la MC mais pas de celle de la RCH (183). En revanche, les femmes arrêtent de fumer plus tôt dans la vie en raison de leurs grossesses. Cela pourrait jouer un rôle dans l'augmentation de l'incidence de la RCH chez les jeunes femmes.

L'appendicectomie est rapportée comme étant associée à une diminution de 69 % du risque de RCH (115,116). En 1997, en France, l'appendicectomie était plus fréquente chez les femmes que chez les hommes (sex-ratio hommes/femmes = 0,85). L'incidence de l'appendicectomie a fortement diminué au cours des deux dernières décennies, en particulier chez les femmes, avec une inversion du sex-ratio hommes/femmes (1,05 en 2012) (184). Ce facteur pourrait

être associé à une augmentation du risque de développer une RCH, en particulier chez les femmes.

L'augmentation de l'incidence des MICI pourrait également être liée à des modifications de facteurs environnementaux au sens large comme ils ont été décrits dans l'introduction générale.

Enfin, l'augmentation de l'incidence des MICI dans les deux sexes dans les deux sexes pourrait être, au moins en partie, associée à une modification de la prise en charge des patients. En effet, le développement de nouveaux outils diagnostiques pourrait avoir conduit à un diagnostic de MICI plus précoce et plus répandu. Cependant, dans notre étude, l'intervalle de temps entre le début des symptômes et le diagnostic de la MICI n'a pas été modifié sur une période de 30 ans, suggérant que les modalités diagnostiques sont restées plus ou moins stables au cours de la période étudiée. Par ailleurs, la MC et la RCH présentent des évolutions temporelles différentes, l'augmentation de l'incidence de la MC ayant précédé celle de la RCH.

EVOLUTION TEMPORELLE DES LOCALISATIONS

Entre 1988 et 2017, la proportion de MC avec atteinte iléocolique (L3) au diagnostic a augmenté de manière significative, passant de 41 % à 57 %. Il est important de noter que cette tendance persistait même en ne considérant que les patients ayant eu une exploration complète du côlon et de l'iléon au moment du diagnostic, ou en ne prenant en compte que les patients ayant subi une entérographie par scanner et/ou IRM (données disponibles depuis 2006 uniquement). Ces variations de localisation de la maladie sur une période de 30 ans sont donc susceptibles de refléter, au moins en partie, une réelle augmentation de l'atteinte iléale de la maladie.

FORCES ET FAIBLESSES

Les principales forces de cette étude résident dans sa conception basée sur un registre épidémiologique en population générale, la grande taille de l'échantillon (22 879 cas incidents de MICI), un enregistrement exhaustif sur une période de 30 ans et l'utilisation de critères diagnostiques validés et publiés. Toutes les données au diagnostic des cas incidents ont été expertisées par deux gastro-entérologues experts. La principale limite concerne l'absence de données sur le statut tabagique ou l'utilisation des contraceptifs oraux ou traitement hormonal substitutif. Enfin l'analyse de la prévalence repose sur plusieurs hypothèses : i) une

augmentation de l'incidence depuis 1975 ; ii) une croissance constante de l'incidence après 2018 et ; iii) un choix du scénario d'évolution de la population de la région concernée. Cependant, l'idée était de donner un ordre de grandeur de cette prévalence dans la zone géographique du registre et d'alerter sur cette augmentation de la prévalence. Des travaux ultérieurs permettront d'affiner ces résultats en faisant varier ces hypothèses ou à l'aide de modèles plus complexes.

Chapitre 2 : Impact sur le niveau d'études et l'insertion professionnelle des MICI à début pédiatrique

1. Contexte et objectifs

Les MICI impactent significativement les patients tout au long de leur vie. Comme pour toute maladie chronique, la MICI diagnostiquée chez un enfant ou un adolescent a un impact sur leur développement physique, émotionnel et social. De plus, les adolescents sont touchés par la maladie pendant une période critique de leur éducation et de l'établissement de leurs objectifs professionnels. Il a été documenté que les patients atteints de MICI ont, comme chez l'adulte, une qualité de vie altérée (185,186). Ces patients font également l'objet d'un plus grand nombre d'hospitalisations liées à la maladie que les patients adultes (187). Ceci entraine un fardeau significatif lié à la maladie tout au long de leur vie et ce d'autant plus longuement que la maladie s'est déclarée tôt dans la vie.

La morbidité sociale dans la MICI à début pédiatrique est peu connue. On peut supposer que les absences récurrentes et parfois prolongées à l'école associées aux exacerbations de la maladie auront des effets néfastes importants sur l'éducation et l'emploi à l'âge adulte.

Les objectifs de notre étude étaient d'évaluer les niveaux d'éducation et l'insertion professionnelle chez les patients jeunes adultes atteints de MICI à début pédiatrique par rapport à la population générale du même âge et du même sexe.

2. Matériel et Méthodes

2.1. Population

Les patients inclus ont été extraits du registre Epimad avec un diagnostic certain ou probable de MICI posé avant l'âge de 17 ans. Afin d'éviter d'inclure des étudiants qui n'auraient pas terminé leurs études, seuls les patients âgés de 25 ans ou plus au moment de l'étude ont été inclus.

2.2. Méthodologie et données collectées

Les patients ont été contactés par voie postale et invités à remplir un auto-questionnaire concernant le dernier diplôme, l'âge au dernier diplôme, le statut professionnel, la profession, la catégorie socio-professionnelle, le statut éventuel d'invalidité au travail, les traitements passés et actuels, les interventions chirurgicales antérieures liées à la MICI, la présence d'une stomie. L'activité de la maladie était mesurée par l'indice de Harvey-Bradshaw (HBI) pour les patients atteints de MC et le Simple Clinical Colitis Activity Index (SCCAI) pour les patients atteints de RCH. La qualité de vie mesurée à l'aide du questionnaire Short-Inflammatory Bowel Disease Questionnaire (SIBDQ) spécifique aux MICI (188–190). Les patients ont également été interrogés sur l'impact de la maladie sur le choix de leurs études, la progression de leurs études et le choix de leur profession.

Une maladie active a été définie par un HBI>3 dans la MC ou un SCCAI>2 dans la RCH. Une qualité de vie "faible" a été définie par un score SIBDQ <45, une qualité de vie "normale" par un score SIBDQ entre 45 et 60, et une qualité de vie "élevée" par un score SIBDQ supérieur à 60.

Les professions ont été classées selon la classification des professions et des catégories socio-professionnelles (PCS) de 2003 de l'INSEE.

Les données sur la MICI et ses caractéristiques au diagnostic ont été recueillies via le dossier médical du patient et extraites de la base de données du registre Epimad. Ces données concernaient : l'âge, le sexe, la date diagnostique, le délai diagnostique, la localisation de la maladie, la présence ou non de lésions ano-périnéales, les signes extra-digestifs, les antécédents familiaux de MICI.

Les patients ont répondu à l'étude entre novembre 2019 et juin 2021. Après un minimum de deux mois, les patients n'ayant pas répondu au questionnaire ont été contactés à nouveau par voie postale afin d'obtenir leur participation à cette étude. Aucune relance supplémentaire n'a été effectuée en l'absence de réponse aux 2 premiers contacts.

Cette étude a obtenu l'accord du Comité de Protection des Personnes Nord-Ouest IV (IRB 2017 A003397 46).

2.3. Définitions

Pour les questions sur les diplômes et les professions, nous avons utilisé, en partie, le questionnaire du recensement de la population de l'INSEE afin de pouvoir disposer de données comparables dans la population générale de la même zone géographique. Nous avons également utilisé les définitions et catégorisations de l'INSEE.

Le niveau d'éducation a été défini comme le diplôme le plus élevé obtenu et divisé en 7 groupes :

- Aucun diplôme ou certificat d'études primaires.
- BEPC, brevet des collèges, Diplôme National du Brevet.
- CAP, BEP ou équivalent.
- Baccalauréat, brevet professionnel ou équivalent.
- Diplôme de l'enseignement supérieur de niveau Bac +2.
- Diplôme de l'enseignement supérieur de niveau Bac + 3 ou Bac + 4.
- Diplôme de l'enseignement supérieur de niveau Bac +5 ou plus.

L'insertion professionnelle a été définie par la situation principale des participants au moment de l'étude : employé, apprenti ou stagiaire rémunéré, étudiant ou stagiaire non rémunéré, chômeur, retraité, parent au foyer.

La population active au sens du recensement de la population comprend les personnes qui déclarent :

- Exercer une profession (salariée ou non) même à temps partiel.
- Aider une personne dans son travail (même sans rémunération).
- Être apprenti, stagiaire rémunéré.
- Être chômeur à la recherche d'un emploi ou exerçant une activité réduite.
- Être étudiant ou retraité mais occupant un emploi.

Cette population correspond donc à la population active occupée à laquelle s'ajoutent les chômeurs en recherche d'emploi (Figure 21). Ne sont pas retenues les personnes qui, bien que s'étant déclarées au chômage, précisent qu'elles ne recherchent pas d'emploi.

La figure 21 illustre ces différentes définitions.

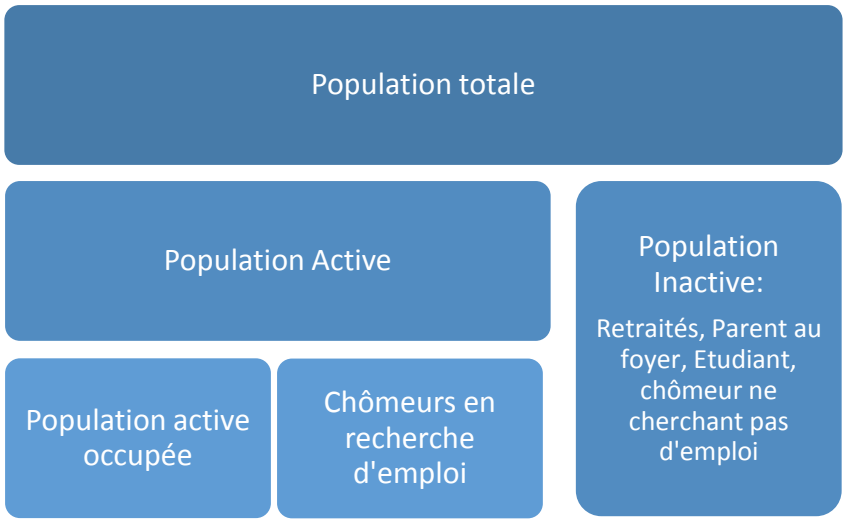
Le taux de chômage au sens du recensement est le rapport entre le nombre de chômeurs en recherche d'emploi et le nombre total de personnes actives (en emploi + au chômage).

Le taux d'activité est le rapport entre la population active et la population totale.

La répartition des emplois a été réalisée selon les 8 catégories socio-professionnelles définies par l'INSEE :

- Agriculteurs exploitants.
- Artisans, commerçants, chefs d'entreprise.
- Cadres et professions intellectuelles supérieures.
- Professions intermédiaires.
- Employés.
- Ouvriers.
- Retraités.
- Autres personnes sans activité professionnelle.

Figure 21. Composition de la population en population active et inactive.



L'emploi dans le secteur de la santé incluait les médecins, dentistes, pharmaciens, infirmiers et autres professions de santé. Ce secteur a été défini à partir des codes PCS débutant par « 311 », « 344 », « 431 », « 432 », « 433 », « 526 » ainsi que le code 525D.

2.4. Données de référence

Les données de référence pour la population générale de même âge, sexe et zone géographique, ont été obtenues auprès de l'INSEE, à partir du recensement de la population pour les données de référence sur le niveau d'éducation, le statut professionnel, la catégorie

socio-professionnelle, le taux d'activité et le taux de chômage ; à partir de la base « tous salariés » de l'INSEE (données individuelles sur chaque salarié, produites à partir des déclarations administratives des employeurs) pour l'emploi dans le secteur public et l'emploi dans le secteur de la santé.

L'année de référence pour ces données était l'année 2019.

2.5. Méthodes statistiques

Les variables quantitatives ont été décrites par la médiane et l'intervalle interquartile (IQR). Les variables qualitatives ont été décrites par la fréquence et le pourcentage.

Afin d'évaluer le biais de non-réponse, les caractéristiques cliniques et démographiques des patients au moment du diagnostic ont été comparées entre les répondants et les non-répondants. Pour évaluer les différences éventuelles de niveau socio-économique, en l'absence de données individuelles, nous avons comparé l'indice de déprivation FDep09 (French Deprivation Index) et le niveau d'urbanisation (grille de densité communale de l'Insee) du lieu de résidence du patient au moment du diagnostic en fonction du statut de répondant ou non-répondant. L'indice FDep09 a été construit par Rey et al. en utilisant quatre variables issues de la base de données de l'INSEE : i) le revenu médian des ménages, ii) la proportion de diplômés du secondaire dans la population âgée de 15 ans et plus, iii) la proportion d'ouvriers qualifiés dans la population active et iv) le taux de chômage (191). Plus l'indice FDep09 est élevé, plus le niveau de déprivation est important.

Afin de comparer les données de l'échantillon des répondants à notre étude aux données de référence, nous avons ajusté nos données pour les rendre concordantes avec la distribution observée dans la population de référence de l'INSEE en termes d'âge et de sexe (données de population de l'INSEE issues du recensement de la population de 2019) par post-stratification. L'idée est de corriger les biais possibles dus à des différences entre l'échantillon étudié et la population totale. Un poids a ainsi été attribué à chaque individu de l'échantillon en fonction de la strate démographique à laquelle il appartient en termes d'âge et de sexe.

Les taux bruts et les taux ajustés tenant compte de ces poids sont présentés dans les tableaux. Dans le texte, seuls les taux ajustés sont donnés.

La comparaison de la distribution des variables observées dans notre étude chez les patients atteints de MICI à début pédiatrique et la distribution dans la population générale a été

réalisée à l'aide de tests du χ^2 de comparaison à une proportion théorique en tenant compte des poids de post-stratification. Les proportions théoriques étaient celles observées dans les données de référence de l'Insee.

L'association entre un niveau d'études supérieures (défini comme un niveau d'études supérieur au baccalauréat) et les caractéristiques de la maladie au diagnostic, l'activité de la maladie, la qualité de vie, les traitements et la chirurgie a été étudiée à l'aide de régressions logistiques univariées. Les résultats sont présentés en termes d'Odds Ratio (OR) accompagnés de leur intervalle de confiance à 95 %.

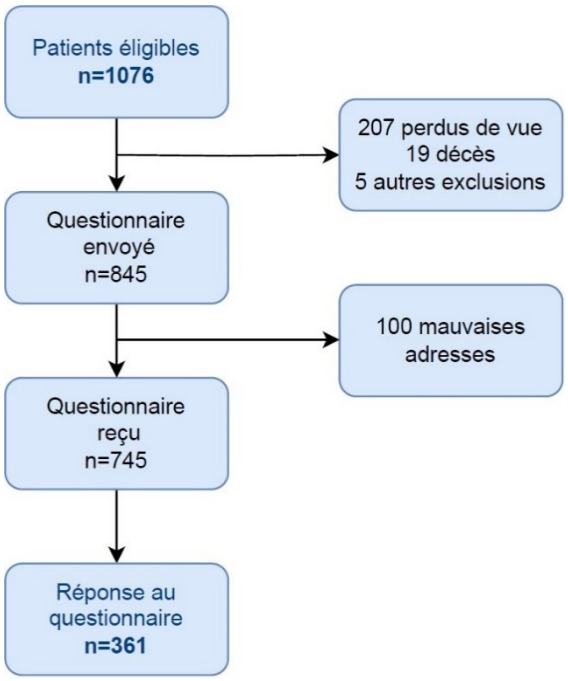
Les analyses statistiques ont été réalisées à l'aide du logiciel SAS 9.4 (SAS Institute Inc., Cary, NC, USA), et le seuil de significativité statistique a été fixé à $p \leq 0,05$.

3. Résultats

3.1. Taux de réponses

Au total, 1076 patients répondant aux critères d'inclusion ont été extraits de la base de données du registre Epimad. Parmi eux, 207 ont été perdus de vue, 19 étaient décédés depuis le diagnostic et 5 étaient dans une situation personnelle compliquée (incarcérés etc.) et n'ont donc pas été contactés (Figure 22). Au final, 844 patients ont été contactés par courrier postal. Parmi eux, 100 n'habitaient plus à l'adresse indiquée (retour de la poste) et 361 ont répondu au questionnaire, ce qui correspond à un taux de réponse de 34 % des patients éligibles, 43 % des patients contactés et 48 % des patients ayant reçu le questionnaire (en excluant les adresses erronées).

Figure 22. Diagramme de l'inclusion des patients dans l'étude.



3.2. Caractéristiques cliniques et démographiques des patients

Les caractéristiques cliniques et démographiques des répondants et des non-répondants sont présentées dans la Table 12. Quarante-sept pour cent (n = 170) des répondants étaient de sexe masculin. L'âge médian au moment du diagnostic et au moment de l'étude étaient de 15,0 ans (IQR : [12,9 ; 16,3]) et 34,2 ans ([29,6 ; 39,5]), respectivement. La durée médiane d'évolution de la maladie était de 20,7 ans ([15,7 ; 26,1]). Ces caractéristiques n'étaient pas différentes entre les répondants et les non-répondants.

Soixante-dix-neuf pour cent (n = 286) des répondants étaient atteints de MC, contre 74 % (n 527) des non-répondants (p=0,047). Dans la MC, les répondants présentaient significativement plus d'atteinte de digestive haute (L4) que les non-répondants : 32 % (n=93) contre 24 % (n=127) (p=0,010). Dans la RCH, les répondants présentaient une atteinte significativement plus étendue que les non-répondants, 12 %, 38 % et 49 % avaient une extension E1, E2 et E3, respectivement, chez les répondants, contre 31 %, 30 % et 38 % chez les non-répondants (p=0,007).

3.3. Caractéristiques au lieu d'habitation au moment du diagnostic

Nous avons ensuite comparé l'indice de déprivation et le niveau d'urbanisation du lieu de résidence au moment du diagnostic entre les répondants et les non-répondants afin d'évaluer un éventuel biais de non-réponse lié au statut socio-économique (Table 13). Nous n'avons observé aucune différence significative selon les quintiles de l'indice de déprivation entre les répondants et non-répondants (p=0,162). Mais, le niveau d'urbanisation différait significativement entre les répondants et les non-répondants (p=0,036). En effet, les répondants résidaient davantage en zones rurales que les non-répondants (28 % contre 22 %) et moins en zones densément peuplées (32 % contre 38 %).

Table 12. Caractéristiques cliniques et démographiques des patients et comparaison entre les répondants (n=361) et les non-répondants (n=715).

	Total (n=1076)	Répondants (n=361)	Non-répondants (n=715)	p-value
Sexe masculin	535 (49,7%)	170 (47,1%)	365 (51,0%)	0,220
Age au diagnostic*	15,0 [12,7 ; 16,2]	15,0 [12,9 ; 16,3]	15,1 [12,7 ; 16,2]	0,842
Durée de la maladie (ans)*	20,5 [15,4 ; 25,9]	20,7 [15,7 ; 26,1]	20,5 [15,3 ; 25,8]	0,666
Age au moment de l'étude*	34,2 [29,6 ; 39,5]	34,2 [29,6 ; 39,5]	34,2 [29,6 ; 39,5]	0,926
Type de MICI				
MC	813 (75,6%)	286 (79,2%)	527 (73,7%)	**0,047**
RCH	263 (24,4%)	75 (20,8%)	188 (26,3%)	
Antécédents familiaux	133 (12,4%)	54 (15,0%)	79 (11,0%)	0,066
Manifestations Extra-digestives	158 (14,7%)	63 (17,4%)	95 (13,3%)	0,068
MC (au diagnostic)	n=813	n=286	n=527	
Localisation				
L1	130 (16,6%)	51 (18,3%)	79 (15,7%)	
L2	197 (25,2%)	70 (25,1%)	127 (25,2%)	0,629
L3	456 (58,2%)	158 (56,6%)	298 (59,1%)	
Localisation haute (L4)	220 (27,1%)	93 (32,5%)	127 (24,1%)	**0,010**
Lésions ano-périnéales	50 (6,1%)	23 (8,0%)	27 (5,1%)	0,100
RCH (au diagnostic)	n=263	n=75	n=188	
Localisation				
E1	67 (26,0%)	9 (12,3%)	58 (31,3%)	
E2	84 (32,6%)	28 (38,4%)	56 (30,3%)	**0,007**
E3	107 (41,4%)	36 (49,3%)	71 (38,4%)	

** médiane [IQR]*

Table 13. Caractéristiques au lieu de résidence et comparaison entre les répondants (n=361) et les non-répondants (n=715).

	Total (n=1076)	Répondants (n=361)	Non-répondants (n=715)	p-value
Index de déprivation FDEP09				
1er quintile	137 (12,8%)	58 (16,2%)	79 (11,1%)	
2ème quintile	149 (13,9%)	46 (12,8%)	103 (14,4%)	
3ème quintile	137 (12,8%)	48 (13,4%)	89 (12,5%)	0,162
4ème quintile	152 (14,2%)	51 (14,2%)	101 (14,2%)	
5ème quintile	496 (46,3%)	155 (43,3%)	341 (47,8%)	
Urbanisation				
Densément peuplé	388 (36,2%)	115 (32,1%)	273 (38,3%)	
Densité intermédiaire	424 (39,6%)	141 (39,4%)	283 (39,7%)	**0,036**
Rural	259 (24,2%)	102 (28,5%)	157 (22,0%)	

3.4. Activité de la maladie et qualité de vie

Quarante-trois pour cent des patients (n=141) étaient en rémission au moment de l'étude, 52 % (n=168) présentaient une maladie active et 5 % (n=16) avaient une maladie sévère. La médiane de la qualité de vie était de 54 [45 ; 61]. Vingt-trois pour cent des patients (n=79) avaient une qualité de vie faible, 51 % (n=172) une qualité de vie normale et 26 % (n=88) une qualité de vie élevée.

3.5. Niveau d'éducation

Le niveau d'éducation a été étudié pour les patients ayant terminé leurs études (soit n=359 sur 361) et est présenté dans la Figure 23 et la Table 14. L'âge médian au dernier diplôme était de 22 ans (IQR : [19 ; 24]). Par rapport à la population générale, les participants avaient un niveau de diplôme plus élevé que celui en population générale, 57 % contre 41 % des participants détenaient un diplôme d'enseignement supérieur (p<0.0001).

Figure 23. Description du dernier diplôme obtenu chez les patients atteints de MICI débutant pendant l'enfance en comparaison avec la population générale.

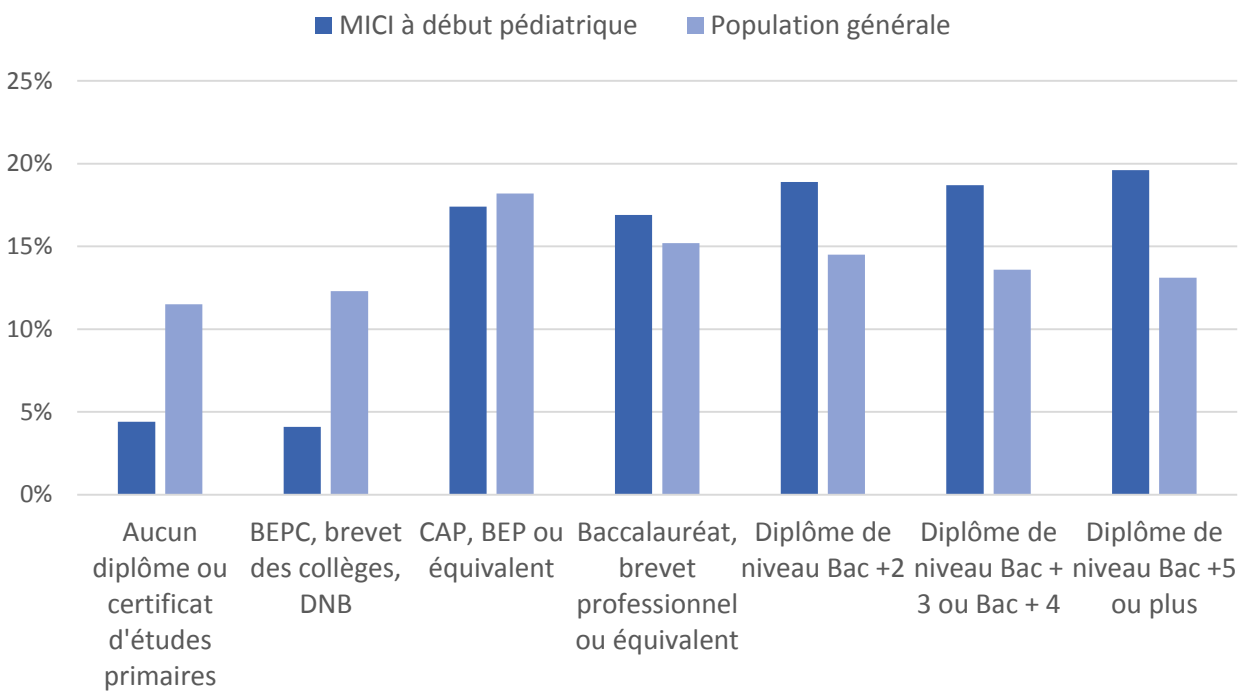


Table 14. Description du dernier diplôme obtenu chez les patients atteints de MICI débutant pendant l'enfance en comparaison avec la population générale.

	Age median au dernier diplôme (IQR)	Age attendu	Patients atteints de MICI à début pédiatrique (n=359)			Données de référence (%) (n=1789473)
			Nombre	Taux brut	Taux ajusté	
Aucun diplôme ou certificate d'études primaires	-		16	4,4%	4,4%	11,5%
BEPC, Brevet des collèges, DNB	15 [15 ; 15]	15	17	4,7%	4,1%	12,3%
CAP, BEP ou équivalent	19 [18 ; 22]	17	61	17,0%	17,4%	18,2%
Baccalauréat, brevet professionnel ou équivalent	19 [18 ; 20]	18-19	62	17,3%	16,9%	15,2%
Diplôme de niveau Bac+2	21 [20 ; 24]	20	70	19,5%	18,9%	14,5%
Diplôme de niveau Bac+3 ou Bac+4	23 [21 ; 25]	21-22	57	15,9%	18,7%	13,6%
Diplôme de niveau Bac+5 ou plus	24 [23 ; 26]	≥ 23	76	21,2%	19,6%	13,1%

3.6. Situation professionnelle

La situation professionnelle principale des patients est présentée dans la Table 15. La majorité des patients avait un emploi au moment de l'étude, représentant 82 % des participants (n=294). Cinq de ces patients étaient en congé maladie ou en congé parental au moment de l'étude.

Table 15. Situation professionnelle principale des patients au moment de l'étude (n=361).

		Patients atteints de MICI à début pédiatrique (n=361)		
		Nombre	**Taux brut**	**Taux ajusté**
Population active	Ayant un emploi	294	81,5 %	82,3%
	Apprentis, stages rémunérés	2	0,5 %	0,5 %
	En recherche d'emploi	31	8,6 %	7,9 %
Population inactive	Etudiants, stages non rémunérés	2	0,5%	0,4 %
	Autres sans emploi*	32	8,9 %	8,9 %

** incluant les parents au foyer, les personnes en invalidité.*

La population active comprenait 91 % des patients (n=327), taux similaire à celui observé dans la population générale du même âge (90 %, p=0,578) (Table 16). Parmi la population active, le taux de chômage était significativement plus faible dans la population de l'étude que dans la population générale (9 % des patients atteints de MICI à début pédiatrique versus 15 % dans la population générale, p=0,001). Au sein de la population active occupée, 91 % (n=271) occupaient un emploi salarié et 82 % (n=238) travaillaient à temps plein, de manière similaire à 92 % de salariés et 84 % travaillant à temps plein dans la population générale (p=0,927 et p=0,172, respectivement).

Les catégories socio-professionnelles des patients sont présentées dans la Figure 24 et la Table 17. La distribution des catégories socio-professionnelles des patients dans l'étude était significativement différente de celle de la population générale (p<0,0001). En effet, les cadres et les professions intellectuelles supérieures ainsi que les professions intermédiaires étaient surreprésentés chez les patients atteints de MICI à début pédiatrique par rapport à la population générale, avec respectivement 22 % et 41 % chez les patients atteints de MICI en comparaison à respectivement 16 % et 28 % dans la population générale.

Table 16. Comparaison du statut professionnel entre les participants à l'étude et la population de référence.

	Patients atteints de MICI à début pédiatrique (n=361)			Données de référence (n=1 827 384)	p
	Nombre	Taux brut	Taux ajusté		
Population active	327	90,6 %	90,7 %	89,8 %	0,578
Sans emploi	31	9,5 %	8,7 %	15,0 %	0,001
Occupée	296	90,5 %	91,3 %	85,0 %	
Emploi salarié[a]	271	91,5 %	90,5 %	91,7 %	0,927
Travail à temps plein[b]	238	81,8 %	81,6 %	84,5 %	0,172

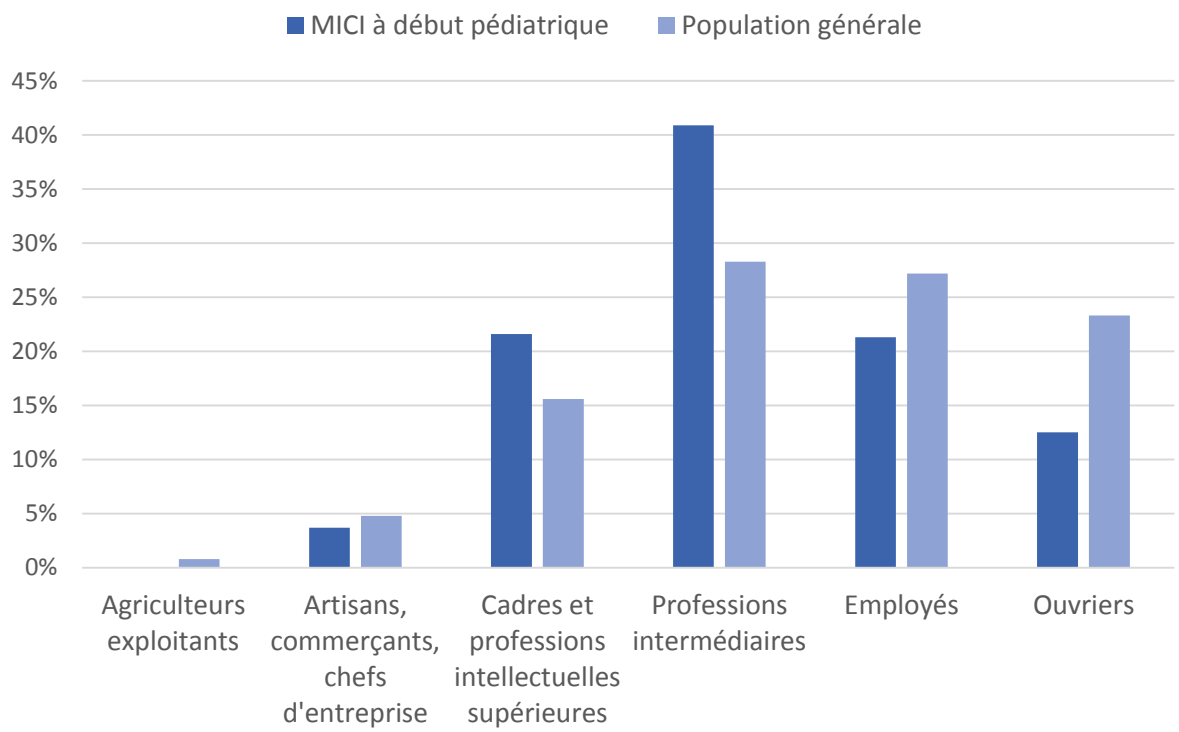
[a] *Le taux d'emploi salarié est calculé par rapport à la population active ayant un emploi (occupée).*
[b] *5 données manquantes. Le taux d'emploi à temps plein est calculé par rapport à la population active ayant un emploi (occupée).*

Table 17. Description des catégories socio-professionnelles dans la population active occupée chez les patients atteints MICI débutant pendant l'enfance en comparaison avec la population générale.

	Patients atteints de MICI à début pédiatrique (n=361)			Données de référence (n=1 396 041)	p
	Nombre	Taux brut	Taux ajusté		
Agriculteurs exploitants	0	0,0 %	0,0 %	0,8 %	
Artisans, commerçants, chefs d'entreprise	10	3,4 %	3,7 %	4,8 %	
Cadres et professions intellectuelles supérieures	66	22,4 %	21,6 %	15,6 %	<0,0001
Professions intermédiaires	119	40,3 %	40,9 %	28,3 %	
Employés	62	21,0 %	21,3 %	27,2 %	
Ouvriers	38	12,9 %	12,5 %	23,3 %	
Population active occupée totale	295[†]			1 396 041	

[†] *1 donnée manquante.*

Figure 24. Description des catégories socio-professionnelles dans la population active occupée chez les patients atteints MICI débutant pendant l'enfance en comparaison avec la population générale.



Au total, 115 (32 %) patients étaient reconnus comme invalides (selon la sécurité sociale ou la CDAPH, Commission des droits et de l'autonomie des personnes handicapées). Parmi eux, 84 % (n=97) faisaient partie de la population active, contre 93 % (n=230) chez les patients sans invalidité reconnue (p=0,006). Le taux de chômage était de 11 % (11/97) dans la population invalide, contre 9 % (20/230) chez les patients sans invalidité, mais cette différence n'était pas significative (p=0,456).

Parmi les patients salariés, respectivement 37 % (n=100, 3 données manquantes), 72 % (n=192, 1 donnée manquante) et 83 % (n=219, 7 données manquantes) avaient informé leur employeur, leurs collègues de travail et leur médecin du travail de leur maladie.

3.7. Facteurs associés à un niveau d'études plus élevé

L'activité de la maladie et la qualité de vie au moment de l'étude étaient les seuls facteurs significativement associés au niveau d'études, une maladie active et une qualité de vie faible étant négativement associées à l'obtention d'un diplôme d'enseignement supérieur en analyse univariée (OR : 0,47 [0,30 ; 0,75],p=0,001 pour une maladie active par rapport à la

rémission; OR : 0,44 [0,26 ; 0,77], p=0,003 pour une faible qualité de vie par rapport à une qualité de vie normale ; OR : 1,20 [0,50 ; 2,05], p=0,499 pour une qualité de vie élevée par rapport à une qualité de vie normale). L'âge précoce au moment du diagnostic (<10 ans), le sexe, le type de MICI, la localisation de la maladie au moment du diagnostic, les interventions chirurgicales antérieures, et les traitements antérieurs par anti-TNF, immunosuppresseurs ou corticoïdes n'étaient pas associés à un niveau d'éducation plus élevé.

3.8. Perception des patients quant à l'impact de la maladie sur leurs études et leur choix professionnel

Trente-deux pour cent (n=117) des patients estimaient que la MICI avait influencé le choix de leurs études, 61 % (n=219) pensaient que leur maladie avait eu un impact sur le déroulement de leurs études. La MICI avait influencé le choix de leur profession pour 36 % d'entre eux (n=130).

Il existait une concordance entre le dernier diplôme obtenu et la perception des patients quant à l'influence de la maladie sur le choix de leurs études (Figure 25, p<0,001) et sur le déroulement de leur cursus scolaire (Figure 26, p<0,008). En effet, les patients affirmant que la MICI avait influencé le choix de leurs études ou le déroulement de leurs études étaient plus fortement représentés dans les niveaux d'éducation inférieurs.

Figure 25. Concordance entre le dernier diplôme des patients atteints de MICI débutant dans l'enfance et la réponse concernant l'impact des MICI sur le choix des études.

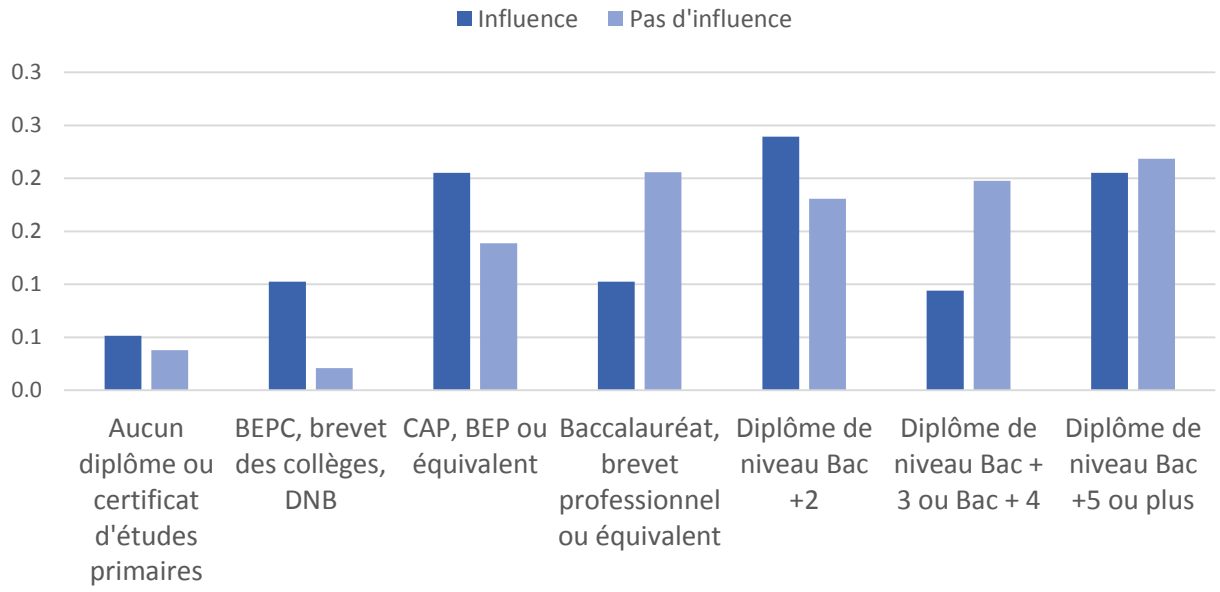
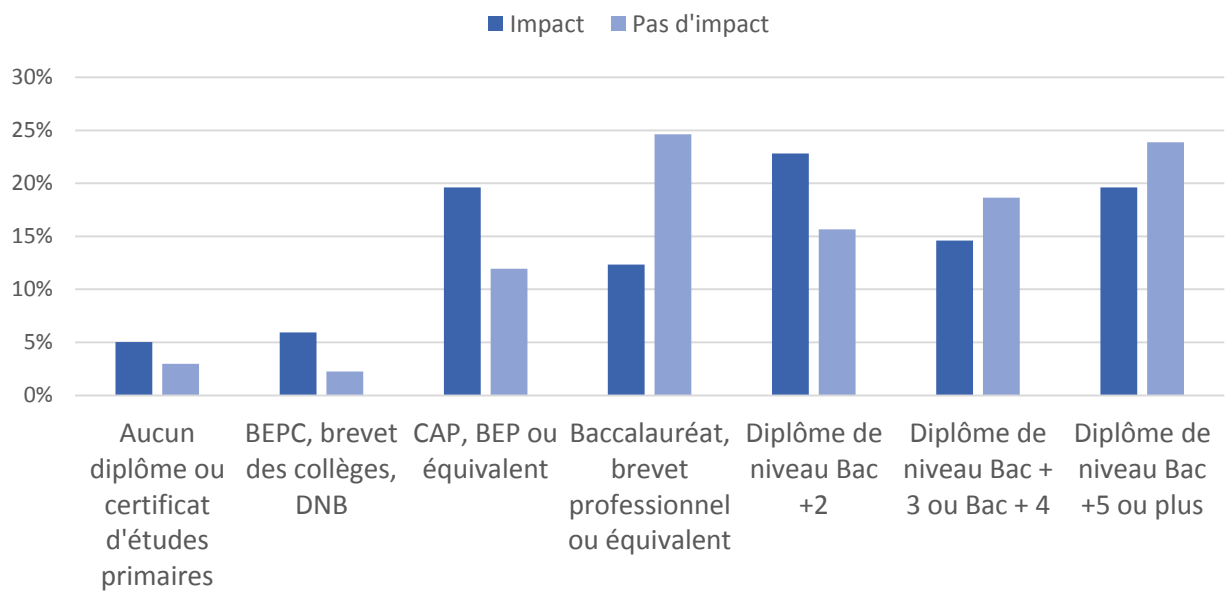


Figure 26. Concordance entre le dernier diplôme des patients atteints de MICI débutant dans l'enfance et la réponse concernant l'impact des MICI sur la progression des études.



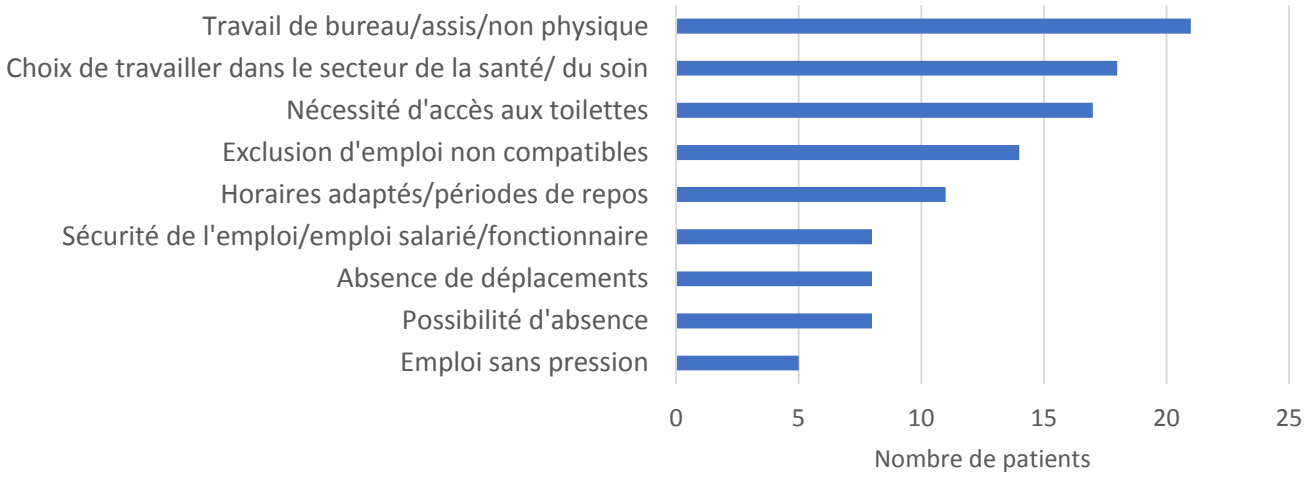
Nous avons ensuite analysé les commentaires libres des patients. Il est à noter que tous les patients n'ont pas fourni de commentaires libres sur ces questions d'éducation et de choix de la profession : 238 patients (66 % du total des patients) ont fourni un commentaire sur leurs études et/ou leur choix de profession.

Parmi les critères de choix de leur profession, les patients ont mentionné, par ordre d'importance : travail de bureau/assis/emploi non-physique (n=21), nécessité d'un accès facile aux toilettes (n=17), exclusion d'emplois non compatibles (n=14), horaires de travail adaptés/périodes de repos (n=11), sécurité de l'emploi/emploi salarié/fonctionnaire (n=8), absence de déplacement (n=8), possibilité d'absence (n=8), emploi sans pression (n=5) (Figure 27). Les professions incompatibles mentionnées étaient principalement l'enseignement, l'armée, la police.

De manière intéressante, 18 patients ont mentionné un impact positif de la maladie sur leur choix d'études et de profession. Leur maladie et le temps passé à l'hôpital ou dans un environnement médical les ont incités à travailler dans le secteur de la santé ou des soins. Ce résultat a été confirmé à l'aide des données quantitatives issues du questionnaire sur les professions, avec au total 17 % (n=56) des personnes travaillant dans le secteur de la santé tel que défini dans la section Méthodes. Parmi les salariés, 14 % (n=43) travaillaient dans le secteur de la santé, contre 9 % dans la population générale (base « tous employés » de l'INSEE,

p=0,005). De plus, 34 % des salariés (n=85) travaillaient dans le secteur public, contre 22 % dans la population générale (p<0,0001).

Figure 27. Résultats des réponses en commentaires libres : principaux critères de choix de la profession.



Les difficultés rencontrées par les patients dans leurs études sont illustrées sur la Figure 28. La principale difficulté rencontrée par les patients était liée à l'absentéisme. Les conséquences de ces difficultés sont présentées sur la Figure 29, impliquant principalement un retard scolaire. L'âge au moment de l'obtention du dernier diplôme a également été collecté dans notre étude. Dans la Table 14, les âges attendus pour chaque niveau de diplôme sont fournis. Les âges des patients semblent correspondre aux âges attendus, peut-être légèrement plus élevés, mais sont difficiles à interpréter en l'absence de données de référence.

Figure 28. Résultats des réponses en texte libres : principales difficultés engendrées par la maladie sur le déroulement des études.

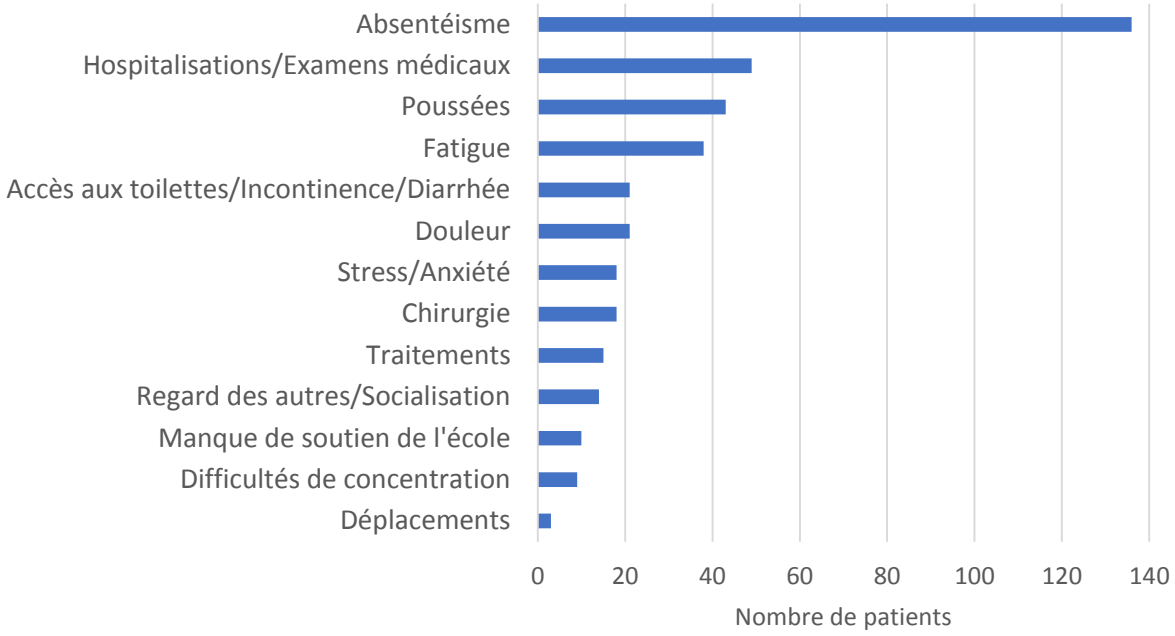
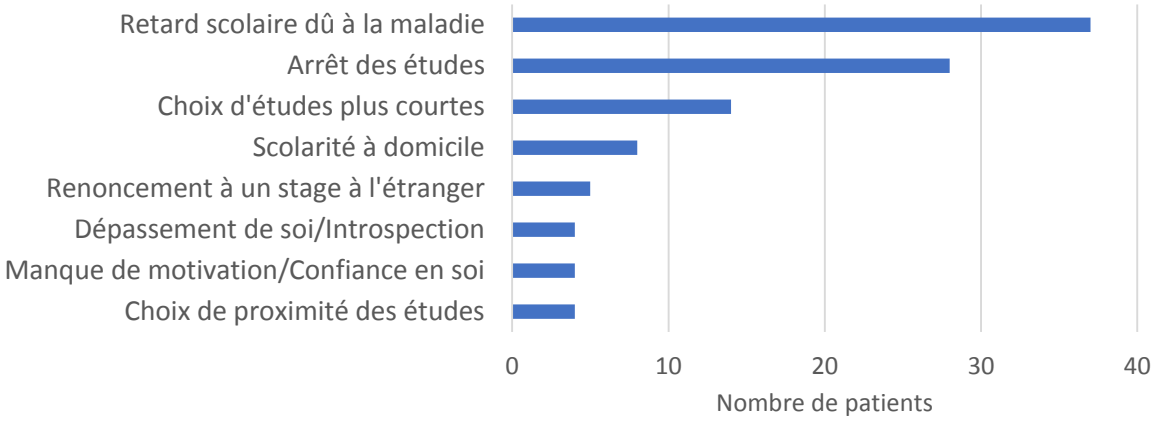


Figure 29. Résultats des réponses en texte libres : principales conséquences de la maladie sur le déroulement des études.



4. Discussion

RESUME DES PRINCIPAUX RESULTATS

Données quantitatives		
	MICI pédiatrique	Population générale de même âge et de même sexe (INSEE)
Taux de chômage	9 %	15 %
Cadres, professions intellectuelles supérieures, professions intermédiaires	63 %	44 %
Diplôme du supérieur	57 %	41 %
Emploi dans le secteur de la santé	14 %	9 %
Emploi dans le secteur public	34 %	22 %
Données qualitatives		
Difficultés dans le déroulement des études	Conséquences sur le déroulement des études	Choix de la profession
		

Cette étude a montré que, comparativement à la population générale du même âge et de la même localisation géographique, le taux de chômage était significativement plus bas parmi les participants à l'étude (9 % contre 15 % dans la population générale). De plus, les patients atteints de MICI débutant dans l'enfance avaient un niveau d'éducation plus élevé : 73 % des patients avaient un niveau d'éducation supérieure ou égal au baccalauréat, et 57 % avaient ensuite obtenu un diplôme de l'enseignement supérieur. En comparaison, dans la population générale, le taux de diplôme de l'enseignement supérieur était de 41 % ($p<0,0001$). De manière intéressante, nous avons également souligné que les patients atteints de MICI débutant dans l'enfance étaient plus enclins à travailler dans le secteur de la santé (14 % des patients atteints de MICI pédiatrique contre 9 % des emplois salariés dans la population générale). Ils sont également plus susceptibles de travailler dans le secteur public (34 % contre 22 % dans la population générale). À notre connaissance, il s'agit de la première étude décrivant ces deux derniers résultats.

COMPARAISON AUX DONNEES DE LA LITTERATURE

EDUCATION

De nombreuses études ont exploré l'impact de la maladie sur le statut professionnel des patients diagnostiqués à l'âge adulte, mais très peu d'études concernaient les patients diagnostiqués dans l'enfance. Ces études étaient souvent basées sur de petits échantillons de patients (192–196). Les critères d'évaluation différaient également d'une étude à l'autre : niveau d'éducation, absentéisme scolaire, performance scolaire ; travail à temps partiel, taux de chômage, revenus.

En ce qui concerne le niveau d'éducation, les études antérieures dans la littérature ont généralement montré un niveau d'éducation similaire à celui de la population générale (192,193,195–198). Nos résultats concordent avec deux études qui ont mis en évidence un niveau d'éducation plus élevé chez les patients atteints de MICI pédiatrique. Au Canada, El-Matary et al. ont comparé 112 patients adultes atteints de MICI pédiatrique (76 MC, 36 RCH et CCI) diagnostiqués de 1978 à 2009 à un groupe témoin de 565 adultes du même âge et du même sexe exempts de maladie chronique. Les patients atteints de MICI étaient significativement plus susceptibles d'obtenir un diplôme universitaire (88 %) que le groupe témoin (73 %) ($p<0,01$). Au Danemark, une étude basée sur des données administratives a comparé 3178 patients atteints de MICI (1344 MC, 1834 RCH) à 28 204 patients témoins appariés (199). Dans cette étude, les patients atteints de MICI pédiatrique avaient une probabilité plus élevée d'obtenir un diplôme de l'enseignement supérieur. Ainsi, le Hazard Ratio d'obtention d'un diplôme de l'enseignement supérieur était de 1,14 (intervalle de confiance à 95% : [1,07-1,21]) pour la MC et de 1,16 [1,10-1,23] pour la RCH.

Dans l'étude de Singh et al. (197), basée sur 337 patients atteints de MICI au Canada, les caractéristiques de la MICI, telles que l'âge au moment du diagnostic, la chirurgie, les hospitalisations, la nécessité d'une corticothérapie ou d'un autre traitement, n'avaient aucun impact sur le niveau d'éducation des patients. Le statut socio-économique ($p<0,0001$) et des problèmes de santé mentale au moment du diagnostic ($p=0,05$) étaient les deux facteurs prédictifs de résultats aux examens plus faibles. Dans la présente étude, seules l'activité de la maladie et la qualité de vie au moment de l'étude étaient associées à l'obtention d'un diplôme de l'enseignement supérieur. Cependant, l'association entre l'activité actuelle de la maladie ou la qualité de vie actuelle et l'activité passée de la maladie est discutable, car c'est l'activité passée qui aurait pu influencer les études.

Dans notre étude, un grand nombre de patients a signalé des difficultés rencontrées lors du déroulement de leurs études. Ceci est concordant avec des études mettant l'accent sur un développement social altéré chez les adolescents atteints de MICI. Dans l'étude de Hummel et al., ces patients étaient ainsi moins susceptibles que leurs pairs de participer à des sports en club pendant plus d'un an (p=0,02), d'aller en boîte de nuit pendant le lycée (p=0,002) et d'avoir une relation amoureuse avant l'âge de 18 ans (p=0,003) (195).

Dans notre étude, la principale difficulté rencontrée pendant les études était liée à l'absentéisme. Ce résultat est cohérent avec les résultats d'une étude comparant 50 patients adolescents atteints de MICI à 42 témoins dans l'Ohio. Dans cette étude, 20 % des patients atteints de MICI étaient absents de l'école pendant plus de 3 semaines par an, contre seulement 4 % des sujets non malades (p<0,05) (193).

EMPLOI

En ce qui concerne l'emploi, notre étude est, à notre connaissance, la première à mettre en évidence un taux de chômage plus faible chez les patients atteints de MICI à début pédiatrique en comparaison à la population générale. Les études précédentes dans la littérature ont généralement montré une insertion professionnelle similaire chez les patients atteints de MICI (194,196,198). A l'inverse, Hummel et al. (195) ont évalué, dans une étude menée aux Pays-Bas, l'activité professionnelle et le développement psychosocial de 62 patients atteints de MICI à début pédiatrique en comparaison à 76 sujets sains : les patients atteints de MICI ayant terminé leurs études universitaires (n=36) étaient significativement moins susceptibles d'avoir un emploi, avec un taux d'emploi de 25 % contre 57 % pour les sujets sains (p=0,004). Cependant, cette étude s'appuie sur un très petit échantillon de patients ayant terminé leurs études.

Les patients atteints de MICI à début pédiatrique avaient, dans l'ensemble, des revenus significativement plus bas au début de l'âge adulte par rapport à des individus de référence de la population générale appariés, dans une étude récente menée en Suède (198), alors que dans la même étude, les taux de chômage étaient comparables. Malheureusement, les données sur les revenus n'étaient pas disponibles dans notre étude. Une autre étude a montré que les jeunes adultes atteints de MICI avaient plus souvent des emplois à temps partiel que les patients témoins (200). Ce résultat n'a pas été observé dans notre étude, dans laquelle le taux de travail à temps plein était similaire à celui de la population générale (82 % contre 84 %).

FORCES ET FAIBLESSES

La principale force de notre étude réside dans le fait que les données proviennent de données en population générale issues d'un registre bien décrit. De plus, les patients étaient diagnostiqués dans l'enfance mais avaient tous plus de 25 ans au moment de l'étude, c'est-à-dire des patients ayant terminé leur scolarité, ce qui nous a permis d'étudier leur statut socio-professionnel. Les comparaisons avec la population générale étaient basées sur des données de référence solides de l'INSEE (recensement de la population et bases de données "tous salariés") de la même tranche d'âge, du même sexe et de la même zone géographique. Les données ont été post-stratifiées selon l'âge et le sexe à partir de la distribution par âge et sexe en population générale. Cette post-stratification n'a eu qu'un impact minime sur les résultats. Notre questionnaire reprenait par ailleurs exactement la même formulation que le questionnaire du recensement de la population sur les diplômes et les professions afin de rendre les données parfaitement comparables.

Une première faiblesse de notre étude réside, comme dans toutes les études réalisées par questionnaire, dans la possibilité de la présence d'un biais de non-réponse et d'un biais de rappel. Cependant, le taux de réponse était satisfaisant - avec 48 % de répondants parmi les patients ayant reçu le questionnaire - et nous avons comparé les caractéristiques au moment du diagnostic entre les répondants et les non-répondants. Nous n'avons observé aucune différence à l'exception d'une légère sur-représentation des patients atteints de MC, d'un taux plus élevé d'atteinte du tractus digestif supérieur chez les répondants atteints de MC et une maladie plus étendue chez les répondants atteints de RCH. Ces deux derniers points iraient plutôt dans le sens d'une maladie plus sévère chez les répondants et ne pourraient donc pas, a priori, expliquer les meilleurs indicateurs d'intégration professionnelle et de niveau d'études observés chez les patients par rapport à la population générale. À noter, la localisation de la maladie au moment du diagnostic n'était pas associée à l'obtention d'un diplôme de l'enseignement supérieur.

Une deuxième faiblesse réside dans l'absence de données sur le statut socio-économique des parents. Ce paramètre peut influencer le niveau d'éducation ainsi que l'avenir professionnel des enfants, indépendamment de l'influence de la maladie. Cependant, des études ont montré que les patients souffrant de MICI à début pédiatrique avaient un niveau socio-économique similaire à celui de la population générale, notamment dans l'étude de Baron et al. basée sur la cohorte pédiatrique du registre Epimad (97).

Conclusions de la première partie

Dans le premier chapitre, nous avons montré que les incidences de MC et de RCH continuaient d'augmenter de manière significative chez les enfants, mais également parmi les jeunes adultes dans le nord de la France. Ce résultat suggère que nous n'atteignons pas encore le plateau d'incidence du stade épidémiologique d'aggravation de la prévalence (compound prevalence) décrit dans la littérature. Dans la RCH une augmentation plus marquée des incidences a été observée chez les femmes, les taux d'incidence chez les femmes atteignant désormais ceux des hommes à la fin de la période d'étude.
Ces constatations suggèrent qu'un ou plusieurs facteurs environnementaux majeurs persistants peuvent prédisposer aux MICI les enfants, les jeunes adultes et les femmes de cette région. Des études épidémiologiques spécifiques et des études fondamentales seraient nécessaires pour évaluer cette hypothèse.

Sur la base de nos résultats, nous projetons que près de 0,6 % de la population du nord de la France sera atteinte de MICI d'ici 2030. Ce résultat souligne l'importance de se préparer au besoin croissant de soins et aux coûts associés, mais aussi au vieillissement de la population atteinte de MICI.

Dans le second chapitre, nous avons montré que les patients souffrant de MICI à début pédiatrique avaient un niveau d'éducation significativement plus élevé que celui de la population générale de la même zone géographique. Cette étude a également mis en évidence un taux de chômage plus bas que dans la population générale. De manière intéressante, les patients étaient également plus susceptibles de travailler dans le secteur public et dans le secteur de la santé. Les patients souffrant de MICI depuis l'enfance présentent une activité de la maladie pouvant interférer avec leurs parcours scolaire et universitaire, entraînant des difficultés de concentration, des retards scolaires et des répercussions sociales. Cependant, leur niveau d'éducation et leur activité professionnelle restent satisfaisants, démontrant une réelle capacité d'adaptation. Certains patients ont souligné un manque d'accompagnement et un manque d'assistance qui leur aurait permis de mener une « scolarité normale ». Cette démarche pourrait être facilitée grâce au soutien des

professionnels de la santé et à une éducation adéquate des patients concernant leur maladie, mais aussi par un meilleur accompagnement au sein des structures éducatives.
Il est important de considérer les enfants et les adolescents atteints de MICI dans leur globalité, tant sur le plan médical, pour minimiser le nombre de rechutes, que sur le plan social, afin de leur permettre de suivre un cursus scolaire aussi normal que possible.

Les résultats d'incidence sur la période 1988-2017 font l'objet d'un manuscrit soumis à la revue The Lancet Regional Health. J'ai réalisé les analyses statistiques et rédigé l'article.

Les résultats sur le niveau d'études et l'insertion professionnelle des patients à début pédiatrique font l'objet d'un manuscrit soumis à la revue United European Gastroenterology Journal (UEGJ). J'ai participé à la rédaction du protocole de l'étude, réalisé le suivi de l'enquête auprès des patients, réalisé l'ensemble des statistiques et rédigé l'article.

Les deux manuscrits soumis sont présentés dans les pages suivantes (versions soumises).

Title: Increasing incidence of inflammatory bowel diseases in children and young adults in Northern France: a 30-year population-based study.

Short Title: Increase in IBD in Northern France.

Authors

Hélène Sarter, MS[1,2], Thibaut Crétin, MD[3,4], Guillaume Savoye, MD, PhD[5], Mathurin Fumery, MD, PhD[6], Ariane Leroyer, MD, PhD[1,2], Luc Dauchet, MD, PhD[1,7,], Thierry Paupard MD[8], Hugues Coevoet[9], MD, Pauline Wils, MD[2,3], Nicolas Richard, MD[5], Dominique Turck, MD[2,10], Delphine Ley, MD, PhD[2,10], Corinne Gower-Rousseau, MD, PhD[11] for EPIMAD study Group*

Affiliations

[1] CHU Lille, Public Health, Epidemiology and Economic Health Unit, EPIMAD Registry, Maison Régionale de la Recherche Clinique, F-59000 Lille, France.

[2] Univ. Lille, Inserm, CHU Lille, U1286 - INFINITE - Institute for Translational Research in Inflammation, F-59000 Lille, France.

[3] Gastroenterology Unit, CHU Lille, University of Lille, F-59000 Lille, France.

[4] Gastroenterology Unit, Saint Philibert Hospital, Catholic University, Lille, France.

[5] Univ Rouen Normandie, INSERM, ADEN UMR1073, "Nutrition, Inflammation and microbiota-gut-brain axis", CHU Rouen, Department of Gastroenterology, F-76000 Rouen, France.

[6] Gastroenterology Unit, Amiens University Hospital, and Peritox, UMRI01, Université de Picardie Jules Verne, Amiens, France.

[7] Univ. Lille, INSERM, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Facteurs de risque et déterminants moléculaires des maladies liées au vieillissement, F-59000 Lille, France.

[8] Gastroenterology Unit, Dunkerque Hospital, France.

[9] Gastroenterology Unit, Les Bonnettes Private Hospital, Arras, France.

[10] CHU Lille, Division of Gastroenterology, Hepatology, and Nutrition, Department of Paediatrics, F-59000 Lille, France.

[11] Research and Public Health Unit, Robert Debré Hospital, Reims University Hospital, France.

Corresponding author

Hélène Sarter, MS,
Public Health, Epidemiology and Economic Health Unit, Registre EPIMAD,
Maison Régionale de la Recherche Clinique, Centre Hospitalier Universitaire Régional, CS 70001, F-59037 Lille Cedex, France.
Phone: +33-320-445-518, Fax: +33-320-446-945
E-mail: helene.sarter@chu-lille.fr.

Abbreviations

APC: Annual percent change

CD: Crohn's disease

CI: Confidence interval

IBD: Inflammatory bowel disease

IBDU: Inflammatory bowel disease unclassified

INSEE: National Institute of Statistics and Economic Studies (Institut National de la Statistique et des Etudes Economiques)

IQR: Interquartile range

IRR: Incidence rate ratio

UC: Ulcerative colitis

Author's contributions

Hélène Sarter, MS: Conceptualization, methodology, software, validation, formal analysis, data curation, visualization, writing- original draft.

Thibaut Crétin, MD: Conceptualization, writing- original draft.

Guillaume Savoye, MD, PhD: Conceptualization, investigation, validation, funding acquisition, writing-review and editing.

Mathurin Fumery, MD, PhD: Conceptualization, validation, investigation, funding acquisition, writing-review and editing.

Ariane Leroyer, MD, PhD: Methodology, data curation, formal analysis, software, writing-review and editing.

Luc Dauchet, MD, PhD: Methodology, writing-review and editing.

Thierry Paupard, MD: Investigation, resources, writing-review and editing.

Hughes Coevoet, MD: Investigation, resources, writing-review and editing.

Pauline Wils, MD: Investigation, resources, writing-review and editing.

Nicolas Richard, MD: Resources, writing-review and editing.

Dominique Turck, MD: Conceptualization, investigation, validation, writing-review and editing.

Delphine Ley, MD, PhD: Conceptualization, investigation, validation, writing-review and editing.

Corinne Gower-Rousseau, MD, PhD: Conceptualization, investigation, funding acquisition, validation, writing-original draft.

HS and CGR verified the underlying data.

Grant support

This work is based on the EPIMAD registry, which is funded by the "Institut National de la Santé et de la Recherche Médicale" (INSERM), Santé Publique France, and Amiens, Lille and Rouen University

Hospitals. The registry also receives logistic support from the DigestScience European charitable foundation (Lille, France) and the François Aupetit association.
All funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Conflict of interest statement

GS has served as speaker for MSD France, Ferring France, Abbvie France, and Vifor France.
MF has received lecture/consultant fees from Abbvie, Ferring, Tillots, MSD, Biogen, Amgen, Fresenius, Hospira, Pfizer, Celgene, Gilead, Boerhringer, Galapagos, Janssen and Takeda.
DL has received consultant fees from Sandoz and AbbVie.
TP has received lecture/consultant fees from Abbvie, Amgen, Takeda, Janssen, Biogen, and Celltrion.
DT has received lecture fees from Sandoz.
NR has received lecture/consultant fees from AbbVie and Takeda.
The other authors state that they have no competing interests regarding this work to disclose.

Writing assistance: None

Data transparency statement

Data sharing requests will be considered by the EPIMAD study group on written request to the corresponding author. De-identified participant data and data dictionary will be available, after approval of a written proposal and a signed data access agreement.

Acknowledgments

We thank the interviewing practitioners and the secretary who have collected data since 1988: S. Auzou, B. Bianco, L. Damageux, B. David, D. De Oliveira, P. Fosse, N. Guillon, V. Jacob, M. Leconte, C. Le Gallo, M. Lemahieu-Inglard, B. Lemaire, H. Pennel, A. Pétillon, D. Rime, I. Rousseau, B. Turck, N. Wauquier, and L. Yzet.

These results have been presented in part at the French Gastroenterology Congress (JFHOD, oral communication, Paris, June 25-28, 2020), the 15th Congress of the European Crohn's and Colitis Organization Congress (oral communication, Vienna, February 13-15, 2020) and the United European Gastroenterology Week (poster, Barcelona, October 19-23, 2019).

ID Approval: 917089

Word count: 2,982

ABSTRACT

Backgroung: In industrialized countries, the incidence of inflammatory bowel disease (IBD) appears stabilized. This study examined the incidence and phenotype of IBD in Northern France over a 30-year period.

Methods: Including all IBD patients recorded in the EPIMAD population-based registry from 1988 to 2017 in Northern France, we described the incidence and clinical presentation of IBD according to age, sex and evolution over time.

Findings: A total of 22,879 incident IBD cases were documented (59% of Crohn's disease (CD), 38% of ulcerative colitis (UC), 3% of IBD unclassified (IBDU)). Over the study period, incidence of IBD, CD and UC was 12·7, 7·2 and 5·1 per 10^5 person-years, respectively. The incidence of CD increased from $5{\cdot}1/10^5$ in 1988–1990 to $7{\cdot}9/10^5$ in 2015–2017 (annual percent change (APC): +1·9%, p<0·0001) while the incidence of UC increased from $4{\cdot}5/10^5$ to $6{\cdot}1/10^5$ (APC: +1·3%, p<0·0001). The largest increase was observed in children (+4·3% in CD, p<0·0001; +5·4% in UC, p<0·0001) followed by young adults aged 17 to 39 years (+1·9% in CD, p<0·0001; +1·5% in UC, p<0·0001). The increase in UC incidence was significantly higher in women than in men (+1·9% in women, +0·8% in men; p=0·006). We estimated that in our area, by 2030, nearly 0·6% of the population will have IBD.

Interpretation: The persistent increase of IBD incidence among children and young adults but also in women with UC in Northern France, suggests the persistence of substantial predisposing environmental factors.

Funding: Santé Publique France; INSERM; Amiens, Lille and Rouen University Hospitals.

Key words: inflammatory bowel disease, incidence, prevalence, population-based registry, Crohn's disease, ulcerative colitis.

Word count (abstract): 250

Research in context

Evidence before this study

We searched PubMed for research articles published from database inception until Dec 31st 2022, with no language restrictions, using the terms "Inflammatory Bowel Disease", "Incidence" and "epidemiology". In industrialized countries, the incidence of IBD appears to have stabilized or even decline in some countries. However, prospective population-based studies with expert-reviewed informations on the patients' diagnoses and disease characteristics are lacking to confirm this change.

Added value of this study

In a large population-based study performed over a 30-year period, we showed that the incidence of CD and UC is still rising in Northern France, particularly in children and young adults. In UC, incidence is rising more sharply in women and is nowadays reaching that in men. We project that by 2030, nearly 0.6% of the whole population in this French area will suffer from IBD (+30% in 10 years).

Implications of all the available evidence

These findings underscore i) the need for healthcare systems to prepare to face the increase of IBD patients, especially in the elderly, in the future ii) the need to a deeper understanding of environmental risk factors for IBD that particularly affect children and young people, but also women in UC.

INTRODUCTION

Inflammatory bowel disease (IBD) encompasses chronic disorders that cause inflammation of the gastro-intestinal tract, Crohn's disease (CD), ulcerative colitis (UC) and IBD unclassified (IBDU).[1,2] Over the past decades, IBD has emerged as a public health challenge worldwide.[3] It is estimated that about 1·3 million people suffer from IBD in Europe, corresponding to 0·2% of the population.[4] Although IBD was initially considered as a Western disease, its epidemiology is changing.[5,6] A recent systematic review of population-based studies (119 studies) revealed that overall incidence rates of IBD have stabilized in Western countries since 1990.[7] However, incidence rates are still rising in some western countries and in children.[8–12] Another concern is the increased prevalence of IBD, especially in the elderly.[13]

Since 1988, a large prospective, population-based registry (EPIMAD) on IBD has been built in Northern France, enabling to study the incidence of IBD and its changes over time.[11,14–16]

The objectives of the present study were to assess the incidence and clinical presentation of IBD in the general population over a 30-year period, to assess temporal trends in incidence according to age and sex and to estimate prevalence in 2030.

PATIENTS AND METHODS

The EPIMAD registry

All patients with IBD recorded in the EPIMAD registry in Northern France from 1988 to 2017 were included.

The EPIMAD registry covered, in 2017, 5,899,200 inhabitants corresponding to 9% of the total French population. The EPIMAD registry's methodology has been described in detail elsewhere.[14,15] Briefly, eight interviewer practitioners collect data from medical chart on all incident IBD patients diagnosed by all gastroenterologists (n=265) from private and public sector, using a standardized questionnaire. The gastroenterologist reports to the EPIMAD registry every patient consulting for the first time with symptoms suggestive of IBD. For each new incident case, an interviewer practitioner visits the gastroenterologist's office and collects the data.

Data collection

The main data recorded are age, sex, date of diagnosis, the time interval between symptoms onset and diagnosis, the clinical presentation, and the radiological, endoscopic and histological findings at the time of diagnosis. A diagnosis of CD, UC or IBDU is established by two expert gastroenterologists.[14] The anatomic sites and CD behavior are defined according to the Montreal Classification.[17]

Statistical analysis

Continuous variables were expressed as the median (interquartile range [IQR]), categorical variables as the frequency (percentage). Intergroup comparisons were performed using the Wilcoxon-Mann-Whitney test for continuous variables and the chi-squared or Fisher's exact tests for categorical variables. Changes over time in categorical variables were assessed using a chi-square test for trend, when appropriate.

Incidence rates were computed as the number of incident cases divided by the population at risk. The incidence rates were standardized by age with weightings for the European population[18] and determined for the population as a whole, for each age group (<17, 17-39, 40-59, and ≥60 years), by sex and for ten consecutive 3-year periods. 95% confidence interval (CI) were based on a gamma distribution.[19] Yearly population data by age-group and sex were obtained from census data from the INSEE (National Institute of Statistics and Economic Studies).

Differences in incidence according to age, sex or time were tested using log-linear Poisson regressions that took account of the number of person-years at risk (introduced as an offset variable) and over-dispersion, when necessary. Differences according to age and sex were presented as Incidence Rate Ratio (IRR) by exponentiation of the coefficients for age or sex. To assess linear time trends, the calendar year was introduced as a predictor variable. Time trends were presented as the annual percentage change (APC) estimated by exponentiation of the time coefficient from the Poisson regression estimates. To assess differences in time trends according to sex or age, interactions year*age and year*sex were included in the model.

To estimate the prevalence of IBD: i) we considered a linear increase of IBD cases in our area from 1975 to 1987 with the same distribution according to type of IBD, age and sex than in 1988-1990 ii) between 1988 and 2017, incident cases were those observed in EPIMAD registry ii) the annual incidence rates from 2018 to 2030 were projected from the APCs observed in 1988-2017 for each age group, sex, and type of IBD and were applied to population projection provided by the INSEE.[20] An aging process was then applied to each incident case from diagnosis year to 2030 using yearly survival rates by age, sex and birth cohort provided by the INSEE.[21]

The statistical analysis was performed using SAS software (v9.4, SAS Institute Inc., Cary, NC, USA) and R software v3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). The threshold for statistical significance was set to $P \leq 0.05$.

The study protocol has been approved by the French Ministry of Health according to the regulation of the registries in general population (Number n°97.107 for Advisory Committee on the Processing of Health Research Information (CCTIRS) and 917089 for the French Data Protection Authority (CNIL)).

RESULTS

Incidence of IBD

Between January 1st, 1988, and December 31st, 2017, we identified 22,879 incident cases of IBD, including 13,445 cases of CD (59%), 8,803 cases of UC (38%), and 631 cases of IBDU (3%). During the study period, the age-standardized incidence of IBD was 12·7/10^5 person-years (95% CI: [12·5-12·8]), and the incidence of CD, UC, and IBDU was respectively 7·2/10^5 [7·1-7·3], 5·1/10^5 [5·0-5·2], and 0·37/10^5 [0·34-0·40].

Clinical presentation at IBD diagnosis (Table 1)

The median [IQR] age at diagnosis was 26 [20-38] years for CD and 35 [25-48] years for UC. Ten per cent of patients (n=2,329) had a family history of IBD. Median time interval between symptoms onset and diagnosis was 3 months [1-9] in CD and 2 months [1-6] in UC (p<0·0001) without change over time. Extra-intestinal manifestations were present in 8% (n=1,899) of the patients.
In CD, 50% of the patients (n=6,467) had ileocolonic (L3) involvement, 30% (n=3,913) had pure colonic (L2) disease, and 20% (n=2,666) had pure ileal (L1) disease; 22% (n=2,951) displayed upper gastrointestinal involvement (L4). Five percent of the patients with CD (n=658) had perianal abscesses and/or fistulae at diagnosis. At diagnosis, 74% of CD patients presented an inflammatory behavior (n=3,840), 18% a stricturing behavior (n=928), and 8% a penetrating behavior (n=433).
At diagnosis, UC was classified as proctitis (E1) in 36% (n=2,741) of the patients, left-sided colitis (E2) in 35% (n=2,640), and extensive colitis (E3) in 29% (n=2,257).

Incidence and clinical presentation by sex

Results of sociodemographic and clinical data by sex are detailed in Table 1.
The incidence rate of CD was significantly higher in women (8·0/10^5 [7·8-8·2]) than in men (6·4 [6·3-6·6]) corresponding to an Incidence Rate Ratio (IRR) of 1·25 [1·19-1·31] (p<0·0001, Supplementary Table 1). Median age at diagnosis did not differ according to sex (p=0·128).
The incidence rate of UC was significantly lower in women (4·5 [4·4-4·7]) than in men (5·7 [5·5-5·9]), corresponding to an IRR of 0·83 [0·78-0·89] (p<0·0001); median age at UC diagnosis was significantly lower in women than in men (32 [24-44] vs. 38 [27-51] years respectively; p<0·0001).

Incidence and clinical presentation by age group

Sociodemographic and clinical data are summarized by age group in Supplementary Table 2. The 1988-2017 incidence rate of CD was significantly higher in the 17-39 years age-group (15·4/10^5 [15·1-15·8]) than in the <17 years (3·7 [3·5-3·9]), the 40-59 years (5·3 [5·1-5·6]) and the ≥60 years age-groups (2·4 [2·3-2·6]) (Supplementary Table 1, p<0·0001). Likewise, in UC, the incidence rate was significantly higher in the 17-39 years age-group (8·7 [8·4-8·9]) than in the <17 years (1·4 [1·3-1·5], the 40-59 years (5·6 [5·4-5·8]) and the ≥60 years age-groups (3·1 [2·9-3·3]; p<0·0001). As shown in Figure 1, there

was a significant interaction between sex and age in both CD (p<0·0001) and UC (p<0·0001). In CD, incidence was significantly higher in men than in women in the <17 years age-group (IRR: 0·8 [0·7-1·0], p=0·012), whereas significantly more women were observed in the 17-39 years age-group (IRR: 1·4 [1·3-1·5], p<0·0001). In UC, an inversion of the sex ratio was also observed with age; incidence was significantly higher in women in the <17 and 17-39 years age-groups (IRR: 1·4 [1·1-1·7], p=0·004 and 1·1 [1·1-1·2], p=0·036 respectively) and significantly higher in men in the 40-59 and ≥60 years age-groups (IRR: 0·6 [0·5-0·6], p<0·0001 and 0·5 [0·4-0·6], p<0·0001 respectively).

Temporal trends of incidences

The overall incidence of IBD increased from 10·4 [9·9-10·9] /10^5 person-years in 1988-1990 to 14·1/10^5 [13·6-14·7] in 2015-2017 (Annual Percent Change, APC: +1·5% [1·2-1·8]; p<0·0001). The incidence of CD increased from 5·1 [4·8-5·5] to 7·9 [7·4-8·3] (APC: +1·9% [1·6-2·2]; p<0·0001), and the incidence of UC also significantly increased from 4·5 [4·1-4·9] to 6·1 [5·7-6·5] (APC: +1·3% [0·9-1·7]; p<0·0001) (Figure 2).

- *Inflammatory bowel disease overall*

The increase in IBD incidence did not significantly differ according to sex (time*sex interaction p=0·229). On contrast, time trends significantly differed according to age (time*age interaction p<0·0001), with stability in the ≥60 years age-group (APC: -0·5% [-1·1-0·1], p=0·131) and the highest rise observed in the <17 years age-group (APC: +4·6% [3·9-5·2], p<0·0001) (Supplementary table 3).

- *Crohn's disease*

The time trends in CD did not differ according to sex (time*sex interaction: p=0·365) (Figure 3A). Among patients with CD, the sex ratio (F/M) was stable over time (p=0·085), fluctuating between 1·1 and 1·4. On contrary, time trends differed according to age (time*age interaction: p<0·0001) (Figure 3B). The highest increase was observed in the <17 years age-group (APC: +4·3% [3·5-5·1], p<0·0001), followed by the 17-39 years age-group (APC: +1·9% [1·5-2·2], p<0·0001), and the 50-59 years age-group (APC: +0·9% [0·3-1·5], p=0·006). In the ≥60 years age-group CD incidence remained stable over time (APC: +0·1% [-1·0-1·1], p=0·893). In each age group, time trends were not statistically different between men and women (time*sex interaction: p=0·387, 0·661, 0·071 and 0·101 in <17, 17-39, 40-59 and ≥60 years age groups, respectively)(Figure 4, Supplementary Figure 1 A, Supplementary Figure 1B and supplementary table 4).

- *Ulcerative colitis*

In UC, time trends differed significantly according to sex (interaction: p=0·006). The increase in UC incidence was significantly higher in women (APC: +1·9% [1·3-2·6]; p<0·0001 versus +0·8% [0·2-1·3]; p=0·006) with incidence in women being close to that of men at the end of study period (Figure 3C). The female/male ratio rose over time, from 0·7 in 1988-1990 to 0·9 in 2015-2017 (trend test: p<0·0001). Time trends in UC significantly differed according to age (p<0·0001, Figure 3D) but also according to age and sex. In <17 years age-group, time trends were not different according to sex

(time*sex interaction: p=0·591) with APC of +5·8% [4·0-7·5] in men (p<0·0001) and +5·2% [3·9-6·6] in women (p<0·0001). We observed significantly different time trends by sex in the 17-39 years (interaction: p<0·001) and 40-59 years age-groups (interaction: p=0·003). The incidence rates increased significantly in women aged 17-39 years and 40-59 years (APC: +2·1% [1·7-2·6], p<0·0001, and 1·7% [1·0-2·5], p=<0·0001, respectively). In men, incidence rates increased in the 17-39 years age-group by +0·9% [0·4-1·4] (p=0·001) and were stable in 40-59 years age-group (+0·1% [-0·6-0·7], p=0·883). In the ≥60 group age-group, time trends remained stable and did not differ significantly when comparing men and women (interaction: p=0·102) (Figure 4, Supplementary Figure 1C, Supplementary Figure 1D and supplementary table 5).

Prevalence

Estimated prevalence was 0·31% of total population in 2010 and 0·43% in 2020. By projecting the above mentioned results, the estimated prevalence would rise to 0·57% in 2030 (i.e. a 30% increase in prevalent cases in 10 years). Notably, the rise in prevalence is also accompanied by an aging of the IBD population (Figure 5), with the prevalence doubling in 20 years in the ≥60 years age-group while slightly rising in 17-39 years.

Temporal evolution of disease sites

The proportion of patients with L3 disease increased significantly between 1988-1990 and 2015-2017 (from 41% to 57%), whereas the proportion with a pure colonic phenotype (L2) decreased significantly from 38% to 22% (trend p<0·0001) (Supplementary Figure 2).

The proportion of perianal lesions remained stable over time (trend p=0·386) as well as the distribution of disease behavior (p=0·494).

In patients with UC, the proportion with E1 disease decreased significantly from 42% to 33%, whereas the proportion with E3 increased significantly from 20% to 30% (p<0·0001) (Supplementary Figure 2).

DISCUSSION

In a large population-based registry including 22,879 incident cases of pediatric- and adult-onset IBD over a 30-year period (1988-2017), the incidence of CD rose steadily from 5·1 to 7·9 (APC: +1·9%/year) and the incidence of UC rose from 4·5 to 6·1 (APC: +1·3%). The increase in incidence was particularly marked in children and young adults. In UC, incidence also increased more sharply in women than in men. Based on these results, we estimate that by 2030, about 0·6 % of the adult population in Northern France will suffer from IBD, with an aging of IBD population.

In the literature, the annual incidence rates in Europe range from 0·4 to 22·8 per 10^5 for CD, and from 2·4 to 44·0 per 10^5 for UC.[4] In the present study, the CD/UC ratio was above 1, contrasting with most of the other population-based studies performed in Western Europe.[4] Since the 1990s, epidemiological studies of Western countries have shown that incidence trends have changed: more than two third of studies in CD or UC reported stable or falling incidences.[6,7] This result contrasts with our present observation of an increase in the overall incidence of IBD in Northern France. Yet, some recent studies also highlighted a continuous rise in high incidence countries.[8,9,22]

Although the incidences appear to have stabilized in some Western countries, the overall estimated prevalence of IBD is still increasing because of a low mortality rate and ageing of the IBD patients.[6] Our estimated IBD prevalence of 0·6% in 2030 is in line with recent publications. Global prevalence of IBD in Europe is currently estimated at 0·2% with ranges from 1·5 to 331 /10^5 persons in CD and from 2·4 to 432 per 10^5 persons in UC.[4] Global Burden of Disease 2017 estimated the highest prevalence in USA and UK with a prevalence of about 0·45%.[23] The results of a Scottish study with a capture-recapture design suggested that the prevalence would rise to 1% within 10 years.[24] We report an aging of the prevalent population that was also described in the literature[24] and has to be taken into account considering the under-representation of elderly in clinical trials and the special features of the elderly population: drug interaction and side-effects raised by polypharmacy, comorbidities and frailty, and higher IBD-related hospitalization costs.[25]

Our study's major finding is that the trends in incidence of CD and UC are varying according to age and sex. The highest rates of increase were observed in children, followed by young adults in both CD and UC and both gender, whereas incidences were stable in 60 years old people and over. These results are in line with a recent systematic review of population-based studies that stated a significant increase in 84% of studies in children.[10] In middle-aged patients aged 40-59 years, incidence rates were only increasing in women and especially in UC. Shah et *al*. described sex differences in the incidence of IBD as a function of age at diagnosis.[26] These findings could suggest that female hormones probably play a role in IBD; we indeed observed a shift in the sex ratio at around the age of puberty and age of menopause, particularly in CD.

Current smoking has been shown to protect against UC while former smoker status was associated with an increased risk of UC - mostly in the first five years after smoking cessation.[27] The rise in regular

tobacco consumption among women since the early 1980s might contribute to the increased incidence of CD but not of UC.[28] Overall, women usually stop smoking earlier in life than men because of pregnancy; this may play a role in the increased incidence of UC in young women.

Appendectomy is reportedly associated with a 69% lower risk of UC.[29] The incidence of appendectomy has declined sharply since 1990, especially in women, with an inversion of the male/female sex ratio (0·85 in 1997, 1·05 in 2012).[30] This factor may have prompted an increase in the risk of developing UC in women.

Lastly, the global increase in incidence could also be related to the development of new diagnostic tools that may have led to an earlier and more widespread diagnosis of IBD. Nevertheless, in our study, the time interval between symptoms onset and IBD diagnosis did not change over time, suggesting that diagnostic modalities remained more or less stable over the study period.

Between 1988 and 2017, the proportion of CD with ileocolonic involvement (L3) rose significantly from 41% to 57%. This trend remained when considering only patients having undergone a full bowel assessment at diagnosis or when considering only patients having undergone CT and/or MRI (sufficient data only since 2006). These variations in disease site are thus likely to reflect a real rise in ileal involvement of the disease.

The present study's major strengths were its population-based design, the large sample size, the exhaustive recording over a 30-year period using the same methodology, and the use of validated, published diagnostic criteria. The major limitation is related to the absence of data on smoking status or oral contraceptive use for the understanding of their respective role in IBD evolution.

In conclusion, the results of this large population-based study over a 30-year period showed that incidences of CD and UC are still rising dramatically in children, but also among young adults in Northern France. Importantly, in UC, with incidences rising more sharply in women, incidence rates for women are reaching those for men at the end of study period. These findings strongly suggest that one or more persistent major environmental factors may predispose children, young adults and women to IBD in this area. Based on our findings, we project that 0·6% of the population in Northern France will experience IBD by 2030. This underscores the importance of preparing for the increasing healthcare demands and associated costs, as well as addressing the ageing of the IBD population.

Table 1: Sociodemographic and clinical data of IBD patients from a prospective population-based registry in Northern France from 1988 to 2017 (n=22,879).

Characteristics at diagnosis	All patients (n=22,879)	Males (n=10,925)	Females (n=11,954)	p-value
Disease Type				
CD	13,445 (58·8 %)	5,910 (54·0 %)	7,535 (63·0 %)	
UC	8,803 (38·5 %)	4,699 (43·0 %)	4,104 (34·3 %)	<0·0001
IBDU	631 (2·8 %)	316 (3·0 %)	315 (2·6 %)	
CD				
Median [IQR] age, years	26 [20-38]	26 [19-39]	26 [20-38]	0·128
Female gender	7,535 (56·0 %)			
Family history of IBD	1,700 (12·6 %)	720 (12·2 %)	980 (13·0 %)	0·154
Median [IQR] time between symptoms onset and diagnosis, months	3 [1; 9]	3 [1; 8]	3 [1; 9]	0·001
Disease site*				
L1	2,666 (20·4 %)	1,176 (20·6 %)	1,490 (20·3 %)	
L2	3,913 (30·0 %)	1,681 (29·4 %)	2,232 (30·4 %)	0·476
L3	6,467 (49·6 %)	2,853 (50·0 %)	3,614 (49·3 %)	
L4	2,951 (21·9 %)	1,418 (24·0 %)	1,533 (20·3 %)	<0·0001
Behaviour* £				
B1	3,840 (73·8 %)	1,443 (73·7 %)	1,769 (74·9 %)	
B2	928 (17·8 %)	357 (18·2 %)	420 (17·8 %)	0·550
B3	433 (8·3 %)	158 (8·1 %)	172 (7·3 %)	
Perianal disease	658 (4·9 %)	374 (6·3 %)	284 (3·8 %)	<0·0001
Extra-intestinal manifestations	1,565 (11·6 %)	647 (10·9 %)	918 (12·2 %)	0·027
UC				
Median [IQR] age, years	35 [25-48]	38 [27-51]	32 [24-44]	<0·0001
Female gender	4,104 (46·6 %)			
IBD family history	594 (6·7 %)	264 (5·6 %)	330 (8·0 %)	<0·0001
Median [IQR] time between symptoms onset and diagnosis, months	2 [1; 6]	2 [1; 6]	2 [1; 6]	0·013
Disease site*				
E1	3,157 (36·3 %)	1,536 (33·1 %)	1,621 (40·0 %)	
E2	3,210 (36·9 %)	1,802 (38·8 %)	1,408 (34·7 %)	<0·0001
E3	2,329 (26·8 %)	1,304 (28·1 %)	1,025 (25·3 %)	
Extra-intestinal manifestations	299 (3·4 %)	138 (2·9 %)	161 (3·9 %)	0·011

** According to the Montreal classification*
£ Recorded in the EPIMAD registry since 2008

FIGURES

Figure 1: Incidences rates of CD (n=13,445, panel a) and UC (n=8,803, panel b) in Northern France over the study period (1988-2017), by sex and 5-year age groups; Women/Men incidence rate ratio (IRR) according to age group in CD (panel c) and UC (panel d).

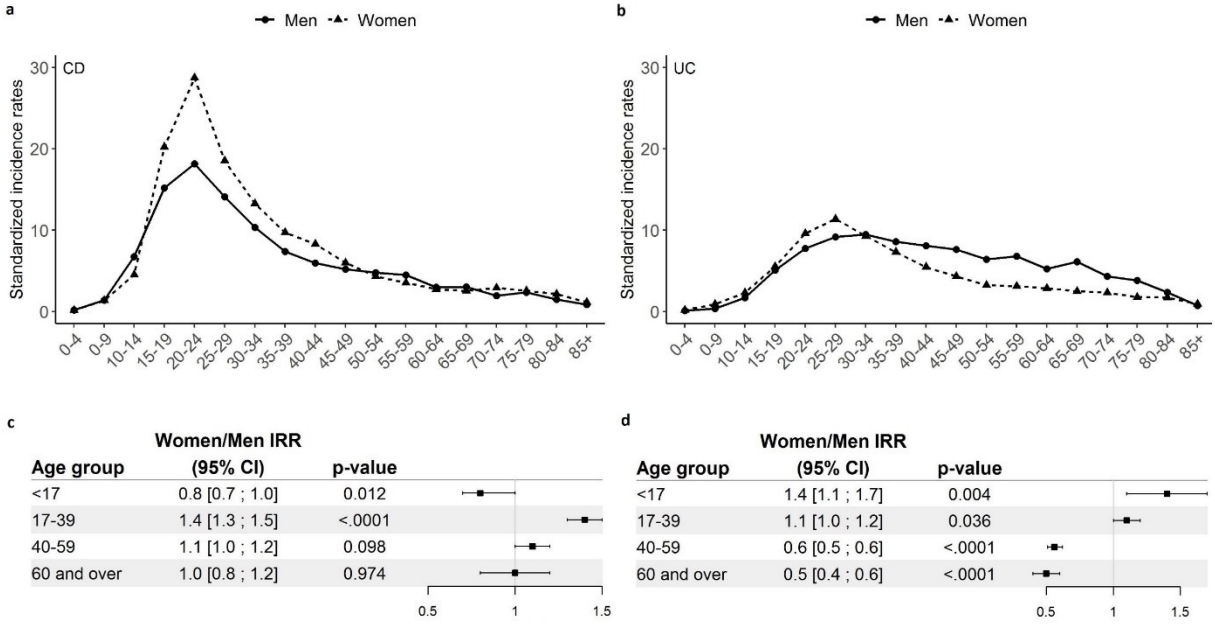


Figure 2: Changes over time in standardized incidence rates for IBD (n=22,879), CD (n=13,445) and UC (n=8,803) in Northern France from 1988 to 2017.

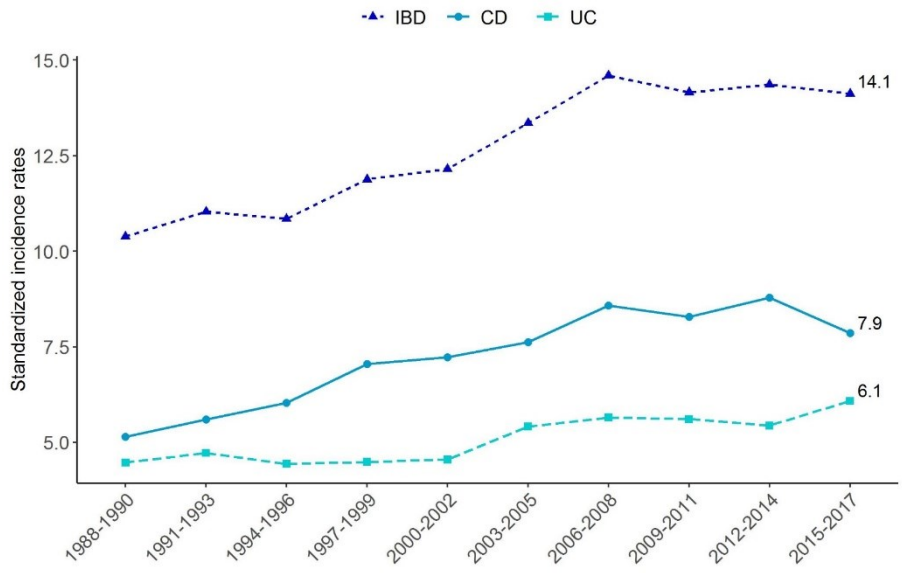


Figure 3: Changes over time in the standardized incidence rates for CD (n=13,445) and UC n=8,803) in Northern France from 1988 to 2017 by sex and age group. a) CD incidence according to sex. b) CD incidence according to age. c) UC incidence according to sex. d) UC incidence according to age.

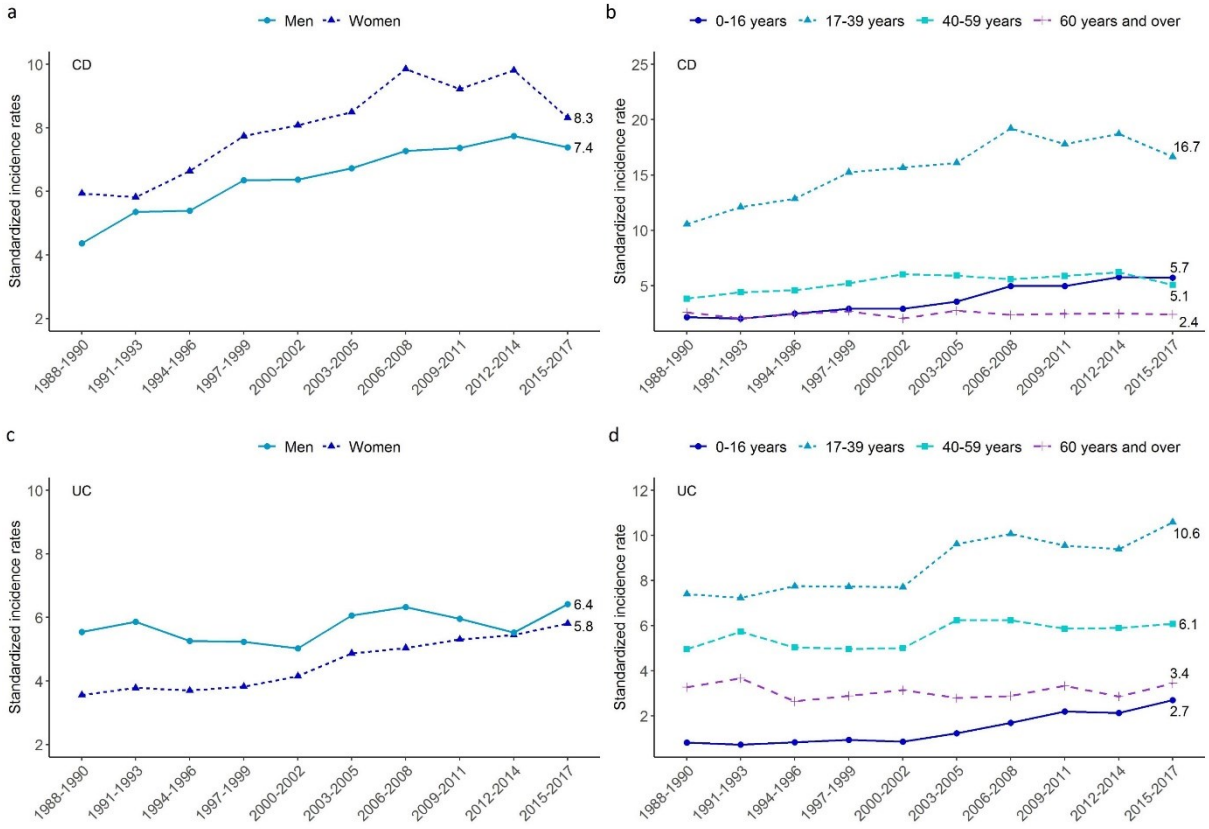


Figure 4: Annual percent change (APC) in %/year over the 1988-2017 period in patients with CD (n=13,445) and patients with UC (n=8,803) in Northern France by age group and sex.

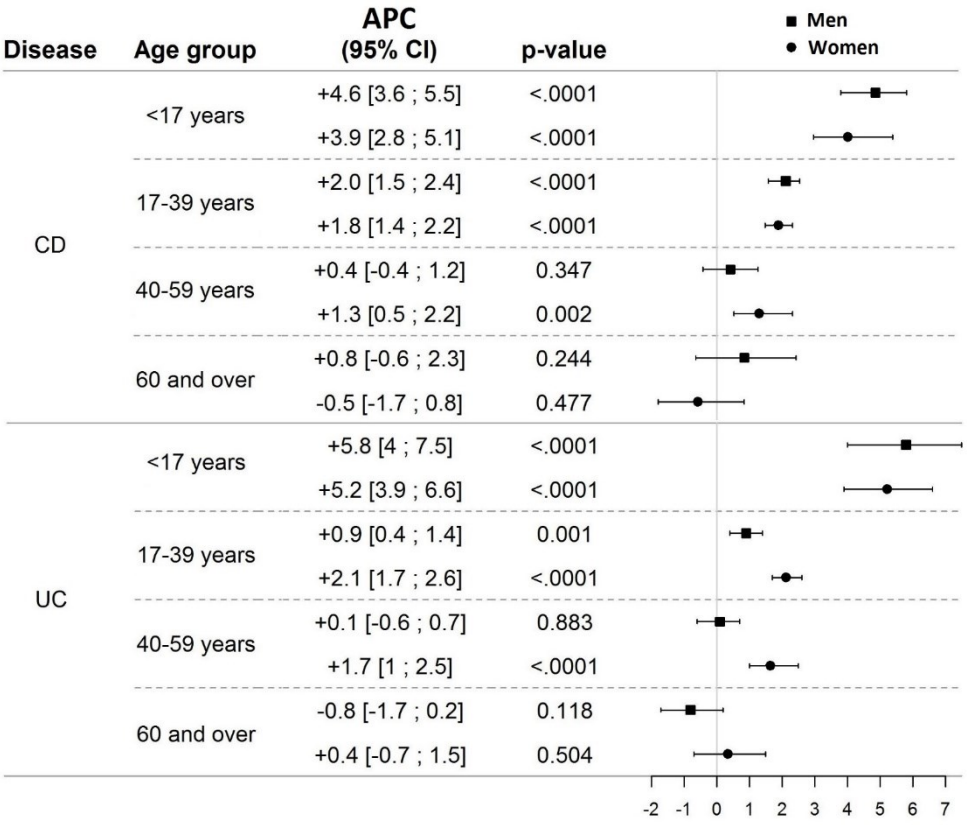
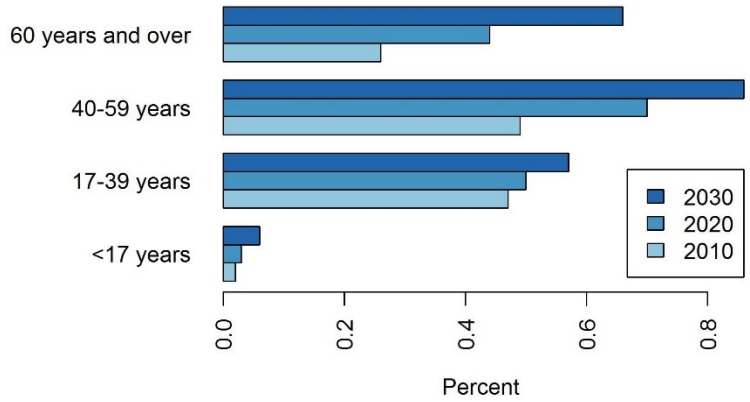


Figure 5: Prevalence according to age in 2010, 2020 and 2030.



REFERENCES

1 Torres J, Mehandru S, Colombel J-F, Peyrin-Biroulet L. Crohn's disease. *Lancet* 2017; **389**: 1741–55.

2 Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel J-F. Ulcerative colitis. *Lancet* 2017; **389**: 1756–70.

3 Kaplan GG. The global burden of IBD: from 2015 to 2025. *Nat Rev Gastroenterol Hepatol* 2015; **12**: 720–7.

4 Zhao M, Gönczi L, Lakatos PL, Burisch J. The burden of inflammatory bowel disease in Europe in 2020. *J Crohns Colitis* 2021; **15**: 1573–87.

5 Ananthakrishnan AN, Kaplan GG, Ng SC. Changing global epidemiology of inflammatory bowel diseases: sustaining health care delivery into the 21st century. *Clin Gastroenterol Hepatol* 2020; **18**: 1252–60.

6 Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* 2021; **18**: 56–66.

7 Ng SC, Shi HY, Hamidi N, *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2017; **390**: 2769–78.

8 Nielsen KR, Midjord J, Nymand Lophaven S, Langholz E, Hammer T, Burisch J. The Incidence and Prevalence of Inflammatory Bowel Disease Continues to Increase in the Faroe Islands - A Cohort Study from 1960 to 2020. *J Crohns Colitis* 2024; **18**: 308–19.

9 Dorn-Rasmussen M, Lo B, Zhao M, *et al.* The incidence and prevalence of paediatric- and adult-onset inflammatory bowel disease in Denmark during a 37-year period: a nationwide cohort study (1980-2017). *J Crohns Colitis* 2023; **17**: 259–68.

10 Kuenzig ME, Fung SG, Marderfeld L, *et al.* Twenty-first century trends in the global epidemiology of pediatric-onset inflammatory bowel disease: systematic review. *Gastroenterology* 2022; **162**: 1147–1159.e4.

11 Ghione S, Sarter H, Fumery M, *et al.* Dramatic increase in incidence of ulcerative colitis and Crohn's disease (1988-2011): a population-based study of french adolescents. *Am J Gastroenterol* 2018; **113**: 265–72.

12 Agrawal M, Christensen HS, Bøgsted M, Colombel J-F, Jess T, Allin KH. The Rising Burden of Inflammatory Bowel Disease in Denmark Over Two Decades: A Nationwide Cohort Study. *Gastroenterology* 2022; **163**: 1547–1554.e5.

13 Coward S, Clement F, Benchimol EI, *et al.* Past and future burden of inflammatory bowel diseases based on modeling of population-based data. *Gastroenterology* 2019; **156**: 1345–1353.e4.

14 Gower-Rousseau C, Salomez JL, Dupas JL, *et al.* Incidence of inflammatory bowel disease in northern France (1988-1990). *Gut* 1994; **35**: 1433–8.

15 Gower-Rousseau C, Vasseur F, Fumery M, *et al.* Epidemiology of inflammatory bowel diseases: new insights from a French population-based registry (EPIMAD). *Dig Liver Dis* 2013; **45**: 89–94.

16 Chouraki V, Savoye G, Dauchet L, *et al.* The changing pattern of Crohn's disease incidence in northern France: a continuing increase in the 10- to 19-year-old age bracket (1988-2007). *Aliment Pharmacol Ther* 2011; **33**: 1133–42.

17 Silverberg MS, Satsangi J, Ahmad T, *et al.* Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol* 2005; **19 Suppl A**: 5A–36A.

18 Revision of the European Standard Population - Report of Eurostat's task force - 2013 edition. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-13-028 (accessed Jan 31, 2024).

19 Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med* 1997; **16**: 791–801.

20 Projections de population 2013-2050 pour les départements et les régions | Insee [Population projections 2013-2050]. https://www.insee.fr/fr/statistiques/2859843 (accessed Jan 31, 2024).

21 Tables de mortalité par sexe, âge et niveau de vie [Mortality tables]| Insee. https://www.insee.fr/fr/statistiques/3311422 (accessed Jan 31, 2024).

22 Kontola K, Oksanen P, Huhtala H, Jussila A. Increasing Incidence of Inflammatory Bowel Disease, with Greatest Change Among the Elderly: A Nationwide Study in Finland, 2000-2020. *J Crohns Colitis* 2023; **17**: 706–11.

23 GBD 2017 Inflammatory Bowel Disease Collaborators. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* 2020; **5**: 17–30.

24 Jones G-R, Lyons M, Plevris N, *et al.* IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. *Gut* 2019; **68**: 1953–60.

25 Ananthakrishnan AN, Donaldson T, Lasch K, Yajnik V. Management of Inflammatory Bowel Disease in the Elderly Patient: Challenges and Opportunities. *Inflamm Bowel Dis* 2017; **23**: 882–93.

26 Shah SC, Khalili H, Gower-Rousseau C, *et al.* Sex-based differences in incidence of inflammatory bowel diseases-pooled analysis ofpopulation-based studies from western countries. *Gastroenterology* 2018; **155**: 1079–1089.e3.

27 Cosnes J. Smoking, physical activity, nutrition and lifestyle: environmental factors and their impact on IBD. *Dig Dis* 2010; **28**: 411–7.

28 OFDT. Tabac : évolution de l'usage occasionnel ou régulier parmi les 18-75 ans [Evolution of tobacco use in 18-75 years]. 2019. https://www.ofdt.fr/pdf/561 (accessed Jan 31, 2024).

29 Andersson RE, Olaison G, Tysk C, Ekbom A. Appendectomy and protection against ulcerative colitis. *N Engl J Med* 2001; **344**: 808–14.

30 Drees, Etudes et Résultats No 868. La longue diminution des appendicectomies en France | Direction de la recherche, des études, de l'évaluation et des statistiques [Diminution of appendicectomies on France]. 2014. https://drees.solidarites-sante.gouv.fr/publications/etudes-et-resultats/la-longue-diminution-des-appendicectomies-en-france-0 (accessed Jan 31, 2024).

Supplementary material

Supplementary Table 1: Incidence rates /10^5 person-years over the 1988-2017 study period and incidence rate ratio (IRR) according to gender and age group in CD and UC.

		1988-2017 Incidence rate /10^5 person-years [95% CI]	Incidence Rate Ratio (IRR) [95% CI]	p-value
CD	**Male**	6·4 [6·3 ; 6·6]	Ref	
	Female	8·0 [7·8 ; 8·2]	1·25 [1·19 ; 1·31]	<0·0001
	<17 years	3·7 [3·5 ; 3·9]	Ref	
	17-39 years	15·4 [15·1 ; 15·8]	4·28 [3·94 ; 4·65]	<0·0001
	40-59 years	5·3 [5·1 ; 5·6]	1·43 [1·29 ; 1·58]	<0·0001
	60 and over	2·4 [2·3 ; 2·6]	0·63 [0·56 ; 0·72]	<0·0001
UC	**Male**	5·7 [5·5 ; 5·9]	Ref	
	Female	4·5 [4·4 ; 4·7]	0·83 [0·78 ; 0·89]	<0·0001
	<17 years	1·4 [1·3 ; 1·5]	Ref	
	17-39 years	8·7 [8·4 ; 8·9]	6·26 [5·41 ; 7·25]	<0·0001
	40-59 years	5·6 [5·4 ; 5·8]	4·07 [3·49 ; 4·74]	<0·0001
	60 and over	3·1 [2·9 ; 3·3]	2·28 [1·92 ; 2·71]	<0·0001

Supplementary Table 2: Sociodemographic and clinical data for patients with IBD from a population-based registry in northern France from 1988 to 2017, by age group (n=22,879).

Variables at diagnosis	<17 (n=2,103)	17-39 (n=13,894)	40-59 (n=4,905)	≥ 60 (n=1,977)	P-value
Type of IBD					
CD	1,510 (71·8%)	8,796 (63·3%)	2,318 (47·3%)	821 (41·5%)	<0·0001
UC	562 (26·7%)	4,766 (34·3%)	2,424 (49·4%)	1,051 (53·2%)	
IBDU	31 (1·5%)	332 (2·4%)	163 (3·3%)	105 (5·3%)	
CD					
Family history of IBD	276 (18·3%)	1,168 (13·3%)	212 (9·1%)	44 (5·4%)	<0·0001
Women	672 (44·5%)	5,150 (58·6%)	1,229 (53·0%)	484 (58·9%)	<0·0001
Median [IQR] time between symptoms onset and diagnosis, months	3 [2 ; 7]	3 [1; 9]	3 [1; 9]	3 [1; 8]	0·101
Disease site*					
L1	222 (15·5%)	1,785 (20·9%)	531 (23·6%)	125 (15·8%)	<0·0001
L2	378 (26·0%)	2,213 (25·9%)	858 (38·0%)	464 (58·5%)	
L3	849 (58·5%)	4,549 (53·2%)	865 (38·4%)	204 (25·7%)	
L4	482 (31·9%)	2,208 (23·1%)	343 (14·8%)	98 (11·9%)	<0·0001
Behavior* £					
B1	665 (82·8%)	2386 (73·1%)	594 (70·2%)	195 (67·5%)	<0·0001
B2	105 (13·1%)	584 (17·9%)	175 (20·7%)	64 (22·1%)	
B3	33 (4·1%)	293 (9·0%)	77 (9·1%)	30 (10·4%)	
Perianal disease	92 (6·1%)	398 (4·5%)	118 (5·1%)	50 (6·1%)	0·019
Extra-intestinal manifestations	340 (22·5%)	924 (10·5%)	246 (10·6%)	55 (6·7%)	<0·0001
UC					
Family history of IBD	80 (14·2%)	372 (7·8%)	109 (4·5%)	33 (3·1%)	<0·0001
Women	317 (56·4%)	2,472 (51·9%)	890 (36·7%)	425 (40·4%)	<0·0001
Median [IQR] time between symptoms onset and diagnosis, months	2 [1 ; 5]	2 [1; 6]	2 [1; 6]	2 [1; 5]	0·636
Disease site*					
E1	122 (22·1%)	1,895 (40·2%)	925 (38·7%)	215 (20·6%)	<0·0001
E2	171 (30·9%)	1,537 (32·7%)	934 (39·0%)	568 (54·6%)	
E3	260 (47·0%)	1,278 (27·1%)	533 (22·3%)	258 (24·8%)	
Extra-intestinal manifestations	34 (6·0%)	162 (3·4%)	71 (2·9%)	32 (3·0%)	0·003

* *According to the Montreal classification*
£ *Recorded in the EPIMAD registry since 2008*

Supplementary Table 3: Time trends in incidence rates of IBD from a prospective population-based registry in northern France from 1988 to 2017 by sex and by age group (n=22,879).

Sex	Age group	Standardized incidence/10^5		APC % / year	P-value
		1988-1990	2015-2017		
Both	<17 years	3·1 [2·6 ; 3·7]	8·6 [7·7 ; 9·5]	**+4·6 [3·9 ; 5·2]*** **	**<·0001**
	17-39 years	18·8 [17·7 ; 20·1]	27·5 [26·1 ; 28·9]	**+1·6 [1·4 ; 1·8]*** **	**<·0001**
	40-59 years	10·0 [8·9 ; 11·2]	11·4 [10·4 ; 12·4]	**+0·5 [0·1 ; 0·9]***	**0·012**
	≥60 years	6·5 [5·5 ; 7·6]	6·0 [5·3 ; 6·8]	-0·5 [-1·1 ; 0·1]	0·131
	All	10·4 [9·9 ; 10·9]	14·1 [13·6 ; 14·7]	**+1·5 [1·2; 1·8] *** **	<0·0001
Women	<17 years	3·2 [2·4 ; 4·1]	8·1 [6·8 ; 9·5]	**+4·3 [3·4 ; 5·2]*** **	**<·0001**
	17-39 years	20·7 [19·0 ; 22·5]	29·5 [27·5 ; 31·7]	**+1·8 [1·5 ; 2·1]*** **	**<·0001**
	40-59 years	7·8 [6·4 ; 9·3]	11·2 [9·9 ; 12·6]	**+1·4 [0·8 ; 2]*** **	**<·0001**
	≥60 years	5·6 [4·4 ; 7·0]	5 [4·1 ; 6·1]	-0·5 [-1·3 ; 0·4]	0·281
	All	10·1 [9·4 ; 10·9]	14·3 [13·6 ; 15·1]	**+1·6 [1·2 ; 2·1]*** **	**<·0001**
Men	<17 years	3·0 [2·3 ; 3·8]	9·1 [7·8 ; 10·5]	**+4·8 [3·9 ; 5·7]*** **	**<·0001**
	17-39 years	17·0 [15·5 ; 18·7]	25·5 [23·6 ; 27·5]	**+1·4 [1 ; 1·7]*** **	**<·0001**
	40-59 years	12·3 [10·6 ; 14·2]	11·5 [10·2 ; 13]	-0·1 [-0·7 ; 0·4]	0·609
	≥60 years	8·0 [6·3 ; 10·2]	7·1 [5·9 ; 8·4]	-0·6 [-1·4 ; 0·3]	0·181
	All	10·9 [10·0 ; 11·8]	14·0 [13·2 ; 14·8]	**+1·3 [0·8 ; 1·7]*** **	**<·0001**

* significant time trend, 0·001≤p<0·05
** significant time trend, 0·0001≤p<0·001
*** significant time trend, p<0·0001
APC: annual percent change estimated using a log-linear Poisson model

Supplementary Table 4: Time trends in incidence rates of CD from a prospective population-based registry in northern France from 1988 to 2017 by sex and by age group (n=13,445).

Sex	Age group	Standardized incidence/10^5		APC % / year	P-value
		1988-1990	2015-2017		
	<17 years	2·2 [1·7 ; 2·7]	5·7 [5·0 ; 6·5]	**+4·3 [3·5 ; 5·1]*****	**<·0001**
	17-39 years	10·6 [9·7 ; 11·5]	16·7 [15·6 ; 17·8]	**+1·9 [1·5 ; 2·2]*****	**<·0001**
	40-59 years	3·8 [3·2 ; 4·6]	5·1 [4·5 ; 5·8]	**+0·9 [0·3 ; 1·5]****	**0·006**
	≥60 years	2·6 [2·0 ; 3·4]	2·4 [2·0 ; 3·0]	+0·1 [-1·0 ; 1·1]	0·893
Both	All	5·1 [4·8 ; 5·5]	7·9 [7·4 ; 8·3]	**+1·9 [1·6 ; 2·2]*****	**<·0001**
	<17 years	2·1 [1·5 ; 2·9]	4·9 [4·0 ; 6·0]	**+3·9 [2·8 ; 5·1]*****	**<·0001**
	17-39 years	12·9 [11·5 ; 14·3]	18·2 [16·6 ; 20·0]	**+1·8 [1·4 ; 2·2]*****	**<·0001**
	40-59 years	4·1 [3·1 ; 5·2]	5·8 [4·9 ; 6·9]	**+1·3 [0·5 ; 2·2]***	**0·002**
Women	≥60 years	2·8 [2·0 ; 3·9]	2·3 [1·7 ; 3·0]	-0·5 [-1·7 ; 0·8]	0·477
	All	5·9 [5·4 ; 6·5]	8·3 [7·7 ; 8·9]	**+1·7 [1·3 ; 2·1]*****	**<·0001**
	<17 years	2·2 [1·6 ; 3·0]	6·5 [5·5 ; 7·8]	**+4·6 [3·6 ; 5·5]*****	**<·0001**
	17-39 years	8·3 [7·2 ; 9·4]	15·1 [13·6 ; 16·7]	**+2·0 [1·5 ; 2·4]*****	**<·0001**
	40-59 years	3·6 [2·7 ; 4·7]	4·4 [3·5 ; 5·3]	+0·4 [-0·4 ; 1·2]	0·347
Men	≥60 years	2·4 [1·4 ; 3·9]	2·5 [1·9 ; 3·4]	+0·8 [-0·6 ; 2·3]	0·244
	All	4·4 [3·9 ; 5·0]	7·4 [6·8 ; 8·0]	**+2·0 [1·6 ; 2·5]*****	**<·0001**

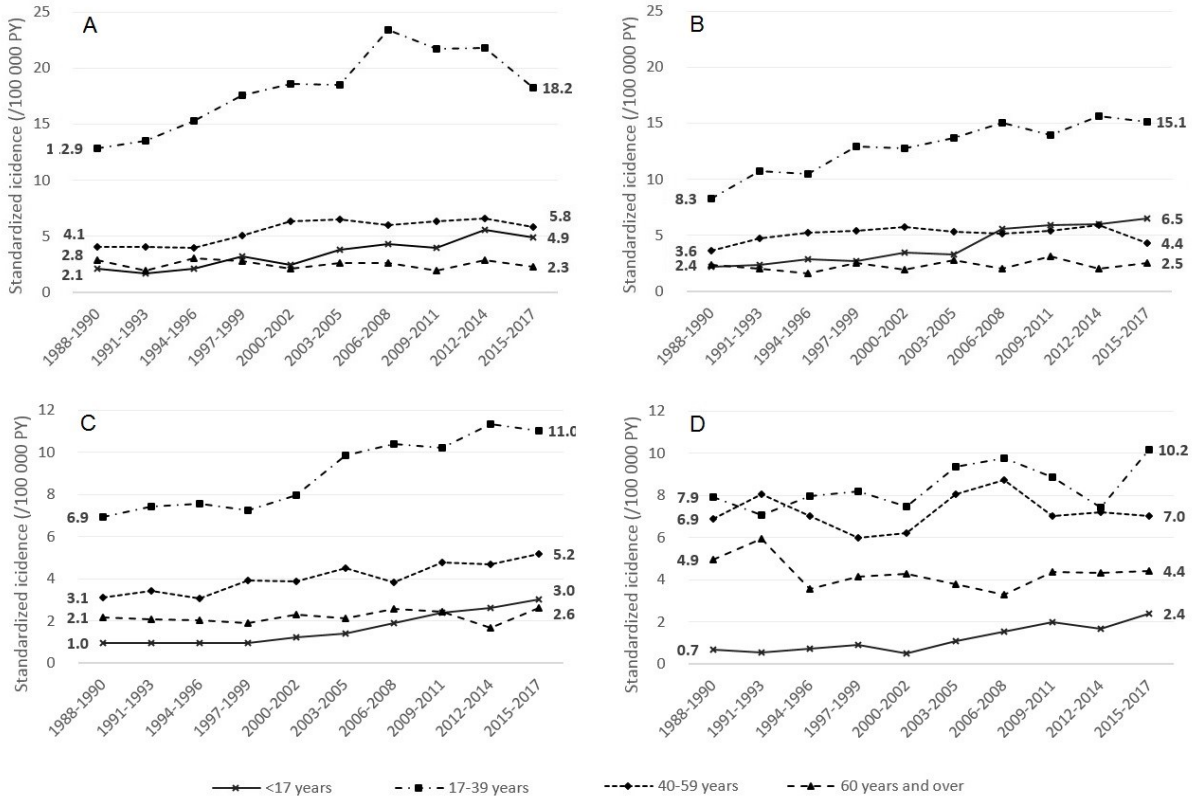
* significant time trend, 0·001≤p<0·05
** significant time trend, 0·0001≤p<0·001
*** significant time trend, p<0·0001
APC: annual percent change estimated using a log-linear Poisson model

Supplementary Table 5: Time trends in incidence rates of UC from a prospective population-based registry in northern France from 1988 to 2017 by sex and by age group (n=8,803).

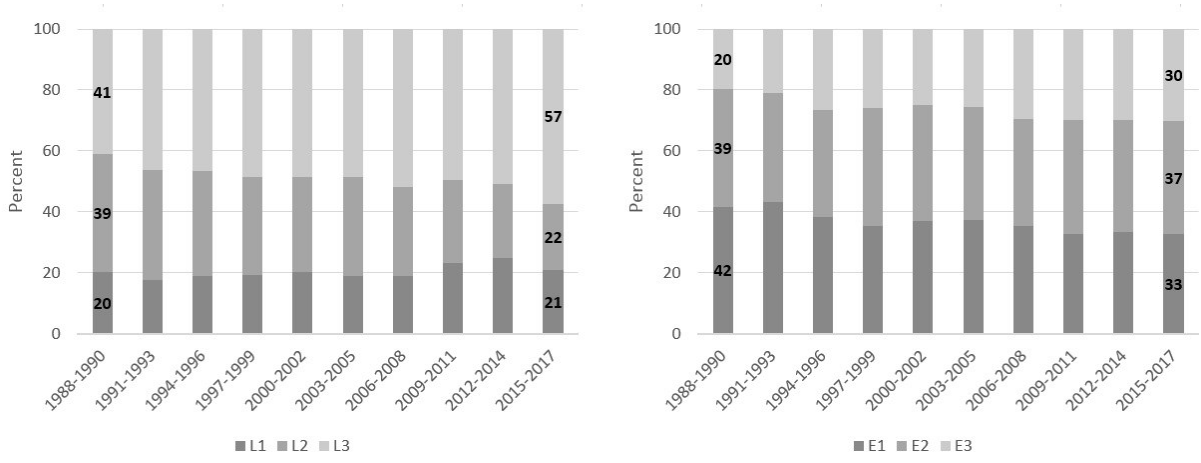
Sex	Age group	Standardized incidence/10^5		APC % / year	P-value
		1988-1990	2015-2017		
Both	<17 years	0·8 [0·6 ; 1·2]	2·7 [2·2 ; 3·3]	**+5·4 [4·3 ; 6·6]*** **	**<·0001**
	17-39 years	7·4 [6·7 ; 8·2]	10·6 [9·7 ; 11·5]	**+1·5 [1·2 ; 1·9]*** **	**<·0001**
	40-59 years	4·9 [4·2 ; 5·8]	6·1 [5·4 ; 6·8]	+0·7 [0·1 ; 1·2]*	0·014
	≥60 years	3·3 [2·6 ; 4·1]	3·4 [2·9 ; 4·1]	-0·2 [-1 ; 0·5]	0·53
	All	4·5 [4·1 ; 4·9]	6·1 [5·7 ; 6·5]	**+1·3 [0·9 ; 1·7]*** **	**<·0001**
Women	<17 years	1·0 [0·6 ; 1·5]	3·0 [2·3 ; 3·9]	**+5·2 [3·9 ; 6·6]*** **	**<·0001**
	17-39 years	6·9 [5·9 ; 8·0]	11·0 [9·8 ; 12·4]	**+2·1 [1·7 ; 2·6]*** **	**<·0001**
	40-59 years	3·1 [2·3 ; 4·1]	5·2 [4·3 ; 6·2]	**+1·7 [1 ; 2·5]*** **	**<·0001**
	≥60 years	2·1 [1·5 ; 3·1]	2·6 [2·0 ; 3·4]	+0·4 [-0·7 ; 1·5]	0·504
	All	3·6 [3·1 ; 4·0]	5·8 [5·3 ; 6·3]	**+1·9 [1·3 ; 2·6]*** **	**<·0001**
Men	<17 years	0·7 [0·4 ; 1·2]	2·4 [1·8 ; 3·2]	**+5·8 [4 ; 7·5]*** **	**<·0001**
	17-39 years	7·9 [6·8 ; 9·1]	10·2 [9 ; 11·5]	**+0·9 [0·4 ; 1·4]***	**0·001**
	40-59 years	6·9 [5·6 ; 8·3]	7·0 [6·0 ; 8·2]	+0·1 [-0·6 ; 0·7]	0·883
	≥60 years	4·9 [3·7 ; 6·7]	4·4 [3·5 ; 5·5]	-0·8 [-1·7 ; 0·2]	0·118
	All	5·5 [5·0 ; 6·2]	6·4 [5·9 ; 7·0]	**+0·8 [0·2 ; 1·3]***	**0·006**

* significant time trend, 0·001≤p<0·05
** significant time trend, 0·0001≤p<0·001
*** significant time trend, p<0·0001
APC: annual percent change estimated using a log-linear Poisson model

Supplementary Figure 1: Changes over time in the standardized incidence rates for CD (n=13,445) and UC n=8,803) in northern France from 1988 to 2017 by sex and by age group, as recorded in the EPIMAD registry. Each data point corresponds to the mean value for a 3-year period. A) CD incidence rates by age in women. B) CD incidence rates by age in men. C) UC incidence rates by age in women. D) UC incidence rates by age in men.



Supplementary Figure 2: Changes in disease sites in patients with CD (n=13,445) and UC (n=8,803) in Northern France from 1988 to 2017.



Title: Young adult patients with paediatric-onset inflammatory bowel disease have a higher educational level and a higher employment rate than the general population

Authors:

Hélène Sarter, MS[1,2], Mathilde Le Coniac MD[3], Ariane Leroyer, MD, PhD[1,2], Guillaume Savoye, MD, PhD[4], Mathurin Fumery, MD, PhD[5], Nathalie Guillon, MD[1,2], Corinne Gower-Rousseau, MD, PhD[6], Delphine Ley, MD, PhD[2,3], Dominique Turck, MD[2,3] for EPIMAD study Group*

Affiliations

[1] CHU Lille, Public Health, Epidemiology and Economic Health Unit, EPIMAD Registry, Maison Régionale de la Recherche Clinique, F-59000 Lille, France

[2] Univ. Lille, INSERM, CHU Lille, U1286 - INFINITE - Institute for Translational Research in Inflammation, F-59000 Lille, France

[3] CHU Lille, Division of Gastroenterology, Hepatology, and Nutrition, Department of Paediatrics, F-59000 Lille, France.

[4] Univ Rouen Normandie, INSERM, ADEN UMR1073, "Nutrition, Inflammation and microbiota-gut-brain axis", CHU Rouen, Department of Gastroenterology, F-76000 Rouen, France.

[5] Gastroenterology Unit, Amiens University Hospital, and Peritox, UMRI01, Université de Picardie Jules Verne, F-80000 Amiens, France.

[6] Research and Public Health Unit, Robert Debré Hospital, Reims University Hospital, F-51100 Reims, France.

Correspondence to: Hélène Sarter, Service de Santé Publique, Epidémiologie, Economie de Santé et Prévention, Registre Epimad, Maison Régionale de la Recherche Clinique, Centre Hospitalier Universitaire Régional, CS 70001, F-59037 Lille cedex, France. helene.sarter@chu-lille.fr

This work was presented in part at a French gastroenterology congress (*Journées Francophones d'Hépato-Gastroentérologie et Oncologie Digestive*, March 17th-20th, 2022, Paris, France).

DECLARATIONS

Ethics approval: The study protocol was approved by an independent ethics committee (CPP Nord-Ouest IV, Lille, France; reference: 2017 A003397 46).

Patient consent: Not applicable.

Availability of data and materials: Data are available on reasonable request to the corresponding author.

Conflicts of interest:

GS has served as a speaker for MSD France, Ferring France, Abbvie France, and Vifor France.

CGR has served as a speaker for Ferring France & International, Takeda France, Tillotts France, Janssen International, and MSD France.

MF has received lecture fees or consultancy fees from MSD, Abbvie, Takeda, Ferring, Gilead, Celgene, Celltrion, Biogen, Amgen,Fresenius, Galapagos, Tillots, Ferring, Janssen, Hospira, Pfizer, Arena, CTMA, and Boehringer.

DL has received lecture fees from Abbvie and Sandoz.

DT has received lecture fees from Sandoz.

HS, ML, and AL have nothing to declare.

Funding: The EPIMAD registry is organized under the terms of an agreement between the Institut National de la Santé et de la Recherche Médicale (INSERM) and Santé Publique France, and also receives financial support from Lille, Amiens, and Rouen University Hospitals and the DigestScience Foundation.

Authors' contributions:

Hélène Sarter, MS: Conceptualization, methodology, software, validation, formal analysis, data curation, visualization, writing- original draft.

Mathilde Le Coniac, MD: Conceptualization, Investigation, writing-review and editing.

Ariane Leroyer, MD, PhD: Methodology, data curation, formal analysis, software, writing-review and editing.

Guillaume Savoye, MD, PhD: Conceptualization, investigation, validation, funding acquisition, writing-review and editing.

Mathurin Fumery, MD, PhD: Conceptualization, validation, investigation, funding acquisition, writing-review and editing.

Nathalie Guillon, MD: Investigation, writing-review and editing.
Corinne Gower-Rousseau, MD, PhD: Conceptualization, investigation, funding acquisition, validation, writing-review and editing.
Delphine Ley, MD, PhD: Conceptualization, investigation, validation, writing-review and editing.
Dominique Turck, MD: Conceptualization, investigation, funding acquisition, validation, writing-original draft.

Acknowledgements:
The authors thank the interviewing practitioners and the clinical research associates who collected data: P. Bechu, N. Guillon, S. Auzou, B. David, H. Pennel, A. Pétillon, and L. Damageux. The authors also thank all the patients, gastroenterologists, and paediatric gastroenterologists who participated in the study.

Abbreviations:
CD, Crohn's disease
HBI, Harvey-Bradshaw Index
INSEE, *Institut National de la Statistique et des Etudes Economiques* (French National Institute of Statistics and Economic Studies)
IBD, inflammatory bowel disease
IQR, interquartile range
SCCAI, Simple Clinical Colitis Activity Index
SIBDQ, Short Inflammatory Bowel Disease Questionnaire
UC, ulcerative colitis

Word count: 3,791

ABSTRACT

Background & aims: There are few published data on the social impact of paediatric-onset inflammatory bowel diseases (IBD). The objective of the present study was to assess the educational level and occupational status of adult patients with paediatric-onset IBD from the EPIMAD Registry.

Methods: The inclusion criteria were (i) a diagnosis of paediatric-onset (<17 years at diagnosis) IBD, and (ii) age 25 or over at the time of the study. The patients answered a self-questionnaire on their educational level and profession. The data were compared with those for members of the general population of the same age and from the same geographic area.

Results: Three hundred and sixty-one patients (286 with CD and 75 with UC) filled out and returned the questionnaire. The median [interquartile] age was 15.0 [12.9; 6.3] at diagnosis and 34.2 [29.6; 39.5] at the time of the study. Patients were more likely to have a higher education degree than the general population (57% *versus* 41%, respectively; p<0.0001). The unemployment rate was significantly lower among study participants than among the general population (9% *versus* 15%, respectively; p=0.001). Salaried patients were significantly more likely to be employed in the healthcare sector (14%, *versus* 9% in the general population; p=0.005) and in the public sector (34%, *versus* 22% in the general population; p<0.0001).

Conclusion: Our results showed that relative to the general population, patients with paediatric-onset IBD had a higher educational level and a higher employment rate and were more likely to work in the healthcare and public sectors.

Keywords: Crohn's disease; profession; education; ulcerative colitis.

Word count (abstract): 250

INTRODUCTION

The incidence of paediatric inflammatory bowel disease (IBD) has increased dramatically over the last few decades.[1,2] Paediatric-onset IBD is usually considered to account for between 10% and 25% of incident IBD cases (depending on the upper age limit used in the definition of onset, which ranges from 15 to 20) and is associated with a high morbidity rate. Patients with paediatric IBD often present with a more aggressive clinical course, more extensive disease, and more frequent complications (such as nutritional impairment, delayed puberty, growth retardation, disability, and the need for surgery).[3–6] These patients also experience more IBD-related hospital admissions than their adult-onset counterparts.[7]

IBD follows a chronic relapsing-remitting course and thus significantly impacts the patients throughout their lives. Furthermore, adolescent patients are affected by IBD during a period that is critical in terms of their education and career plans. As with many chronic illnesses, IBD has significant impact on a child's or adolescent's physical, emotional and social development. Furthermore, it is well established that the symptoms of IBD worsen a patient's quality of life (QoL).[8,9]

The putative relationship between paediatric-onset IBD and social problems has not been extensively studied. One can hypothesize that the recurrent and sometimes long absences from school or work associated with disease exacerbations have severe negative impacts effects on a person's education or employment. The objective of the present study was to assess educational levels and employment rates in young adult patients with paediatric-onset IBD *versus* members of the general population of the same age range and sex.

PATIENTS AND METHODS

Study population

The main study inclusion criteria were (i) documentation in the Epimad registry (see below), (ii) Inflammatory Bowel Disease (IBD) diagnosed before the age of 17, and (iii) age 25 or over at the time of study (to ensure that all participants had completed their education). The Epimad registry was initiated in 1988 and has been described in detail elsewhere.[10] Briefly, all gastroenterologists and paediatric gastroenterologist practicing in the private or public sectors in northern France (n=265) have reported on all patients consulting for the first time for symptoms suggestive of IBD. On a regular basis (three times a year, on average), staff from the Epimad office visit the gastroenterologists' offices and collect information from the patients' medical records. The final diagnosis of IBD is made by two expert gastroenterologists, according to previously published and validated criteria.[10]

Data collection

Patients were contacted by mail and asked to fill out a self-questionnaire concerning their highest educational qualification or degree, their age on obtainment of the highest educational qualification or

degree, profession, current occupational status, socioprofessional category, work disability status, previous and ongoing medical treatments, previous surgery related to IBD, the presence of stoma, disease activity (according to the Harvey-Bradshaw Index (HBI) for patients with Crohn's disease (CD) and the Simple Clinical Colitis Activity Index (SCCAI) for patients with ulcerative colitis (UC), and QoL (according to the Short Inflammatory Bowel Disease Questionnaire (SIBDQ)).[11–13] Active disease was defined as an HBI >3 for CD or an SCCAI >2 for UC. QoL was defined as "poor" (with a SIBDQ score below 45), "normal" (with a SIBDQ score between 45 and 60) or "high" (with a SIBDQ score over 60).[14] Patients were also asked about the impact of IBD on their choice of educational courses, the progress of their educational courses, and their choice of profession.

Professions were classified according to the socio-professional categories defined by the French National Institute for Statistics and Economic Studies (*Institut National de la Statistique et des Etudes Economiques* (INSEE)).

Demographic and clinical characteristics at diagnosis were extracted from the EPIMAD registry's database: age, sex, the date of diagnosis, the time interval between symptom onset and diagnosis, clinical presentation at diagnosis, and any family history of IBD. The CD's location was defined according to the Montreal classification[15] as pure ileal involvement (L1), pure colonic involvement (L2), ileocolonic involvement (L3), and/or upper gastrointestinal disease (L4, which could be combined with L1, L2 or L3). Patients with ileocecal involvement were classified as L3. The presence or absence of perianal lesions was noted. The CD's behaviour was not assessed because this variable has only been recorded since 2008. The UC's location was defined according to the Montreal classification as proctitis (with disease limited to the rectum (E1)), left-sided colitis (with disease limited to the colon below the splenic flexure (E2)), or extensive colitis (with involvement of the colon beyond the splenic flexure (E3)).

Patients who had not sent back the completed self-questionnaire within two months were chased up by mail, in order to request their participation in the study.

Definition of outcomes

The definitions used in the present study were taken from the INSEE's 2020 census of the population in France.

Educational level was defined as the highest educational qualification or degree obtained, in five groups: (i) No formal education or primary education only; (ii) junior secondary education; (iii) short vocational secondary education; (iv) upper secondary education (a baccalaureate high school leaving certificate, or equivalent); and (v) higher education. The latter category was subdivided into 2 years of higher education, 3 or 4 years of higher education, and 5 or more years of higher education.

The occupational status was defined as the participant's primary situation at the time of the study: employed, apprentice or paid trainee, student or unpaid trainee, unemployed, retired, or homemaker. The occupied labour force was defined as all people exercising a profession or helping other people in their work (whether paid or not); employed people, apprentices, trainees, employed students, and

employed retirees (Figure 1). The labour force was defined as the occupied labour force plus the active unemployed population (i.e. unemployed people actively looking for work). The unemployment rate was defined as the number of unemployed people as a proportion (in percent) of the labour force.
Professional occupations were classified with regard to the six socioprofessional categories defined by the INSEE: (i) farmers; (ii) craftspeople, shopkeepers, and managers; (iii) executives and intellectual professions; (iv) intermediate professions; (vii) employees; and (viii) workers. Healthcare professionals were defined as physicians, dentists, pharmacists, nurses, and other healthcare providers.

Reference data

Reference data for the general population of the same age and sex and from the same geographic area were obtained from the INSEE's population census database (for reference data on the educational level, occupational status, socioprofessional category and unemployment rate) and "all employees" database (for employment in the public and in the healthcare sector). The reference year for these data was 2019.

Statistical analysis

Data management and statistical analysis were carried out using SAS software (version 9.4, SAS Institute Inc., Cary, NC, USA). The threshold for statistical significance was set to $p \leq 0.05$.
Quantitative variables were quoted as the mean (standard deviation) or the median [interquartile range (IQR)], depending on the distribution. Qualitative variables were quoted as the frequency (percentage).
In order to evaluate non-response bias, the respondents and non-respondents were compared with regard to their clinical and demographic characteristics at diagnosis. To assess differences in the socio-economic level, we examined the FDEP09 deprivation index [16] and the level of urbanization of the place of residence at diagnosis.
In order to facilitate comparisons, we post-stratified our sample by five-year age group and by sex so that it was proportionally consistent with the 2019 INSEE national census population. Below, the tables give the raw and adjusted rates and the text presents the adjusted rates only. The paediatric-onset IBD patient population and the reference general population were compared by applying chi-square tests of comparison to a theoretical proportion (goodness of fit test) taking into account post-stratification weights. Theoretical proportions were those observed in reference data.

The study protocol was approved by an independent ethics committee (*CPP Nord-Ouest IV*, Lille, France; reference: 2017 A003397 46).

RESULTS

The response rate

A total of 1,076 patients registered in the Epimad database met the inclusion criteria. Of these, 207 were lost to follow-up, 19 had died since the diagnosis of IBD, and 5 had a complicated personal situation (e.g. incarceration) and so were not contacted (Figure 2). Ultimately, 845 patients were contacted by mail. One hundred patients no longer lived at the mailing address given in the database, and 361 patients (WW with CD and YY with UC) sent back the completed questionnaire: this corresponded to 34% of the eligible patients, 43% of the patients sent a questionnaire (including those with an incorrect mailing address), and 48% of the patients who actually received the questionnaire. The patients filled out the self-questionnaire between November 2019 and June 2021.

Clinical and demographic characteristics of the patients

Forty-seven percent (n=170) of the respondents were men (Table 1). The median [IQR] age was 15.0 [12.9; 16.3] at diagnosis and 34.2 [29.6; 39.5] at the time of the study. The median disease duration was 20.7 years [15.7; 26.1]. The differences in these variables between respondents and non-respondents were not statistically significant. Seventy-nine percent (n=286) of respondents and 74% (n=527) of non-respondents had CD (p=0.047). Concerning the patients with CD, L4 involvement was significantly more common among respondents (32%, n=93) than among non-respondents (24%, n=127; p=0.010). Concerning the patients with UC, the disease was significantly more extensive in respondents than in non-respondents: E1, E2 and E3 locations were reported in respectively 12%, 38% and 49% of the respondents and 31%, 30% and 38% of the non-respondents (p=0.007).

Characteristics of the place of residence

In order to assess a potentially biased relationship between response status and socio-economic status, we compared respondents and non-respondents with regard to the deprivation index (in quintiles) and the level of urbanization of the place of residence (Table 1). The intergroup difference in the deprivation index was not significant (p=0.162). However, the intergroup difference in the urbanization level was significantly different (p=0.036); relative to non-respondents, respondents were more likely to live in a rural area than (22% *versus* 28%, respectively) and less likely to live in a densely populated area (38% *versus* 32%, respectively).

Disease activity and QoL

Forty-three percent of the patients (n=141) were in remission at the time of the study, 52% (n=168) had active disease, and 5% (n=16) had severe disease. The median QoL rating was 54 [45 ; 61]. Twenty-three percent (n=79) of the patients had a low QoL, 51% (n=172) had a normal QoL and 26% (n=88) had a high QoL (Table 1).

Educational level

For the 359 patients having finished their formal education, the median [IQR] age at highest diploma was 22 [19; 24] (Figure 3A and Supplementary Table 1). Compared with the general population, patients had a higher educational level; the proportions with a degree from a higher education establishment were respectively 41% and 57% (p<0.0001).

Professional status

The great majority of the patients (82%, n=294) were in work at the time of the study (Table 2). Five of the 294 were on sick leave or maternity/paternity leave at the time of the study.

The labour force comprised 91% (n=327) of the patients; this proportion was similar in the general population (90%, p=0.578) (Table 3). Among the labour force population, the unemployment rate was lower among the patients than in the general population (9% *versus* 15%, respectively; p=0.001). Ninety-one percent (n=271) of the occupied labour force were salaried employees, and 82% (n=238) had full-time work; these proportions were similar in the general population (92% (p=0.927) and 84% (p=0.172), respectively).

The distribution of socioprofessional categories in the patient group differed significantly (p<0.0001) from that in the general population (Figure 3B and Supplementary Table 2). Indeed, the proportion of executives and higher intellectual professions and the proportion of intermediate professions were higher in the patient group than in the general population (22% *versus* 16%, and 41% *versus* 28%, respectively).

A total of 115 (32%) patients were officially registered as disabled. The proportion in the labour force population was lower for patients with a disability (84%, n=97) than for patients without a disability (93%, n=230; p=0.006). The unemployment rate was higher among patients with a disability (11%, n=11 out of 97) than among patients without a disability (9%, n=20 out of 230), although this difference was not statistically significant (p=0.456).

Of the salaried patients, 37% (n=100, 3 missing data) had informed their employer about their disease, 72% (n=192, 1 missing data) had informed their work colleagues, and 83% (n=219, 7 missing data) had informed their occupational physician.

Factors associated with a higher educational level

Disease activity and quality of life at the time of the study were the only factors significantly associated with educational level. In a univariate analysis, active disease and low quality of life were negatively associated with a higher educational level (odds ratio [95% confidence interval (CI)] = 0.47 [0.30; 0.75] for active disease *versus* remission, p=0.001; 0.44 [0.26; 0.77] for low QoL *versus* normal QoL, p=0.003; and 1.20 [0.50; 2.05] for high QoL *versus* normal QoL, p=0.499). A young age (<10 years) at diagnosis, sex, type of IBD, disease location at diagnosis, previous surgery, and previous treatment with an anti-TNF, immunosuppressant or corticosteroid were not associated with a higher education.

The patients' perception of the impact of IBD on their education and the choice of a profession

Thirty-two percent of the patients (n=117) believed that their IBD had influenced their educational choices, 61% (n=219) felt that IBD had influenced the progression of their education, and 36% (n=130) felt that IBD had influenced their choice of a profession.

It is noteworthy that the educational level was associated with the patient's feelings about the influence of the disease on their educational choices (Figure 4A, $p<0.001$) and the progression of their education (Figure 4B, $p<0.008$). Indeed, patients who said that IBD had influenced their educational choices or the progression of their education were more likely to have a lower educational level.

The patients' free-form, written comments were also analyzed. It is noteworthy that only 66% (n=238) of the patients gave free-form, plain-language comments on their education and/or profession. The main IBD-related educational difficulty faced by patients was absenteeism (mentioned by 136 patients; Figure 5A). It is noteworthy that 10 patients mentioned a lack of school support and only three patients mentioned that their educational courses had been adapted as a function of their health. The main consequences of these difficulties were academic delays (n=37) and dropping out of school or university (n=28, Figure 5B). The age on obtainment of the highest educational qualification or degree (Supplementary Table 1) appeared to match the ages expected for each educational level.

The patients mentioned the following criteria for their choice of a profession, in decreasing order of importance: office work/seated work/a non-physical job (n=21), the need for an easy access to a toilet (n=17), the exclusion of incompatible occupations (n=14), specifically adjusted working hours/rest periods (n=11), job security/salaried employment/civil servant (n=8), no travel (n=8), ability to take leave (n=8), and a no-pressure occupation (n=5) (Figure 5C). The incompatible occupations most frequently mentioned were teaching, firefighting, or serving in the army or police. Interestingly, 18 patients stated that their disease had had a positive impact on their educational and professional choice: the time spent in hospital or in a medical environment made them want to work in the healthcare sector. This result was confirmed by the quantitative data from the self-questionnaire: 17% (n=56) of the labour forcehad a job in the healthcare sector. Fourteen percent (n=43) of the salaried patients worked in the healthcare sector; this compared with 9% (according to the INSEE) in the general population ($p=0.005$). Furthermore, 34% (n=85) of the salaried patients worked in the public sector, compared with 22% in the general population ($p<0.0001$).

DISCUSSION

Our assessment of a sample of patients with paediatric-onset IBD from a population-based registry showed that the latter's employment rate was significantly higher than that of members of the general population of the same age and from the same geographic area. Furthermore, the educational level was higher among patients with paediatric-onset IBD than among the general population. It is noteworthy that patients with paediatric-onset IBD were more likely to be working in the healthcare sector and in the public sector.

Many studies have investigated the impact of adult-onset IBD on the patients' education and professional status. In contrast, patients with paediatric-onset IBD have not been extensively assessed in this respect, and the few published studies had often small sample sizes.[17–21] The evaluation criteria also differed from one study to another: the educational level, school attendance and/or school performance, part-time working, the unemployment rate, and earnings.

Most of the literature data show that the educational level distribution is similar in patients with paediatric-onset IBD *versus* the general population.[18–23] Our results are in line with two studies that found a higher educational level in patients with paediatric-onset IBD. In Canada, El Matary *et al.* compared 112 adult patients with paediatric-onset IBD (76 with CD and 36 with UC or with IBD unclassified) diagnosed between 1978 and 2007 with a control group of 565 age- and sex-matched adults free of chronic disease.[20] The IBD patients were significantly more likely to have a university degree (88%) than the control group (73%) ($p<0.01$). In Denmark, a study based on administrative data compared 3178 patients with paediatric-onset IBD (1344 with CD and 1834 with UC) with 28204 individuals from a matched, population-based cohort; the hazard ratio [95%CI] for achieving an upper secondary educational level was higher among patients with IBD (1.14 [1.07; 1.21) for CD and 1.16 [1.10; 1.23] for UC).[24] In our study, the only factors associated with a higher educational level were quality of life and disease activity at the time of the study. Yet, the association between current disease activity and past disease activity - which might have influenced educational choices in the past - is hard to demonstrate. In our study, the educational level was not associated with the patients' clinical characteristics at diagnosis, including age, disease location, previous surgery, and previous treatment with an anti-TNF, immunosuppressant or corticosteroid. These results are in line with a Canadian study of 337 patients with paediatric-onset IBD, in which age at diagnosis, surgery, hospital admissions, and the need for corticosteroid therapy or any other therapy had no impact on the patients' educational level.[22]

In our study, the major practical difficulty encountered by patients during their education was absenteeism. This is in line with the results of a study of 50 IBD adolescents and 42 healthy adolescents in Ohio; the investigators found that 20% of the patients with IBD and only 4% of healthy subjects were absent from school for more than 3 weeks per year ($p<0.05$).[21] In our study, a large number of patients also reported impairments in their social interactions. Likewise, Hummel *et al.*'s study of 62 patients with paediatric-onset IBD and 76 healthy subjects in The Netherlands reported that the patients were

less likely to attend sports clubs (p=0.02), go to nightclubs during high school (p=0.002), and have a romantic relationship (p=0.003).[19]

To the best of our knowledge, the present study is the first to have found a higher employment rate among patients with paediatric-onset IBD. The literature data generally showed similar employment rates in patients with IBD *versus* the general population.[17,20,23] Conversely, Hummel *et al.* found that patients with IBD who had completed their university education were significantly less likely to be employed than healthy people (with employment rates of 25% and 57%, respectively; p=0.004).[19] However, Hummel *et al.*'s study featured a small number (n=36) of patients having completed their education. It is noteworthy that according to a recent study in Sweden, young adults with IBD earned significantly less than matched individuals from the general population reference but had similar employment rates.[23] Unfortunately, data on earnings were not available in our study. Another study found that young adults with IBD were more likely than control individuals to have a part-time job.[25] We did not confirm this because the frequency of full time working was similar in our patients with IBD *versus* the general population.

Interestingly, our patients with paediatric-onset IBD were more likely to work in the public sector and in the healthcare sector. To the best of our knowledge, the present study is the first to have obtained these two results. The patients reported that being in frequent contact with healthcare professionals during their childhood made them want to help others. The choice of the public sector might be related to job security and greater provisions for sick leave.

Our study had a number of strengths. Firstly, the study data came from a well-documented, population-based registry. Secondly, the patients had been diagnosed with IBD in childhood but were all aged over 25 at the time of the study; the fact that they had completed their education enabled us to study their socioprofessional status. Comparisons with the general healthy population were based on solid reference data from the INSEE (population census and employee databases) on people of the same age and in the same geographic area. Post-stratification of these data had only a limited impact on the results. Lastly, the study questionnaire used exactly the same questions as the census questionnaire, which facilitated direct comparisons of the groups' data.

Our study also had some limitations. Firstly, and as in all studies based on a self-questionnaire, the results might have been subject to recall and non-response bias. However, the response rate was satisfactory (48% of the people who actually received the questionnaire), and we compared the characteristics at diagnosis of the respondents and non-respondents: no differences were observed, other than a slightly higher proportion of patients with CD, a higher rate of CD upper digestive tract involvement, and more extensive UC among the respondents. The last two points suggest that the respondent patients would be in favor of a more severe disease; hence, less severe disease cannot explain the higher employment rate and the higher educational level observed in the patients with IBD *versus* the general population. Moreover, the disease location at diagnosis was not associated with a higher educational level. The last limitation relates to the lack of data on the socio-economic status of the

patients' parents. These variables might influence a person's educational level and professional future, independently of the disease. However, it has been shown that patients suffering from paediatric-onset IBD have a similar socio-economic level to that of the general population.[26]

Conclusion

Our results showed that relative to healthy counterparts, patients with paediatric-onset IBD have a higher educational level and a higher employment rate – thereby demonstrating a true ability to cope with the disease. However, the patients who mentioned difficulties during their education had a lower educational level, on average. The latter finding emphasizes the need for a holistic approach to the treatment of children and adolescents with IBD, in order to minimize the relapse frequency and enable patients to follow a school curriculum that is as normal as possible.

References

1 Ghione S, Sarter H, Fumery M, *et al.* Dramatic increase in incidence of ulcerative colitis and Crohn's disease (1988-2011): a population-based study of french adolescents. *Am J Gastroenterol* 2018; **113**: 265–72.

2 Kuenzig ME, Fung SG, Marderfeld L, *et al.* Twenty-first century trends in the global epidemiology of pediatric-onset inflammatory bowel disease: systematic review. *Gastroenterology* 2022; **162**: 1147–1159.e4.

3 Duricova D, Burisch J, Jess T, Gower-Rousseau C, Lakatos PL, ECCO-EpiCom. Age-related differences in presentation and course of inflammatory bowel disease: an update on the population-based literature. *J Crohns Colitis* 2014; **8**: 1351–61.

4 Herzog D, Fournier N, Buehr P, *et al.* Prevalence of intestinal complications in inflammatory bowel disease: a comparison between paediatric-onset and adult-onset patients. *Eur J Gastroenterol Hepatol* 2017; **29**: 926–31.

5 Ley D, Duhamel A, Behal H, *et al.* Growth Pattern in Paediatric Crohn Disease Is Related to Inflammatory Status. *J Pediatr Gastroenterol Nutr* 2016; **63**: 637–43.

6 Gower-Rousseau C, Sarter H, Savoye G, *et al.* Validation of the Inflammatory Bowel Disease Disability Index in a population-based cohort. *Gut* 2017; **66**: 588–96.

7 Buie MJ, Coward S, Shaheen A-A, *et al.* Hospitalization rates for inflammatory bowel disease are decreasing over time: a population-based cohort study. *Inflamm Bowel Dis* 2023; **29**: 1536–45.

8 Smyth M, Chan J, Evans K, *et al.* Cross-sectional analysis of quality of life in pediatric patients with inflammatory bowel disease in British Columbia, Canada. *J Pediatr* 2021; **238**: 57–65.e2.

9 Chouliaras G, Margoni D, Dimakou K, Fessatou S, Panayiotou I, Roma-Giannikou E. Disease impact on the quality of life of children with inflammatory bowel disease. *World J Gastroenterol* 2017; **23**: 1067–75.

10 Gower-Rousseau C, Salomez JL, Dupas JL, *et al.* Incidence of inflammatory bowel disease in northern France (1988-1990). *Gut* 1994; **35**: 1433–8.

11 Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut* 1998; **43**: 29–32.

12 Irvine EJ, Zhou Q, Thompson AK. The Short Inflammatory Bowel Disease Questionnaire: a quality of life instrument for community physicians managing inflammatory bowel disease. CCRPT Investigators. Canadian Crohn's Relapse Prevention Trial. *Am J Gastroenterol* 1996; **91**: 1571–8.

13 Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet* 1980; **1**: 514.

14 Williet N, Sarter H, Gower-Rousseau C, *et al.* Patient-reported Outcomes in a French Nationwide Survey of Inflammatory Bowel Disease Patients. *J Crohns Colitis* 2017; **11**: 165–74.

15 Silverberg MS, Satsangi J, Ahmad T, *et al.* Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol* 2005; **19 Suppl A**: 5A–36A.

16 Rey G, Rican S, Jougla E. Measurement of mortality inequalities by cause of death - Ecological approach using a social disadvantage index. [Mesure des inégalités de mortalité par cause de décès -

Approche écologique à l'aide d'un indice de désavantage social.]. *Bull Epidémiologique Hebd* 2011; 87–90.

17 Mayberry MK, Probert C, Srivastava E, Rhodes J, Mayberry JF. Perceived discrimination in education and employment by people with Crohn's disease: a case control study of educational achievement and employment. *Gut* 1992; **33**: 312–4.

18 Ferguson A, Sedgwick DM, Drummond J. Morbidity of juvenile onset inflammatory bowel disease: effects on education and employment in early adult life. *Gut* 1994; **35**: 665–8.

19 Hummel TZ, Tak E, Maurice-Stam H, Benninga MA, Kindermann A, Grootenhuis MA. Psychosocial developmental trajectory of adolescents with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 2013; **57**: 219–24.

20 El-Matary W, Dufault B, Moroz SP, Schellenberg J, Bernstein CN. Education, employment, income, and marital status among adults diagnosed with inflammatory bowel diseases during childhood or adolescence. *Clin Gastroenterol Hepatol* 2017; **15**: 518–24.

21 Mackner LM, Bickmeier RM, Crandall WV. Academic achievement, attendance, and school-related quality of life in pediatric inflammatory bowel disease. *J Dev Behav Pediatr* 2012; **33**: 106–11.

22 Singh H, Nugent Z, Brownell M, Targownik LE, Roos LL, Bernstein CN. Academic performance among children with inflammatory bowel disease: a population-based study. *J Pediatr* 2015; **166**: 1128–33.

23 Malmborg P, Everhov ÅH, Söderling J, Ludvigsson JF, Bruze G, Olén O. Earnings during adulthood in patients with childhood-onset inflammatory bowel disease: a nationwide population-based cohort study. *Aliment Pharmacol Ther* 2022; **56**: 1007–17.

24 Rasmussen J, Nørgård BM, Nielsen RG, *et al.* Implication of inflammatory bowel disease diagnosed before the age of 18 for achieving an upper secondary education: a nationwide population-based cohort study. *Inflamm Bowel Dis* 2024; **30**: 247–56.

25 Calsbeek H, Rijken M, Dekker J, van Berge Henegouwen GP. Disease characteristics as determinants of the labour market position of adolescents and young adults with chronic digestive disorders. *Eur J Gastroenterol Hepatol* 2006; **18**: 203–9.

26 Baron S, Turck D, Leplat C, *et al.* Environmental risk factors in paediatric inflammatory bowel diseases: a population based case control study. *Gut* 2005; **54**: 357–63.

Table 1: Clinical and demographic characteristics of patients with paediatric-onset IBD overall, and a comparison of respondents (n=361) and non-respondents (n=715). Significant differences are given in bold type.

	Total (n=1076)	Respondents (n=361)	Non-respondents (n=715)	p-value	Missing data
Individual characteristics					
Male sex	535 (49.7%)	170 (47.1%)	365 (51.0%)	0.220	0 (0.0%)
Age at diagnosis (y) †	15.0 [12.7 ; 16.2]	15.0 [12.9 ; 16.3]	15.1 [12.7 ; 16.2]	0.842	0 (0.0%)
Disease duration (y) †	20.5 [15.4 ; 25.9]	20.7 [15.7 ; 26.1]	20.5 [15.3 ; 25.8]	0.666	0 (0.0%)
Age at the time of the study (y) †	34.2 [29.6 ; 39.5]	34.2 [29.6 ; 39.5]	34.2 [29.6 ; 39.5]	0.926	0 (0.0%)
Disease					
CD	813 (75.6%)	286 (79.2%)	527 (73.7%)	**0.047**	0 (0.0%)
UC	263 (24.4%)	75 (20.8%)	188 (26.3%)		
Family history of IBD	133 (12.4%)	54 (15.0%)	79 (11.0%)	0.066	0 (0.0%)
Extra-intestinal symptoms	158 (14.7%)	63 (17.4%)	95 (13.3%)	0.068	0 (0.0%)
Disease activity					36 (10.0%)
Remission		141 (43.4%)			
Active disease		168 (51.7%)			
Severe disease		16 (4.9%)			
SIBDQ		54 [45 ; 61]			22 (6.1%)
Low QoL		79 (23.3%)			
Normal QoL		172 (50.7%)			
High QoL		88 (26.0%)			
CD (at diagnosis)	n=813	n=286	n=527		
Disease location					
L1	130 (16.6%)	51 (18.3%)	79 (15.7%)		
L2	197 (25.2%)	70 (25.1%)	127 (25.2%)	0.629	30 (2.8%)
L3	456 (58.2%)	158 (56.6%)	298 (59.1%)		
Upper location (L4)	220 (27.1%)	93 (32.5%)	127 (24.1%)	**0.010**	0 (0.0%)
Perineal lesions	50 (6.1%)	23 (8.0%)	27 (5.1%)	0.100	0 (0.0%)
UC (at diagnosis)	n=263	n=75	n=188		
Disease location					
E1	67 (26.0%)	9 (12.3%)	58 (31.3%)		
E2	84 (32.6%)	28 (38.4%)	56 (30.3%)	**0.007**	5 (1.9%)
E3	107 (41.4%)	36 (49.3%)	71 (38.4%)		
Characteristics of the place of residence at diagnosis					
Deprivation index					
1st quintile	137 (12.8%)	58 (16.2%)	79 (11.1%)		
2nd quintile	149 (13.9%)	46 (12.8%)	103 (14.4%)		
3rd quintile	137 (12.8%)	48 (13.4%)	89 (12.5%)	0.162	5 (0.5%)
4th quintile	152 (14.2%)	51 (14.2%)	101 (14.2%)		
5th quintile	496 (46.3%)	155 (43.3%)	341 (47.8%)		
Level of urbanization					
Densely populated	388 (36.2%)	115 (32.1%)	273 (38.3%)		
Intermediate density	424 (39.6%)	141 (39.4%)	283 (39.7%)	**0.036**	5 (0.5%)
Rural	259 (24.2%)	102 (28.5%)	157 (22.0%)		

† Median [IQR]. y: years; SIBDQ: Short Inflammatory Bowel Disease Questionnaire; QoL: quality of life; CD: Crohn's disease; UC: ulcerative colitis.

Table 2: Main occupational status of the patients with paediatric-onset IBD at the time of the study (n=361)

		Paediatric-onset IBD patients (n=361)		
		Frequency	Raw rate	Adjusted rate
Labour force population	Exercising a profession	294	81.5 %	82.3%
	Apprentice or paid trainee	2	0.5 %	0.5 %
	Unemployed job seeker	31	8.6 %	7.9 %
Out of the labour force	Student or unpaid trainee	2	0.5%	0.4 %
	Other unemployed†	32	8.9 %	8.9 %
Total		361	100%	100%

† Including homemakers and people with a disability.

Table 3: Professional status in patients with paediatric-onset IBD *versus* the general population.

	Patients with paediatric-onset IBD (n=361)			Reference data (n=1,827,384)	
	Frequency	Raw rate	Adjusted rate		p-value
Labour force population	327	90.6 %	90.7 %	89.8 %	0.578
Unemployed	31	9.5 %	8.7 %	15.0 %	0.001
Occupied	296	90.5 %	91.3 %	85.0 %	
Salaried job†	271	91.5 %	90.5 %	91.7 %	0.927
Full-time job‡	238	81.8 %	81.6 %	84.5 %	0.172

† The proportion of salaried employees was estimated for the occupied labour force.
‡ Data were missing for five participants. The proportion of full-time worker was estimated for the occupied labour force.

Figures:

Figure 1: Composition of the labour force and out of labour force populations.

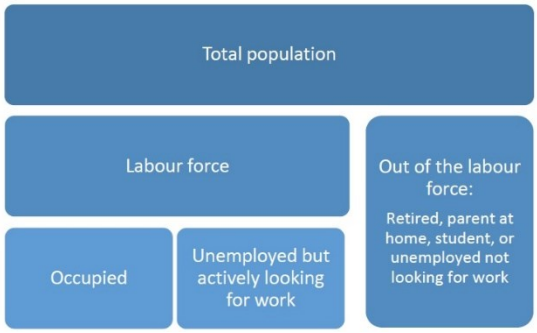


Figure 2: Study flow chart.

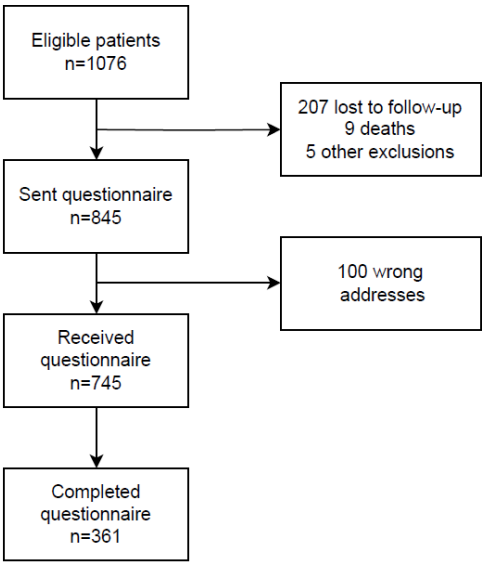


Figure 3: The highest educational qualification or degree obtained (Panel A) and the socioprofessional categories in the occupied labour force population (Panel B) among patients with paediatric-onset IBD *versus* the general population.

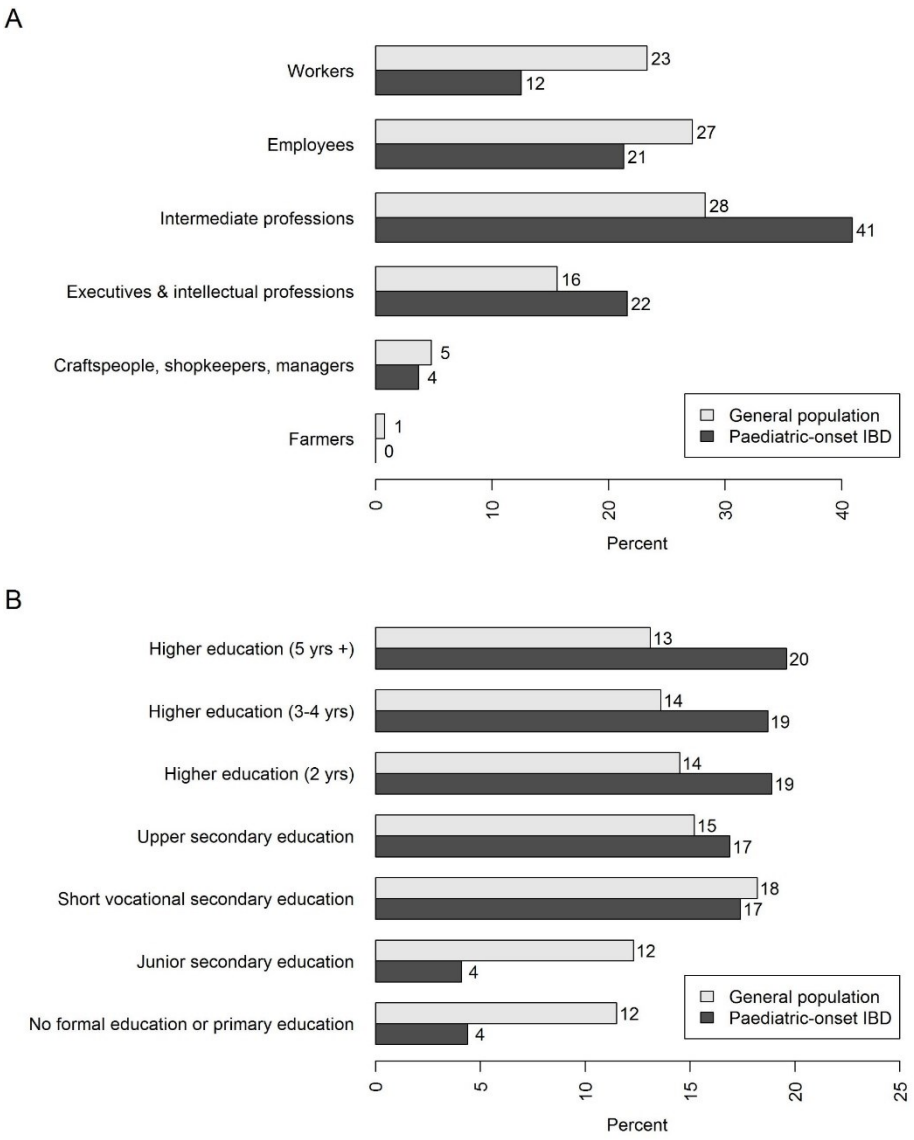


Figure 4: The relationship between the highest educational qualification or degree obtained and the reply concerning the impact of IBD on educational choices (Panel A) and the progression of education (Panel B) among patients with paediatric-onset IBD.

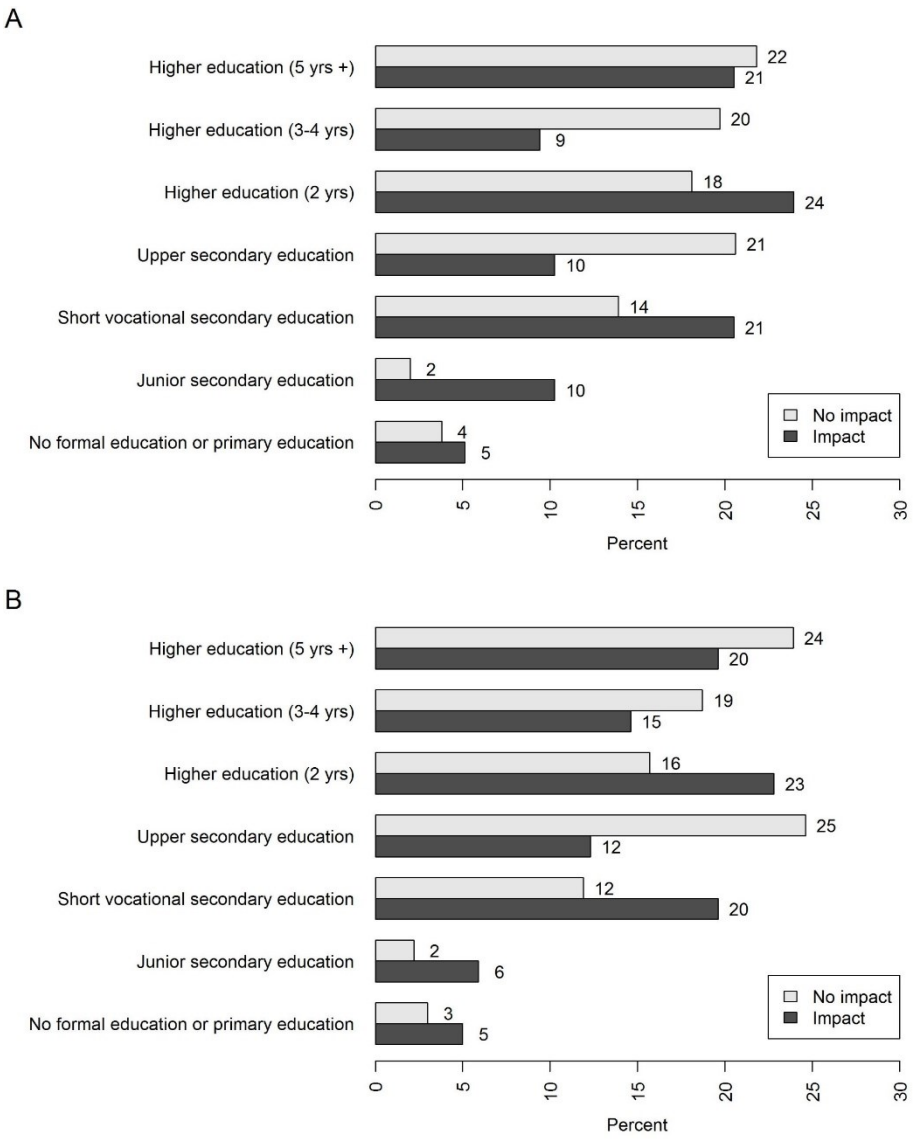
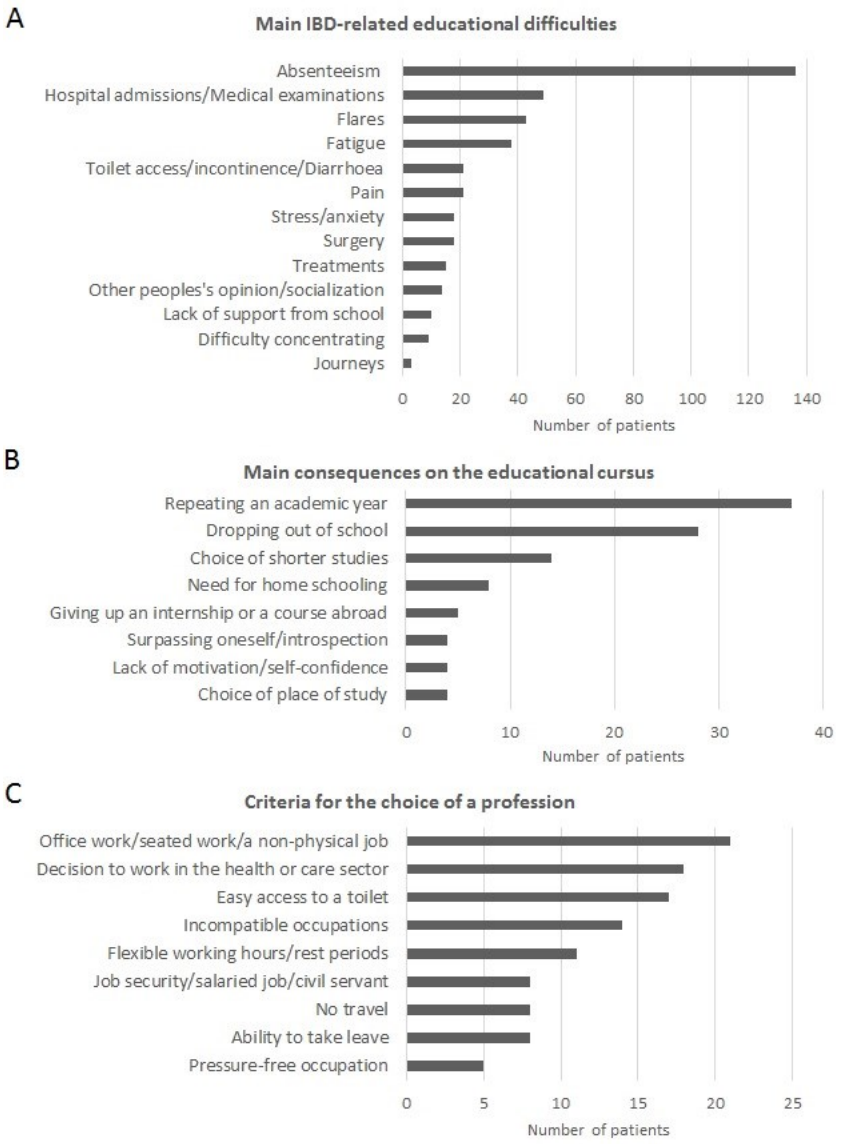


Figure 5: Data from the analysis of the patients' free-form, plain-language comments on the study questionnaire. A) The main difficulties encountered during the patients' education. B) The main impacts of IBD on the progression of patients' education. C) The patients' criteria for choosing a profession.



Supplementary Material

Supplementary Table 1: The highest educational level attained by patients with paediatric-onset IBD and who had completed their education (n=359), relative to reference data on the French general population.

	Median [interquartile range] age on obtainment of the highest educational qualification	Expected age	Patients with paediatric-onset IBD (n=359)			Reference data % (n=1,789,473)
			n	Raw rate	Adjusted rate	
No formal education or primary education only	-		16	4.4%	4.4%	11.5%
Junior secondary education	15 [15; 15]	15	17	4.7%	4.1%	12.3%
Short vocational secondary education	19 [18; 22]	17	61	17.0%	17.4%	18.2%
Upper secondary education	19 [18; 20]	18-19	62	17.3%	16.9%	15.2%
Higher education (2 years)	21 [20; 24]	20	70	19.5%	18.9%	14.5%
Higher education (3 to 4 years)	23 [21; 25]	21-22	57	15.9%	18.7%	13.6%
Higher education (5 years or more)	24 [23; 26]	23 and more	76	21.2%	19.6%	13.1%

Supplementary Table 2: Socioprofessional categories in the occupied labour force population of patients with paediatric-onset IBD (n=361), relative to reference data on the French general population.

	Patients with paediatric-onset IBD (n=361)			Reference data (n=1,396,041)	p-value
	Frequency	Raw rate	Adjusted rate		
Farmers	0	0.0%	0.0%	0.8%	<0.0001
Craftspeople, shopkeepers, and managers	10	3.4%	3.7%	4.8%	
Executives and intellectual professions	66	22.4%	21.6%	15.6%	
Intermediate professions	119	40.3%	40.9%	28.3%	
Employees	62	21.0%	21.3%	27.2%	
Workers	38	12.9%	12.5%	23.3%	
Total occupied labour force population	295[†]			1,396,041	

[†] missing data, n=1.

PARTIE 2 : Intégration de données cliniques et omiques

INTRODUCTION

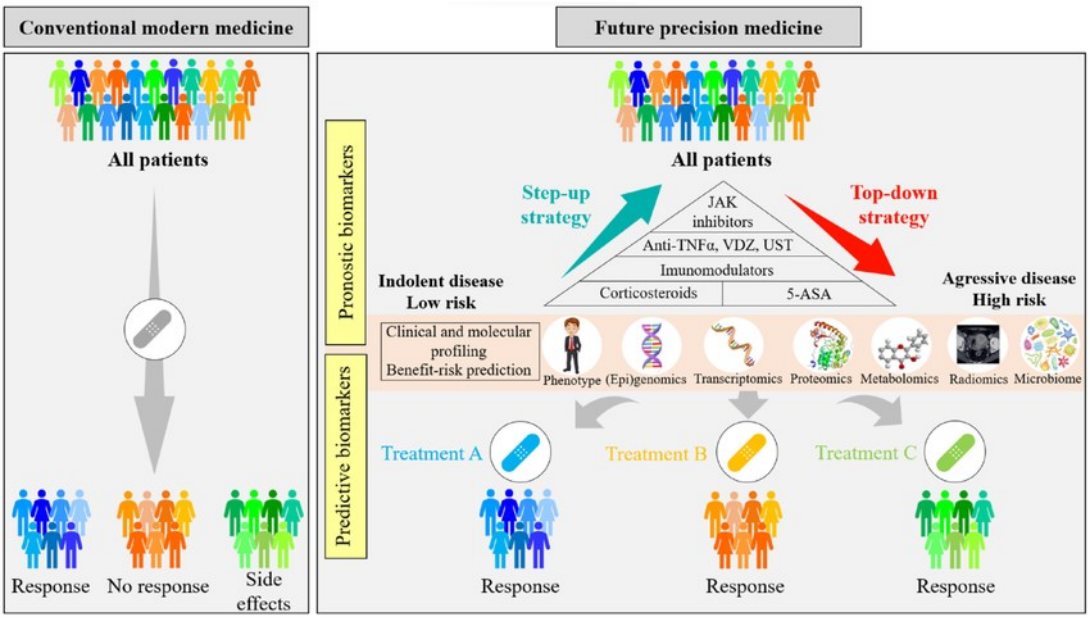
1. Médecine de précision

La médecine de précision, aussi appelée médecine personnalisée, vise à prendre en compte les particularités de chaque patient afin d'offrir une prise en charge individualisée (201). Elle s'oppose à l'approche standardisée des soins et a pour objectif d'améliorer le diagnostic, d'orienter les décisions thérapeutiques ou préventives en identifiant de manière plus précise les variabilités entre les patients. Par conséquent, une approche thérapeutique ciblée présente non seulement l'avantage d'optimiser son efficacité, mais aussi de réduire l'exposition inutile à des traitements parfois coûteux et/ou aux effets secondaires potentiellement graves.

L'objectif de la médecine de précision est ainsi de définir des sous-groupes ou strates de patients ayant des caractéristiques similaires concernant leur susceptibilité à une maladie, la gravité ou l'évolution de la maladie, ainsi que l'efficacité attendue d'un traitement. Bien que la médecine de précision soit déjà couramment utilisée en oncologie, notamment pour le traitement du cancer du sein, son intégration dans la pratique clinique demeure relativement limitée dans d'autres domaines. Cependant, la recherche dans ce domaine est en constante évolution et est particulièrement dynamique, en raison notamment de la capacité croissante d'analyser de vastes ensembles de données biologiques et génétiques, tels que le génome, le transcriptome, le protéome, le métabolome, etc., en combinaison avec les données cliniques.

On distingue généralement 3 types de biomarqueurs : diagnostiques, pronostiques et prédictifs. Les biomarqueurs « diagnostiques » sont utilisés pour identifier la présence de la maladie chez un individu, les biomarqueurs « pronostiques » visent à prédire la progression de la maladie, et enfin les biomarqueurs dits « prédictifs » s'attachent à prédire la réponse au traitement et à définir le traitement approprié pour chaque malade.

Figure 30. Médecine de précision dans les MICI (tiré de Vieujean et al, 2023)(93).



La médecine de précision en est toujours au stade d'objectif dans les MICI (92). En effet, les MICI de par leur complexité biologique et leur étiopathologie multifactorielle rendent la recherche en médecine de précision particulièrement difficile. De nombreux travaux dans la littérature ont porté sur la recherche des trois types de biomarqueurs. Les enjeux dans ce domaine sont représentés sur la figure 30.

Premièrement, des biomarqueurs diagnostiques ont été recherchés afin de prédire le déclenchement de la maladie. Ces recherches se sont tout d'abord orientées vers des facteurs génétiques puis vers des biomarqueurs biologiques, notamment sérologiques semblant plus prometteurs. Une étude à partir de patients identifiés au sein de la cohorte EPIC (European Prospective Investigation into Cancer and Nutrition) a mis en évidence que les combinaisons de marqueurs sérologiques (pANCA, ASCA, anti-CBir1 et anti-OmpC) permettaient de prédire les cas incidents de MC et de RCH avec une aire sous la courbe de 0,68 et 0,66, respectivement. La valeur prédictive augmentait lorsque le délai avant le diagnostic de la MC ou de la RCH diminuait (202). Plus récemment, il a été confirmé que l'augmentation des anticorps antimicrobiens était associée au risque de développement futur de la MC chez des apparentés sains et ce indépendamment des biomarqueurs de la fonction anormale de la barrière intestinale, de l'inflammation subclinique et des risques génétiques liés à la MC (203). Enfin, l'étude PREDICTS (Proteomic Evaluation and Discovery in an IBD Cohort of Tri-service Subjects)

a isolé un ensemble d'anticorps et de protéines sériques permettant d'identifier les patients qui seront diagnostiqués avec la MC dans les 5 années suivantes avec une bonne précision (AUC : 0,76 à 5 ans, 0,87 à 1 an) (204). En revanche, ces biomarqueurs ne permettaient pas de prédire avec précision un futur diagnostic de RCH. Ces résultats pourraient permettre dans le futur d'identifier les individus présentant un risque élevé de MC chez qui des stratégies de prévention ou d'atténuation de la maladie pourraient être mises en œuvre (prévention primaire). En effet, une intervention précoce dans la MC étant associée à de meilleurs résultats (85), on peut raisonnablement penser qu'une intervention dès la phase préclinique de la maladie pourrait plus facilement abroger ou moduler la réponse immunitaire.

Deuxièmement, des travaux ont porté sur la recherche de facteurs et de biomarqueurs dits « pronostiques » dans le but d'identifier les patients à risque de complications afin de leur proposer un traitement plus précoce et/ou plus agressif (stratégie TOP-DOWN). Les principaux facteurs cliniques associés à une évolution péjorative de la maladie de la MC sont un âge plus jeune au diagnostic (avant 40 ans), la présence de lésions anopérinéales, la localisation iléale, la localisation digestive haute (205–207). Dans la RCH, l'étendue des lésions, un âge plus jeune au diagnostic (<40 ans) et les marqueurs d'inflammation (CRP, ESR) ont été trouvés associés au risque de colectomie (208). Les contributions génétiques à l'évolution de la MC se sont principalement concentrées sur NOD2 ou plus généralement sur les variants de susceptibilité (209–216). Le rôle de NOD2 est discordant dans la littérature. L'association de NOD2 avec la sévérité de la maladie a cependant été confirmée dans une revue systématique et une méta-analyse (217). Les scores génétiques combinant tous les loci de susceptibilité connus n'ont pas prouvé être prédictifs du pronostic de la maladie (218,219) mais certains variants de susceptibilité ont été associés aux phénotypes de la maladie et à son évolution (211,220), notamment des variants de MHC, IBD5, DLG5, ATG16L1, IL23R. Plus récemment, une analyse GWAS (219) a mis en évidence 4 variants génétiques associés au pronostic de la maladie qui étaient distincts des variants de susceptibilité (gènes FOXO3, XACT, IGFBP1, MHC). Les marqueurs sérologiques, principalement les pANCA, ASCA, anti-Cbir1 et anti-OmpC, sont également associés à la maladie pénétrante/sténosante ou à la chirurgie et pourraient être de bons prédicteurs de l'évolution de la maladie. Un plus faible nombre d'études se sont intéressées à la RCH. Un variant de MHC (HLA-DRBA) a été trouvé comme associé au risque de colectomie dans la RCH (221).

La recherche d'une combinaison de marqueurs cliniques, sérologiques et génétiques a mené à la proposition de différents outils parmi lesquels on peut citer :

i) PROSPECT combinant la localisation, des marqueurs sérologiques (ASCA, anti-CBir1, pANCA) et NOD2 pour prédire le risque de complications de la MC (214).

ii) RISK : Kugathasan et al. ont proposé un modèle de stratification des risques pour prédire le phénotype compliqué de la MC. Les principales conclusions étaient qu'un âge plus avancé au moment du diagnostic, une origine afro-américaine et la séropositivité aux anticorps anti-Saccharomyces cerevisiae (ASCA) et à CBir1 étaient associés aux complications de la maladie. Une signature de gènes de la matrice extracellulaire dans les biopsiées iléales au moment du diagnostic était associée à une maladie sténosante (218). Une seconde étude basée sur des patients de l'étude RISK a identifié une signature de protéines dans le sang permettant de prédire les complications pénétrantes et sténosantes (222).

iii) PredictSURE IBD : Plus récemment, une équipe a démontré qu'une signature transcriptionnelle dans les lymphocytes T CD8 pouvait prédire l'évolution de la maladie à la fois dans la MC et dans la RCH (223). Un modèle a ensuite été développé par qPCR sur sang total et a isolé 17 gènes permettant de prédire de manière fiable le pronostic chez les patients atteints de MC et de RCH dès le diagnostic, sans nécessiter de séparation cellulaire (224). Ce premier biomarqueur pronostique est actuellement évalué prospectivement dans la MC et la RCH dans l'essai randomisé PROFILE en Grande-Bretagne (PRedicting Outcomes For Crohn's dIsease using a moLecular biomarkEr) et l'étude PRECIOUS (Predicting Crohn's and Colitis Outcomes) aux Etats-Unis. Les premiers résultats de l'étude PROFILE font cependant état d'une absence d'utilité clinique de ces biomarqueurs (225).

Troisièmement, la recherche de facteurs dits « prédictifs » associés à la réponse au traitement s'est intensifiée. Concernant la recherche de variants génétiques associés à la réponse au traitement, peu de variants ont été identifiés et leur pouvoir prédictif demeure faible. Une méta-analyse a montré que les polymorphismes de TLR2, TLR4, TLR9, TNFRSF1A, IFNG, IL6 et IL1B étaient associés à la réponse au traitement par anti-TNF (226). Les SNPs les plus significatifs dans ce contexte sont liés à l'immunité, la production de cytokines et la reconnaissance immunitaire. Des études transcriptomiques ont ensuite été menées : l'expression de TREM-1 dans les muqueuses et le sang total serait un biomarqueur de la

rémission endoscopique suite au traitement par anti-TNF. Des études ont également porté sur la dose nécessaire et l'optimisation du traitement (227). En effet, une méta-analyse récente a montré que la perte annuelle de réponse était de 10 % pour l'infliximab et de 13 % pour l'adalimumab et que les taux annuels d'escalade de dose étaient de 14 % (infliximab) et de 21 % (adalimumab), avec un bénéfice clinique dans 72 % et 52 % des cas, respectivement (228).

2. Intégration de données hétérogènes

Dans ce contexte de médecine de précision, l'essor des données "omiques" (génomique, transcriptomique, protéomique, métabolomique etc) a considérablement modifié l'échelle des données disponibles et a ouvert la voie à une meilleure compréhension des mécanismes en jeu dans l'apparition et le développement de certaines maladies complexes. L'exploitation de ces données constitue un enjeu statistique majeur et nécessite d'adapter et de développer continuellement des méthodes en bio-informatique et en statistique.

L'exploitation de ces données pose différents défis, notamment :

a) La grande dimension (p»n) (229): le nombre de variables est largement supérieur au nombre d'individus. En effet, on mesure généralement des milliers de variables sur seulement quelques dizaines d'individus, à la fois parce que les malades peuvent être « rares » mais aussi parce que ces analyses sont coûteuses. Dans ce contexte, les modèles de régression statistique classiques ne sont plus valides. L'estimateur des moindres carrés, par exemple, n'est plus défini de manière unique.
b) La corrélation entre les variables : d'une part les variables peuvent être biologiquement corrélées ; d'autre part, le grand nombre de variables implique une redondance mathématique dans les données. Les estimations des coefficients peuvent être entachées d'une forte variance.

Plusieurs types de méthodes permettent de répondre à ces deux premiers problèmes, notamment les méthodes de réduction de la dimension (création de combinaisons linéaires de variables : régression sur composantes principales (RCP) ou partial least square regression (PLS) (230), par exemple) et les méthodes de régression pénalisées ou régularisées dont les plus populaires sont Lasso (231), Rigde (232) et Elastic

Net (233). La méthode Lasso permet la sélection de variables en grande dimension et comporte de nombreuses variantes. La régression Ridge pallie les problèmes dus à la corrélation et la régression Elastic Net offre un compromis entre Lasso et Ridge.

c) La présence de valeurs manquantes : du fait du grand nombre de variables, peu de sujets sont exempts de valeurs manquantes, or les méthodes de régression classique excluent généralement les patients présentant des valeurs manquantes.
d) L'hétérogénéité des données issues de différentes sources avec des technologies différentes : les données peuvent être de type et de qualité différentes, le nombre de variables et les tailles d'effet peuvent considérablement varier selon le type de données.
e) La nécessité d'extraire l'information pertinente d'un grand volume de données : seul un petit nombre de variables est réellement associé à la réponse étudiée (principe de parcimonie). Il devient nécessaire de mettre en œuvre des méthodes permettant de sélectionner ces variables, noyées dans le « bruit » des autres variables. Cette sélection présente par ailleurs l'avantage de proposer un modèle plus facile à interpréter mais également de réduire le risque de sur-ajustement inhérent à la grande dimension.

Au moment où nous avons débuté ces travaux de thèse, différentes propositions avaient émergé pour analyser conjointement un ensemble de données cliniques et un ensemble de données omiques (234). Une première stratégie, appelée stratégie « naïve » consiste à considérer l'ensemble des données, qu'elles soient cliniques ou omiques, de la même façon. Cette stratégie n'est cependant pas satisfaisante car les variables cliniques sont susceptibles d'être noyées dans le bruit des variables omiques. D'autres stratégies ont été envisagées et comparées, notamment la stratégie « résiduelle » ou « clinical offset » en deux étapes : dans une première étape, un modèle classique de régression est réalisé sur le groupe de données cliniques ; dans une seconde étape, le prédicteur linéaire de la première étape est introduit comme « offset » dans le modèle contenant les données omiques (c'est-à-dire que le coefficient du prédicteur linéaire des données cliniques est fixé à 1). Une autre stratégie est la stratégie « favorisante » (« favoring ») qui favorise les variables du «petit» jeu de données clinique qui n'est pas (ou moins) pénalisé dans les méthodes pénalisées de type Lasso. Une stratégie appelée « réduction de dimension » consiste à réduire la dimension du jeu de données omiques et à introduire un score résumant ces données omiques dans un modèle

contenant les variables cliniques. De Bin et al. ont comparé ces différentes stratégies à l'aide de différentes méthodes statistiques (sélection univariée, Lasso, boosting, sélection forward etc.) sur deux jeux de données incluant des données cliniques et omiques sur le cancer du sein et le neuroblastome (235). Ils ont conclu qu'il était difficile d'établir la supériorité d'une méthode par rapport aux autres sur ces deux jeux de données. Ils ont ensuite ultérieurement comparé différentes méthodes statistiques et stratégies d'analyse sur des jeux de données simulés (236). D'une manière générale, le choix de la meilleure combinaison de stratégie et de méthode d'analyse dépendait de la structure de corrélation des données. Concernant les stratégies d'analyse, ils n'ont pas observé de différences majeures selon les stratégies, bien que les stratégies « naive » et de « réduction de dimension » présentaient presque toujours les plus mauvaises performances. Concernant les méthodes statistiques, les méthodes Lasso, boosting (237) et elastic-net présentaient des résultats similaires, quel que soit le degré de corrélation entre les variables. La régression Ridge quant à elle donnait les moins bons résultats.

D'autres auteurs ont suggéré d'analyser chaque modalité séparément, puis de fusionner les résultats (238). D'autres méthodes ont ensuite émergé pour analyser plus de 2 groupes de données omiques et en une seule étape : notamment l'IPF-Lasso (239) et SGCCA (240). Cependant, malgré ces travaux, la littérature restait incomplète quant à l'utilisation en pratique des différentes stratégies selon les données disponibles.

3. Contexte et objectifs

Les travaux présentés dans cette deuxième partie découlent du recueil de données génétiques et sérologiques auprès d'un échantillon de patients atteints de maladie de Crohn à début pédiatrique dans le but de proposer un score combinant données cliniques, sérologiques et génétiques pour prédire la complication de la maladie. En tant que statisticienne du registre Epimad, les écueils présentés en introduction pour l'analyse de ces données m'ont amenée à m'interroger sur les méthodes adaptées pour analyser ces données.

Les objectifs de cette partie sont ainsi doubles.

Dans un premier chapitre nous comparerons différents algorithmes pour intégrer ces données avec la finalité d'en extraire l'information pertinente (sélection de variables). Nous nous sommes focalisés dans ce travail sur l'hétérogénéité en termes de nombre de variables

et de taille d'effet, les jeux de données cliniques étant généralement de plus petite taille avec des effets plus forts que les données omiques. En termes de méthodes statistiques, nous nous sommes intéressés aux méthodes de régression pénalisées et plus particulièrement au Lasso (231) et ses différentes variantes. La finalité de ces travaux méthodologiques était de choisir une méthode adaptée à l'analyse des données du registre Epimad.

Dans un second chapitre nous analyserons les données du registre Epimad, en tirant parti des résultats du premier chapitre, afin d'intégrer des données cliniques, génétiques et sérologiques dans l'objectif de proposer un score de prédiction de complication de la MC à début pédiatrique pour une prise en charge plus adaptée des malades.

Chapitre 1 : Comparaison de méthodes d'intégration de données cliniques et omiques

Cette partie présente des travaux réalisés à partir de simulations de données.

Après une introduction aux méthodes de régressions pénalisées, une première partie présente des simulations de données mises en œuvre pour le choix du seuil pour la méthode « stability selection ». La seconde partie présente la méthode et les résultats de simulations de données réalisées pour la comparaison de méthodes d'analyse de données multi-blocs (plusieurs jeux de données) et utilise les résultats des premières simulations pour le choix du seuil pour la stability selection.

Dans ce qui suit, nous appellerons "blocs" les différents groupes de variables (également appelés modalités dans l'IPF-Lasso) qui pourraient en situation réelle correspondre à un bloc de données cliniques ou un bloc de données « omiques ».

1. Introduction aux méthodes de régressions pénalisées

Dans un cadre de grande dimension (p»n), l'hypothèse de parcimonie suppose que, parmi le grand nombre de variables disponibles, seul un nombre restreint est réellement associé à la variable à prédire. La difficulté est donc de mettre en œuvre des méthodes statistiques efficaces pour sélectionner ces variables. Cette hypothèse de parcimonie est à la base des méthodes de régression pénalisées présentées brièvement ci-dessous.

1.1. Le Lasso

Dans ce paragraphe, nous introduisons l'estimateur Lasso pour la régression linéaire. On souhaite prédire ou expliquer les valeurs d'une variable $\in \mathbb{R}^n$, appelée variable dépendante ou variable à expliquer, observées sur n individus à partir de p variables, dites indépendantes ou explicatives, représentées par les vecteurs colonnes X_1, ..., X_p. On note β le vecteur des coefficients de dimension p. β est parcimonieux au sens où seul un nombre s (s < p) de ses composantes sont égales à 0. On note S = {k : $\beta_k \neq 0$} le support de β, soit l'ensemble des valeurs non nulles de β. L'objectif est de sélectionner les variables X_j avec $j \in S$. On appellera

par la suite une variable de ce support « vraie variable » ou « variable pertinente » par opposition aux « fausses variables », hors de ce support. Le modèle linéaire s'écrit :

$$y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon$$

où ε est l'erreur résiduelle centrée.

L'estimateur des moindres carrés ordinaires (OLS) consiste à minimiser la somme des carrés des erreurs, soit :

$$\hat{\beta}^{OLS} = \underset{\beta \in \mathbb{R}^p}{argmin} \|y - X\beta\|_2^2$$

Si rang(X) = p, c'est-à-dire que les variables X_1, ..., X_p sont linéairement indépendantes, alors la solution $\hat{\beta}^{OLS}$ est unique. Dans un contexte de grande dimension, où p > n, rang(X) est inférieur à p et la solution n'est plus unique.

Afin de pallier à ce problème, des méthodes de régressions pénalisées (ou contraintes) ont été proposées. En particulier, le Lasso (Least Absolute Shrinkage and Selection Operator) impose une contrainte de norme L1 sur les coefficients β.

L'estimateur Lasso peut s'écrire :

$$\hat{\beta}^{Lasso} = \underset{\beta \in \mathbb{R}^p}{argmin} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 + \lambda ||\beta||_1 \right\}$$
$$\text{avec } ||\beta||_1 = \sum_{j=1}^{p} |\beta_j|$$

λ est le paramètre de pénalisation.

L'avantage du Lasso est qu'il force un certain nombre de coefficients à valoir 0 (selon la valeur du paramètre de régularisation λ), permettant ainsi la sélection de variables et une plus grande interprétabilité. Plus le paramètre λ augmente et plus la parcimonie augmente. En pratique, l'algorithme LARS (241) permet d'apporter une réponse à ce problème d'optimisation pour le Lasso.

Des extensions du Lasso pour les modèles linéaires généralisés et le modèle de Cox ont également été proposées (242–244).

1.2. Choix du paramètre de pénalité

Le choix du paramètre de pénalité dépend du contexte dans lequel on se place : prédiction ou sélection de variables. Dans un contexte de prédiction, la validation croisée V-Fold est généralement utilisée (en général V=5 ou 10). Cette méthode consiste à scinder l'échantillon en V groupes, à estimer le modèle sur V-1 groupes et le tester sur le groupe restant, puis à répéter cette opération pour chaque bloc servant de groupe test. Dans un cadre de sélection de variables, il a cependant été montré que lorsque le choix du paramètre de pénalité était basé sur un critère de qualité de prédiction, en général le modèle n'était pas consistant pour la sélection de variables, c'est-à-dire qu'il ne permettait pas de retrouver le vrai support (218).

Dans un cadre de sélection de variables une alternative est la *stability selection* introduite par Meinshausen et Bühlmann dans laquelle on s'intéresse non pas à la recherche d'un paramètre de régularisation optimal mais à la probabilité de sélection de chaque variable sur des rééchantillonnages de l'échantillon de départ (245).

Brièvement, soient Λ un ensemble de valeurs possibles pour λ et B réechantillonnages sans remise de taille n/2 des données. Pour chaque variable k et chaque $\lambda \in \Lambda$, on définit la probabilité de sélection :

$$\hat{\Pi}_k^{\lambda} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{k \in \hat{S}_b^{\lambda}\}$$

où $\hat{S}_b^{\lambda}$ est l'ensemble des variables sélectionnées pour un paramètre lambda donné et un rééchantillonnage b.

La sélection stable est définie par :

$$\hat{S}^{stable} = \{k : \underset{\lambda \in \Lambda}{max} \hat{\Pi}_k^{\lambda} \geq \pi_{thr}\}$$

où π_{thr} est un seuil prédéfini.

Selon Meinshausen et Bühlmann (245) les résultats sont peu sensibles au choix d'un seuil entre 0,6 et 0,9 ainsi qu'au support Λ. Le choix de ce seuil fait l'objet du paragraphe 2.2.

1.3. Extension du lasso permettant de prendre en compte des données groupées

Afin d'analyser des groupes de variables plusieurs variantes du Lasso ont été proposées, notamment :

- Le *group-Lasso* permet la sélection de groupes de variables définis a priori. L'ensemble du groupe est sélectionné. Le group-Lasso permet de regrouper des variables corrélées ou de regrouper des variables indicatrices lorsque l'on souhaite étudier des variables qualitatives (246).
- Le *sparse-group-Lasso* (SGL) est une extension du group-lasso introduisant également une pénalité au sein des groupes. La sélection au sein des groupes est régie par un unique paramètre de pénalisation, la parcimonie au sein des groupes est donc la même dans tous les groupes (247).
- L'*IPF-Lasso* introduit également une pénalisation au sein des groupes mais d'une manière plus flexible. La pénalité n'est plus unique mais varie selon les groupes. Les pénalisations sont choisies a priori ou par validation croisée (239).
- Le *multi-layer group Lasso* (MLGL) permet de sélectionner des groupes de variables corrélées à différents niveaux de la hiérarchie d'une classification hiérarchique, sans devoir fournir des groupes a priori. L'exploitation de la hiérarchie permet de réduire le temps de calcul induit par la flexibilité offerte (248).

1.4. Adaptive lasso

La sélection de variables par la méthode du Lasso a été montrée comme étant consistante sous certaines conditions, c'est à dire qu'elle retrouve les vraies variables. Zou et al. ont proposé une variante du Lasso appelée « adaptive Lasso» ayant les propriétés Oracle, c'est-à-dire étant consistante pour la sélection des variables et étant optimale pour l'estimation des paramètres (249).

L'estimateur de l'adaptive Lasso s'écrit :

$$\hat{\beta}^{Adapt} = \underset{\beta \in \mathbb{R}^p}{argmin} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j| \right\}$$

$$\text{où } w_j = 1/|\hat{\beta}_j|^{\gamma},\ \gamma > 0$$

Les $\hat{\beta}_j$ sont issus d'une première estimation : moindres carrés ordinaires, ridge (232) etc. L'idée générale de cette méthode est que l'on favorise les prédicteurs ayant un effet plus fort et qu'à l'inverse on défavorise les variables de bruit.

1.5. Une méthode multi-blocs : SGCCA

SGCCA (Sparse Generalised Canonical Correlation Analysis) (240) est une généralisation de la PLS (230) pour l'analyse de plus de deux jeux de données. Cette méthode permet en outre l'analyse des corrélations entre des variables issues de jeux de données hétérogènes tout en maximisant la discrimination entre plusieurs groupes. SGCCA est défini par le problème d'optimisation suivant, *Xj* étant le j-ème bloc de données :

$$\underset{a_1,a_2,\ldots,a_J}{argmax} \sum_{j\neq k} c_{jk}\, g(\mathrm{cov}(X_j a_j, X_k a_k))$$

$$\text{sous contraintes} \begin{cases} \|a_j\|_2^2 = 1,\ j = 1\ldots J \\ \|a_j\|_1 \leqslant \lambda_j,\ j = 1\ldots J \end{cases}$$

$$\text{avec } c_{jk} = \begin{cases} 1 & \text{si les blocs sont connectés} \\ 0 & \text{sinon} \end{cases}$$

La fonction *g* peut-être soit $g(x) = x$ ou x^2 ou $|x|$.

Les composantes latentes (combinaisons linéaires de variables) sont construites de telle sorte que la somme des covariances entre toutes les paires de blocs de données soit maximisée. Les covariances sont pondérées selon la matrice de design C. La sélection des variables est réalisée avec une pénalisation L1 sur le vecteur de coefficients des variables définissant les combinaisons linéaires. Dans un cadre supervisé, un des blocs X est remplacé par les variables indicatrices de la variable de goupe Y. Une matrice de design C contenant essentiellement des 1 privilégie les interactions entre les différents blocs alors qu'une matrice C n'ayant des 1 que pour les liens avec la réponse va privilégier des variables prédictives de la réponse.

1.6. Approches par blocs

Les méthodes SGL, IPF-Lasso et SGCCA ont l'avantage de prendre en compte la structure en blocs des données en une seule étape. Le Lasso avec validation croisée, la stability selection ou l'adaptive Lasso peuvent être appliqués à l'ensemble des données regroupant tous les blocs de données (c'est-à-dire que toutes les variables se voient accorder la même importance, méthode « naïve » présentée en introduction). Une autre approche, proposée

par Zhao et al. (238), consiste à effectuer une procédure en deux étapes : dans la première étape, les variables sont sélectionnées séparément dans chaque bloc de données ; dans la deuxième étape, les variables sélectionnées sont combinées dans une régression pénalisée.

2. Choix du seuil de stabilité

2.1. Objectif

Dans un contexte où la sélection des variables pertinentes nous intéresse plus particulièrement, la *stability selection* semble une méthode prometteuse. Cependant, le choix du seuil pour la sélection de variables reste difficile. Dans l'article original, Meinshausen et Bühlmann proposent un seuil permettant de contrôler le FWER (Family-Wise Error Rate, c'est-à-dire la probabilité d'avoir au moins 1 faux positif, c'est-à-dire une variable sélectionnée à tort) mais ce seuil i) est valide sous des conditions difficiles à vérifier ii) n'est pas facile à appliquer en pratique. Nous souhaitons donc dans un premier temps utiliser des données simulées pour proposer une méthode de choix d'un seuil qui serait déterminé à partir des données.

2.2. Méthodologie

Pour chaque variable k, la *stability selection* donne une probabilité de sélection maximale sur l'ensemble des valeurs de λ :

$$\hat{\Pi}_k^{max} = \underset{\lambda \in \Lambda}{max} \hat{\Pi}_k^{\lambda}$$

PROPOSITIONS DE METHODES DE CHOIX DU SEUIL

Pour la détermination du seuil, nous nous intéressons à la suite ordonnée de ces probabilités de sélection dont un exemple obtenu à partir de données simulées est présenté en Figure 31.

Nous proposons les seuils suivants :

i) Méthode du plus grand saut : seuil correspondant au plus grand saut observé dans les probabilités de sélection ordonnées (Figure 32).

ii) Méthode parallèle : cette méthode a été imaginée par analogie à la méthode parallèle proposée pour déterminer le nombre de composantes dans une ACP (analyse en composantes principales) (250). L'idée est de sélectionner les variables dont la probabilité de sélection

maximale dépasse celle obtenue sur des rééchantillonnages des données (Figure 33). Ces rééchantillonnages permettent de conserver la même structure des données tout en supprimant l'effet des variables. En pratique, les rééchantillonnages sont réalisés par tirage avec remise dans l'échantillon de départ.

Figure 31. Transformation du chemin de stabilité issu de la stability selection en probabilités de sélection maximales ordonnées. A) Les chemins de stabilité représentent les probabilités de sélection obtenues sur 200 réplications bootstrap de la stability selection pour chaque variable en fonction du paramètre de pénalisation λ du Lasso. Chaque courbe est associée à une variable. B) On calcule ensuite pour chaque variable le maximum de la probabilité de sélection sur l'ensemble des valeurs λ. Le graphique de gauche représente ces probabilités de manière ordonnée de la plus petite à la plus grande. Chaque point représente une variable.

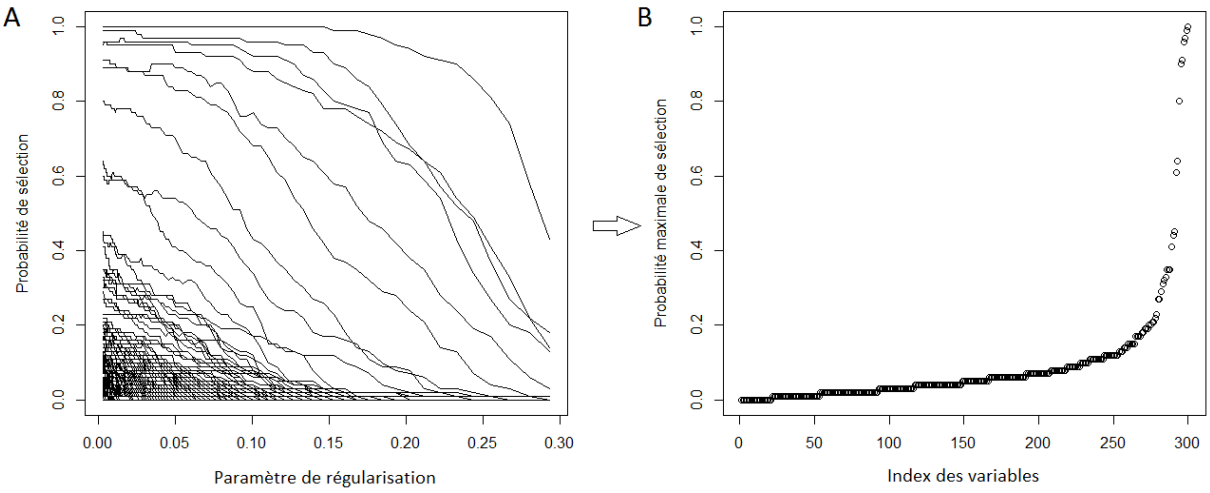


Figure 32. Illustration du principe de la méthode du plus grand saut. A) Chemins de stabilité issus de la stabilty selection. Dans cet exemple, les variables en rouge sont celles ayant été simulées comme ayant un lien avec la variable réponse. B) Le seuil choisi correspond au plus grand saut observé entre deux valeurs consécutives de la série ordonnée des probabilités de sélection maximales.

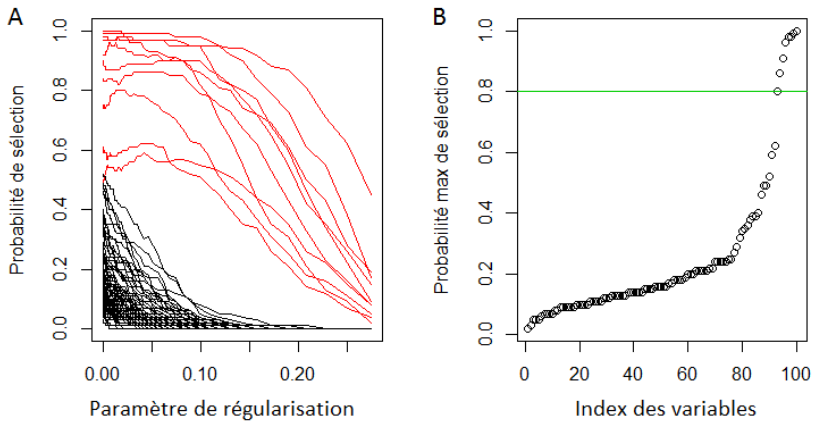
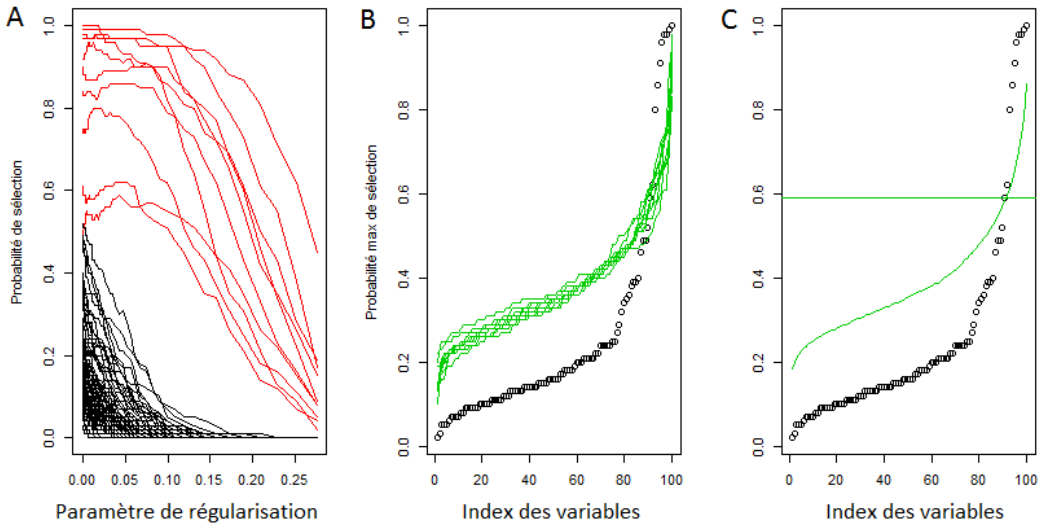


Figure 33. Illustration du principe de la méthode « parallèle ». A) Chemins de stabilité issus de la stabilty selection. Dans cet exemple, les variables en rouge sont celles ayant été simulées comme ayant un lien avec la variable réponse. B) Les points noirs correspondent à la série ordonnées des probabilités maximales de sélection observées. Les courbes vertes sont les probabilités de sélection maximales observées sur les ré-échantillonnages des données. Chaque courbe correspond à la série ordonnée pour un rééchantillonnage C) Le seuil choisi (ligne horizontale) correspond au seuil auquel la courbe moyenne des simulations parallèles et de la série ordonnée des probabilités de sélection maximales se croisent.



SCHEMAS DE SIMULATION

Différents jeux de données sont simulés en faisant varier les caractéristiques suivantes :

- Nombre de sujets : n=100
- Variable réponse Y : P(Y = 1) = 0,5
- Nombre total de variables : p = 100, 300, 500, 800, 1000
- Nombre de vraies variables (c'est-à-dire de variables ayant un réel lien avec Y) :

  p^r = (5, 10, 20, 40) pour p≤800, p^r= (10, 20, 40,100) pour p = 1000
- Taille d'effet : β = 0,3 ; 0,5 ; 1 ; 1,5

La réponse binaire Y est tirée de la distribution de Bernoulli avec une probabilité de succès de 0,5. Les variables X sont simulées à partir de distributions multinormales :

$$Y \sim \mathcal{B}(\tau = 0.5)$$
$$X_1, \ldots, X_p | Y = 0 \sim \mathcal{MN}(0_p, \Sigma)$$
$$X_1, \ldots, X_p | Y = 1 \sim \mathcal{MN}(\mu, \Sigma)$$

$$\mu' = \left(\underbrace{\beta, \ldots, \beta}_{p^r}, 0, \ldots, 0 \right)$$

où Σ est la matrice identité.

Au total, en combinant ces différents paramètres nous obtenons 80 schémas de simulation différents. Pour chaque schéma, nous réalisons 50 simulations. Pour la méthode parallèle, 30 ré-échantillonnages sont réalisés pour chaque jeu de données simulé.

METHODES COMPAREES

Sur ces simulations, les méthodes de choix du seuil suivantes sont testées :

- Seuil fixé à 0,6 ; 0,7 ou 0,8
- Méthode du plus grand saut
- Méthode parallèle

Afin de comparer les méthodes entre elles, sont calculés pour chaque simulation deux seuils de référence : i) le seuil obtenu à partir de la courbe ROC (compromis entre la sensibilité et la spécificité pour la sélection des variables pertinentes) et ii) le seuil autorisant au maximum 1 faux positif (une variable sélectionnée à tort).

CRITERES DE COMPARAISON

Nous comparons pour chaque méthode :

- Le nombre total de variables sélectionnées
- La sensibilité (probabilité pour une vraie variable d'être sélectionnée)
- Le taux de vrais positifs (parmi les variables sélectionnées)

2.3. Résultats

Les résultats de simulations montrent que :

- Le seuil dépend du nombre total de variables, du nombre de variables pertinentes et de la taille d'effet (Figure 34) ;

- Le seuil peut être inférieur à 0,5 ;
- Une taille d'effet faible (0,3) mène souvent à la sélection d'aucune variable (notamment pour un seuil fixe à 0,8 et pour la méthode parallèle). Le tableau 12 résume les résultats pour des tailles d'effet ≥ 0,5.
- Un seuil fixe supérieur ou égal à 0,6 est un seuil conservateur permettant de limiter les faux positifs mais peu sensible ;
- La méthode du "plus grand saut" ne semble pas adaptée car le plus grand saut a tendance à apparaître trop rapidement, notamment si une variable a un effet plus fort que les autres. Cette méthode est donc elle aussi conservatrice ;
- Sur des jeux de données de petite taille, la sélection « parallèle » ne permet pas de retrouver les vraies variables.
- La méthode "parallèle" est plus sensible au prix d'un plus grand nombre de faux positifs. Cependant, les caractéristiques de cette méthode en termes de sensibilité et de taux de faux positifs sont plus proches de celle obtenues à partir des seuils de "référence" (seuil obtenu par courbe ROC et seuil limitant le nombre de faux positifs à 1). (Table 18).

Figure 34. Distribution du seuil obtenu sur l'ensemble des simulations en fonction du nombre total de variables, du nombre de variables pertinentes et de la taille d'effet.

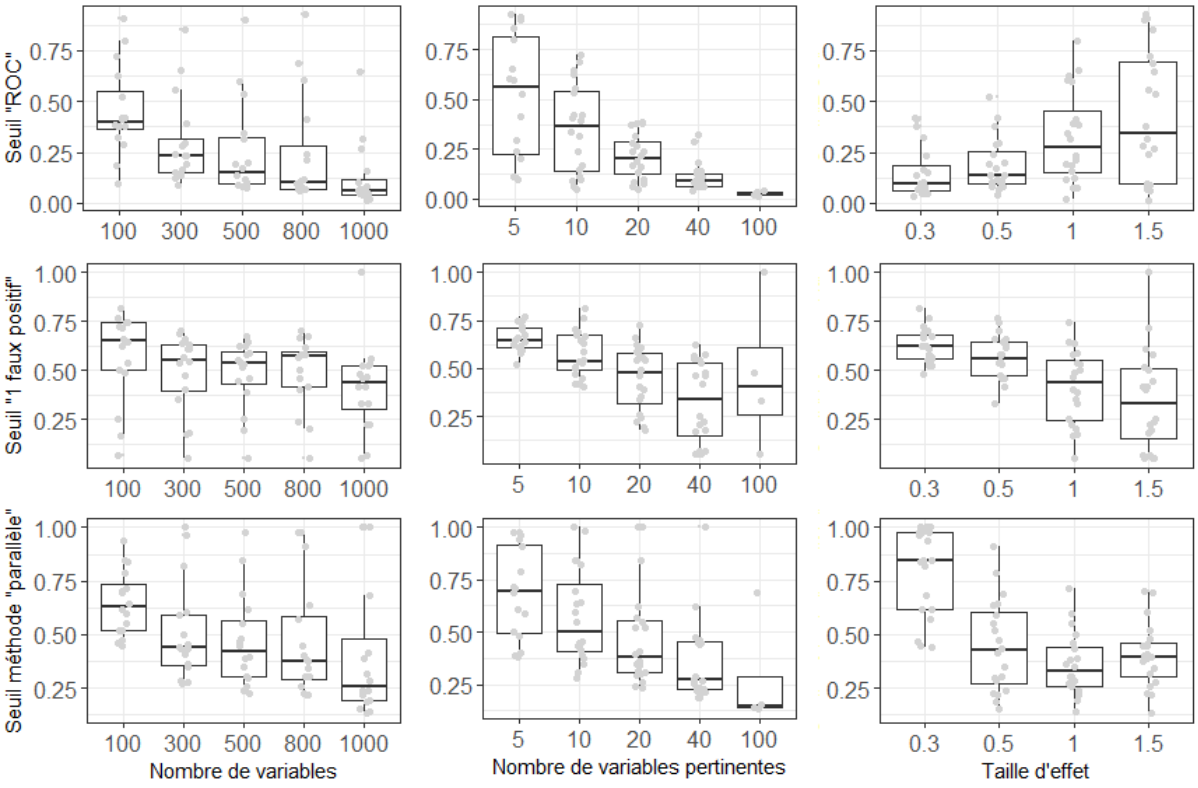


Table 18. Comparaison des méthodes de choix du seuil pour la stability selection (médianes [IQR] sur 60 schémas de simulation, excluant les schémas pour β=0,3).

	Seuil fixé à 0,6	Seuil fixé à 0,7	Seuil fixé à 0,8	Méthode du "plus grand saut"	Méthode "parallèle"	Seuil ROC	Seuil autorisant 1 faux positif
Nombre total de variables sélectionnées	9 [5-15]	7 [4-11]	5 [2-8]	6 [3-12]	17 [10-28]	29 [13-53]	11 [6-21]
Sensibilité	0,52 [0,26-0,85]	0,40 [0,20-0,74]	0,26 [0,10-0,60]	0,64 [0,14-0,90]	0,87 [0,57-1,00]	1,00 [0,90-1,00]	0,90 [0,35-1,00]
Taux de vrais positifs	1,00 [0,91-1,00]	1,00 [1,00-1,00]	1,00 [1,00-1,00]	1,00 [1,00-1,00]	0,81 [0,62-0,98]	0,82 [0,36-1,00]	0,01 [0,83-0,95]
Valeur du seuil	0,60	0,70	0,80	0,76 [0,68-0,82]	0,38 [0,27-0,50]	0,24 [0,10-0,45]	0,47 [0,25-0,59]

En conclusion, la méthode "parallèle" que nous avons proposée nous semble être un bon compromis en termes de capacité de sélection des variables pertinentes avec un nombre limité de faux positifs. La sélection de variables à partir d'un seuil fixe supérieur à 0,6 permet également de limiter le nombre de faux positifs. Ces travaux pourraient être approfondis en envisageant d'autres méthodes, notamment des méthodes basées sur le changement de variance ou de changement de pente dans la distribution des probabilités ordonnées.

3. Comparaison de méthodes de sélection de variables

L'objectif du travail suivant est de comparer différentes méthodes pour sélectionner les variables pertinentes dans un contexte de variables groupées (plusieurs blocs de données). Nous nous sommes concentrés sur une réponse binaire et sur la capacité à sélectionner les variables pertinentes dans un contexte prédictif. En effet, nous avons émis l'hypothèse que seules quelques variables sont pertinentes (hypothèse de parcimonie) : identifier uniquement les variables pertinentes pourrait améliorer les performances prédictives et rendre les outils de diagnostic/prédiction plus utilisables en pratique.

Nous nous sommes concentrés sur l'hétérogénéité des données en termes de nombre de variables et de différences de taille d'effet des variables pertinentes. Différentes méthodes ont été testées sur des ensembles de données simulées.

3.1. Méthodes

SCHEMAS DE SIMULATION

Nous utilisons 6 schémas de simulation proposés dans l'article sur l'IPF-Lasso (239) :

- Nombre de sujets : n=100
- Variable réponse binaire Y telle que P(Y = 1) = 0,5
- 2 blocs de données dont les caractéristiques suivantes varient :
    - Nombre total de variables dans chaque bloc (p_1, p_2)
    - Nombre de variables pertinentes dans chaque bloc (p_1^r,p_2^r)
    - Taille d'effet des variables pertinentes fixée au sein de chaque bloc (β_1, β_2)

Pour chaque schéma de simulation, nous avons dans un premier temps créé B = 50 jeux de données simulés en raison du temps de calcul important. Ce nombre sera par la suite augmenté à 100. La réponse binaire Y est tirée de la distribution de Bernoulli avec une probabilité de succès de 0,5. Les variables X sont simulées à partir de distributions multinormales, de la même façon que dans le paragraphe 2.2. :

$$
\begin{aligned}
&Y \sim \mathcal{B}(\tau = 0.5)\\
&X_1, \ldots, X_{p_1+p_2} | Y = 0 \sim \mathcal{MN}(0_{p_1+p_2}, \Sigma)\\
&X_1, \ldots, X_{p_1+p_2} | Y = 1 \sim \mathcal{MN}(\mu, \Sigma)
\end{aligned}
$$

$$
\mu' = \left(\underbrace{\beta_1, \ldots, \beta_1}_{p_1^r}, 0, \ldots, 0, \underbrace{\beta_2, \ldots, \beta_2}_{p_2^r}, 0, \ldots, 0 \right)
$$

Les caractéristiques des 6 principaux schémas de simulations sont présentées dans la table 19.

Ces schémas sont tout d'abord réalisés avec une matrice de corrélation identité (aucune corrélation entre les variables) puis avec une matrice de corrélation définissant une corrélation de 0,4 entre certaines variables d'un même bloc ainsi qu'entre des variables de deux blocs. Boulesteix et al. ayant conclu qu'il y avait peu de différence entre les résultats obtenus sur données corrélées et non corrélées, nous augmenterons cette corrélation à 0,6.

Table 19. Schémas de simulation de deux blocs de données.

	p_1	p_2	p_1^r	p_2^r	β_1	β_2
Schéma A	1000	1000	10	10	0,5	0,5
Schéma B	100	1000	3	30	0,5	0,5
Schéma C	100	1000	10	10	0,5	0,5
Schéma D	100	1000	20	0	0,3	
Schéma E	20	1000	3	10	1	0,3
Schéma F	20	1000	15	3	0,5	0,5

Le schéma A correspond à la situation de deux blocs de données identiques en termes de taille, de nombre et de proportion de variables pertinentes et de taille d'effets.

Dans le schéma B, la proportion de variables réellement pertinentes est la même dans les deux blocs (3 % des variables) et leurs effets sont également égaux, mais le bloc 1 est plus petit (p_1 = 100) que le bloc 2 (p_2 = 1000).

Dans le schéma C, les blocs sont de tailles différentes mais les proportions de variables réellement pertinentes sont différentes dans les deux blocs (10 % versus 1 %). Les tailles d'effet sont similaires.

Cette différence de proportion est plus prononcée dans le schéma D : ce schéma reflète une situation assez probable en pratique dans laquelle on observerait un bloc de données omiques sans effet.

Le schéma E reflète également une situation courante et qui nous intéresse plus particulièrement dans le cadre de ce travail : un petit bloc contenant des prédicteurs forts - ce qui est souvent le cas des variables cliniques ou d'un petit groupe de biomarqueurs sélectionnés – et un bloc de taille importante contenant des prédicteurs aux effets faibles correspondant à un bloc omique.

Dans le schéma F, les tailles des blocs sont les mêmes que dans le schéma E mais il y a plus de variables réellement pertinentes dans le bloc 1 et leurs effets sont forts dans les deux blocs, situation également attendue en pratique.

METHODES COMPAREES

Les méthodes comparées sont les suivantes :

- Méthodes ne tenant pas compte de la structuration en blocs de données :

    - Le Lasso standard : le choix du paramètre de régularisation est réalisé par validation croisée (5-fold, 10 répétitions)
    - Le Lasso avec stability selection au seuil 0,6
    - Le Lasso avec statibity selection et seuil « parallèle »
    - L'adaptive Lasso : la première étape est réalisée par régression ridge.
- Méthodes tenant compte de la structuration en blocs de données :
    - Le Sparse group Lasso (SGL) : le choix du paramètre de régularisation est réalisé par validation croisée 5-fold
    - L'IPF-Lasso
    - SGGCA avec corrélation entre les blocs ou non (appelés SGCCA-0 pour les blocs non corrélées et SGCCA-1 pour les blocs corrélés).
- Méthodes en deux étapes :
    - S : on réalise des Lasso séparés sur chaque bloc puis on introduit les deux prédicteurs linéaires en résultant dans une régression logistique
    - Sélection de variables dans chaque bloc par stability selection puis adaptive Lasso sur l'ensemble des variables sélectionnées. Les deux seuils i) fixe à 0,6 et ii) seuil « parallèle », seront testés.

CRITERES DE COMPARAISON

Pour chaque méthode testée, les résultats sont donnés en termes de :

- Performances prédictives : le taux d'erreur de classement estimé sur un échantillon test de taille 5000 simulé de la même façon que l'échantillon d'apprentissage
- Nombre total de variables sélectionnées : dans un objectif de parcimonie, cet indicateur est intéressant à observer en regard des autres critères
- Capacité à sélectionner les variables pertinentes : sensibilité (probabilité de sélection des « vraies » variables) et taux de vrais positifs (proportion de vraies variables parmi les variables sélectionnées).

Ces performances sont moyennées sur l'ensemble des simulations réalisées.

Les données sont analysées à l'aide des package glmnet, SGL, ipflasso et mixOmics pour SGCCA. La sability selection et l'adaptive Lasso ont été programmés à partir de la fonction glmnet et cv.glmnet du package glmnet.

3.2. Résultats

Les résultats pour chacun des 6 schémas sont présentés sur les figures 35 à 40. On peut noter les résultats suivants :

- Schéma A (deux blocs similaires) : les performances prédictives de l'ensemble des méthodes sont proches. SGL sélectionne un plus grand nombre de variables, alors que la stability selection au seuil 0,6 est la méthode la plus parcimonieuse (résultats similaires sur l'ensemble des données ou pour l'analyse par bloc). Il en résulte un taux de variables pertinentes (parmi les variables sélectionnées) largement plus élevé pour la stability selection au seuil 0,6.
- Lorsque la taille d'effet est faible (0,3), l'ensemble des méthodes a des difficultés à retrouver les vraies variables. Pour le schéma D, c'est SGCCA qui parvient à retrouver le plus de vraies variables mais au prix d'un plus grand nombre de variables sélectionnées.
- Schémas E et F : les performances prédictives sont plus contrastées entre les méthodes. La stability selection avec "méthode parallèle" semble ne pas fonctionner en petite dimension (p=20 variables) : une sélection univariée pourrait être plus adaptée pour ce cas. Pour le schéma E, l'ensemble des méthodes parvient à sélectionner les 3 variables du bloc 1 (effet fort) mais toutes ont des difficultés à sélectionner les variables à effet faible du bloc 2. SGL parvient à en sélectionner un plus grand nombre mais au prix d'un modèle moins parcimonieux. Pour le schéma F (tailles d'effet similaires dans les deux blocs mais taille des blocs et nombre de variables pertinentes différents), l'IPF-Lasso, la stability selection par bloc au seuil 0.6 et SGCCA ont des performances prédictives similaires et permettent de limiter les faux positifs. La stability selection par bloc au seuil de 0,6 permet d'obtenir une plus forte proportion de variables pertinentes parmi les variables sélectionnées.
- La corrélation entre les variables (résultats non présentés), même augmentée à 0,6 ne change pas substantiellement les résultats.
- La méthode "parallèle" présente des performances similaires à l'IPF-Lasso et SGCCA pour les schémas B et C.

D’une manière globale, la stability selection au seuil supérieur à 0,6 permet de limiter le taux de faux positifs parmi les variables sélectionnées pour un coût en taux d’erreur de classement qui semble limité.

2. Conclusion du chapitre 1

Dans l’étude qui suit sur des données réelles du registre Epimad, nous proposons de sélectionner les variables séparément sur chaque bloc de données, par des analyses univariées pour les données cliniques et par stability selection sur les données génétiques. Pour la stability selection, nous utiliserons un seuil stringeant à 0,7 afin de limiter la sélection de fausses variables. La méthode parallèle, bien que prometteuse, nécessiterait plus de travaux méthodologiques ainsi qu’une publication scientifique avant utilisation.

Figure 35. Comparaison des méthodes de sélection de variables pour le Schéma A

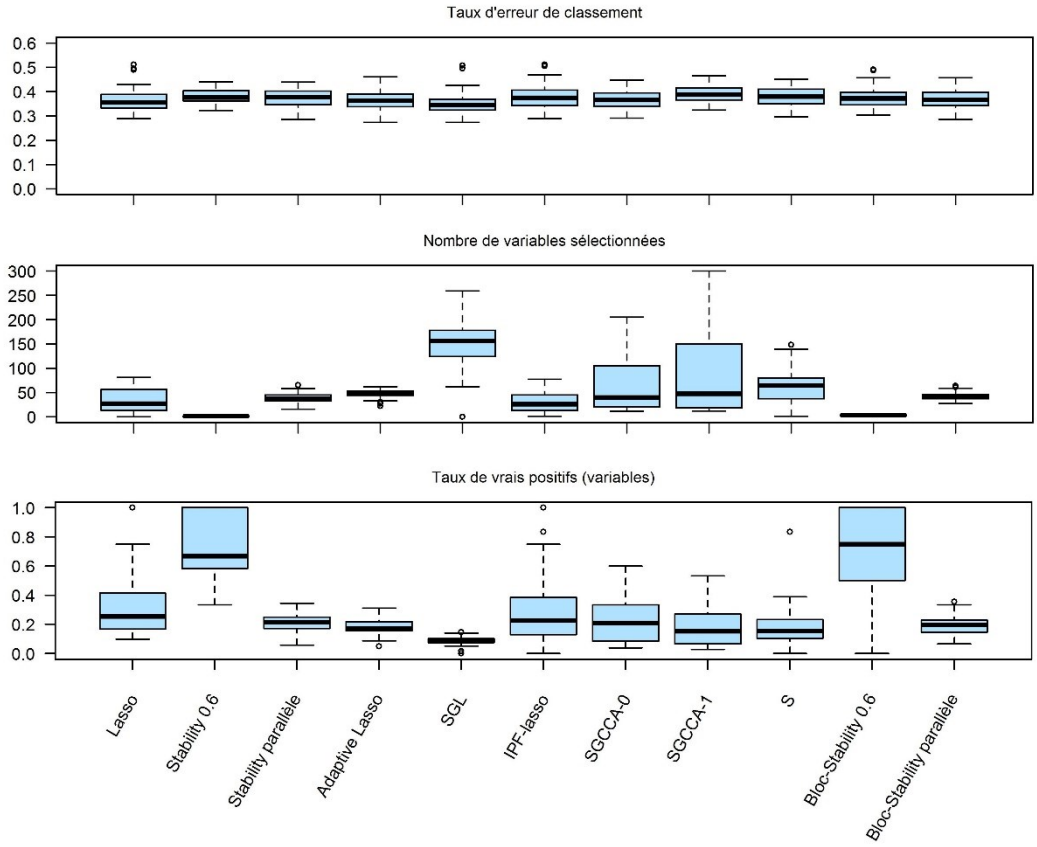


Figure 36. Comparaison des méthodes de sélection de variables pour le Schéma B

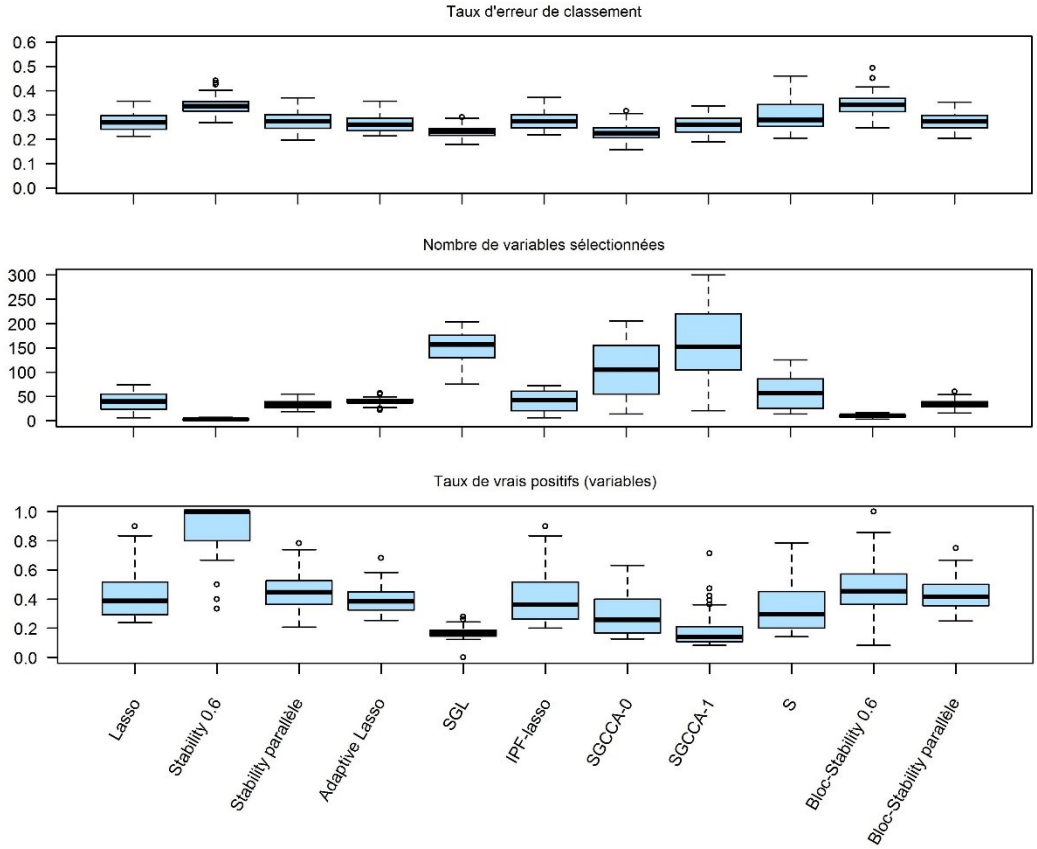


Figure 37. Comparaison des méthodes de sélection de variables pour le Schéma C

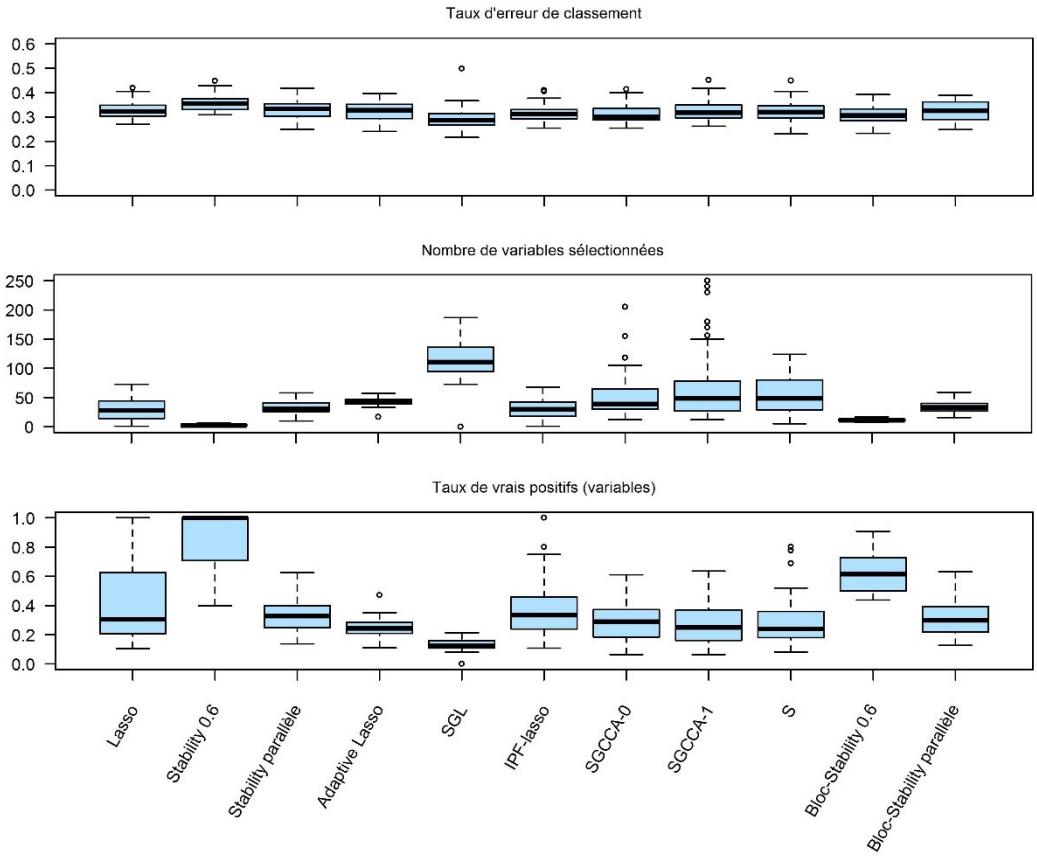


Figure 38. Comparaison des méthodes de sélection de variables pour le Schéma D

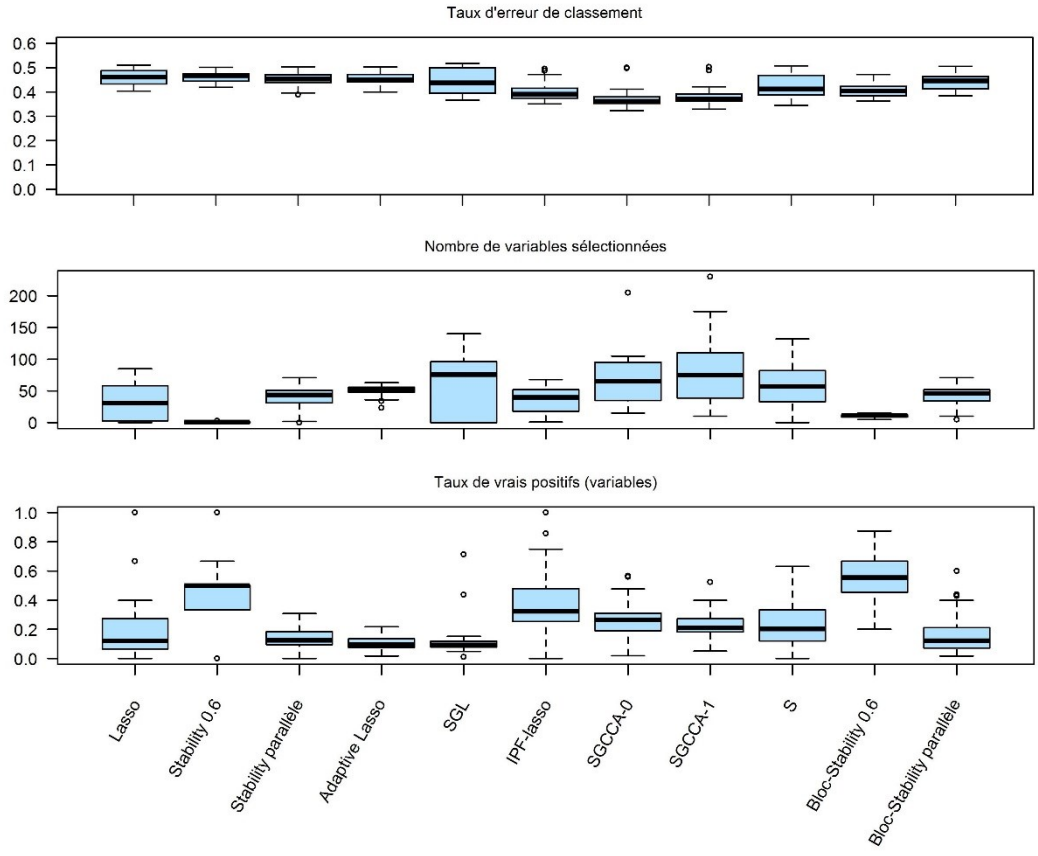


Figure 39. Comparaison des méthodes de sélection de variables pour le Schéma E

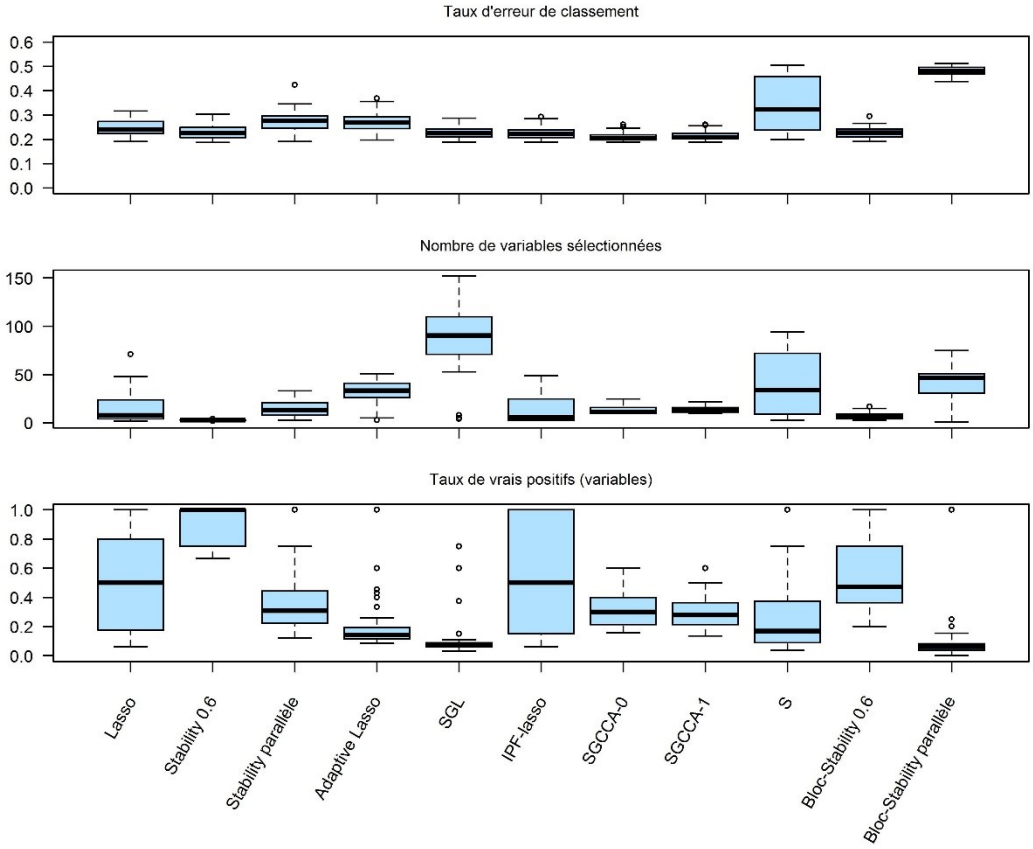
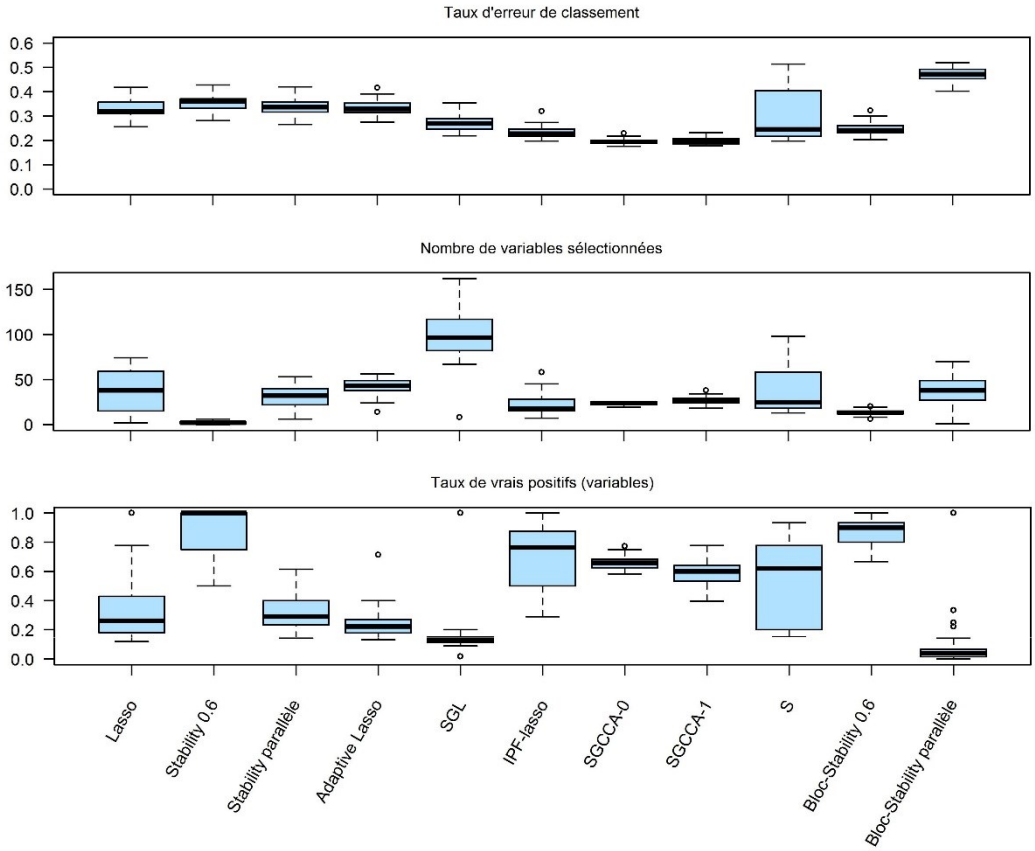


Figure 40. Comparaison des méthodes de sélection de variables pour le Schéma F



Chapitre 2 : Construction d'un score de complication de la maladie de Crohn à début pédiatrique

1. Introduction

La MC à début pédiatrique a été décrite comme une maladie plus grave que la MC adulte. Jusqu'à un tiers des enfants présente une maladie compliquée (sténosante ou pénétrante) dans les cinq années suivant le diagnostic et subit une résection intestinale précoce (76,251,252). Les évolutions les plus graves de la MC pédiatrique peuvent entraîner un retard de croissance et de puberté ainsi qu'une mauvaise qualité de vie et un niveau élevé de handicap fonctionnel, c'est-à-dire une limitation importante dans différents domaines de la vie quotidienne (limitation physique, sociale, psychologique). De nouvelles stratégies thérapeutiques basées sur l'utilisation intensive et précoce de traitements par immunosuppresseurs et/ou anti-TNF sont apparues au cours des dernières décennies afin d'obtenir une rémission sans corticostéroïdes et de prévenir la progression de la maladie et les lésions intestinales irréversibles (253,254). L'instauration d'un traitement par anti-TNF dans les 2 ans suivant le diagnostic a été associée à un faible taux de complications liées à la MC et à des taux élevés de rémission clinique et endoscopique (86). Ces thérapies sont toutefois limitées par un risque accru d'infection, de manifestations paradoxales ou de cancer (255–257). L'évolution de la maladie étant hétérogène entre les patients, il est crucial de pouvoir personnaliser les stratégies thérapeutiques et d'identifier les patients à haut risque de complications nécessitant un traitement intensif précoce dès l'apparition de la maladie.

Comme cela a été décrit en introduction générale de ce chapitre, les facteurs prédictifs d'une évolution compliquée de la MC ont fait l'objet de nombreuses études. Les modèles de prédiction existants présentent cependant un certain nombre de limites. Notamment, ils ont été développés à partir de centres uniques et/ou de centres de référence et principalement dans la population adulte, ou ne sont basées que sur un nombre limité de variables cliniques, sérologiques et génétiques. Les études spécifiquement axées sur la MC à début pédiatrique, pour laquelle l'impact de la génétique pourrait être différent, sont rares (214,218,222,258–262).

Récemment, une étude a montré que les études existantes, y compris le modèle RISK (218), n'étaient pas reproductibles (263).

L'objectif de ce travail était de développer un score basé sur des données cliniques, sérologiques et génétiques pour prédire, dès le diagnostic de la maladie, une évolution compliquée de la MC à début pédiatrique, à partir de patients en population générale issus du registre Epimad.

2. Matériel et méthodes

2.1. Population de l'étude (cohorte de découverte)

La population d'étude concerne les patients inclus dans le registre Epimad et répondant aux critères d'inclusion suivants :

- Maladie de Crohn certaine ou probable selon les critères du registre Epimad ;
- Patient présentant un phénotype non compliqué de la maladie au diagnostic (maladie purement inflammatoire, non sténosante non pénétrante : B1) ;
- Diagnostic entre le 1er janvier 1988 et le 31 décembre 2004 ;
- Suivi pendant au moins 5 ans après le diagnostic ;
- Maladie diagnostiquée avant l'âge de 17 ans ;

L'ensemble de ces patients constitue ce qu'on appelle l'échantillon de découverte (aussi appelé échantillon d'apprentissage), c'est-à-dire la cohorte à partir de laquelle le modèle prédictif est construit.

2.2. Collecte des données phénotypiques

Les données ont été extraites du registre Epimad : âge au diagnostic, sexe, localisation de la maladie, antécédents familiaux, symptômes extra-digestifs. Le phénotype de la maladie a été défini selon la classification de Montréal : B1, maladie inflammatoire (non sténosante, non pénétrante) ; B2, maladie sténosante ; et B3, maladie pénétrante. Les catégories B2 et B3 ont été regroupées et définies comme "phénotype compliqué".

Un suivi des patients a été réalisé à parti des dossiers médicaux des patients afin de recueillir les traitements, l'évolution du phénotype et le recours à une chirurgie en lien avec la MC. La chirurgie était limitée à la résection intestinale.

2.3. Définition des critères d'évaluation de la maladie compliquée

Le critère principal était défini par le critère composite : phénotype compliqué (B2 ou B3) et/ou résection intestinale dans les cinq ans suivant le diagnostic. Les critères secondaires étaient : i) la nécessité d'une résection intestinale dans les cinq ans suivant le diagnostic et, ii) un phénotype compliqué à cinq ans (B2 ou B3), analysés séparément.

2.4. Données sérologiques

Des échantillons sérums et de plasma ont été prélevés à l'inclusion. La détermination sérologique des IgA et IgG anti-Saccharomyces-cerevisiae (ASCA), de l'anticorps cytoplasmique antineutrophile périnucléaire (pANCA), du précurseur de la protéine C de la membrane externe (anti-OmpC) et des anti-flagellines (anti-CBir1, anti-Fla1 et anti-FlaX) a été effectuée par le laboratoire Prometheus (San Diego, CA). Les résultats ont été donnés en termes de positivité par rapport à un seuil spécifique à chaque anticorps déterminé par le laboratoire Prometheus.

2.5. Données génétiques

La liste complète des SNP (Single Nucleotide Polymorphisms) figure dans le matériel supplémentaire de l'article présenté en Annexe 1 de ce mémoire. La sélection des SNP a été effectuée comme suit. Tout d'abord, les voies d'intérêt pour la présente étude ont été identifiées par discussion entre le généticien (Francis Vasseur) et des gastro-entérologues experts. Ensuite, les SNP ont été identifiés à partir d'une revue systématique de la littérature effectuée par le généticien au moment de la conception de l'étude. Une recherche exhaustive dans plusieurs bases de données électroniques a été effectuée, sans restriction de langue. Un examinateur (FV) a évalué de manière indépendante le titre et le résumé des études identifiées lors de la recherche primaire en vue de leur inclusion, et le texte intégral des articles restants a été examiné afin de déterminer s'ils répondaient aux critères d'inclusion.

Au total, 373 SNP ont été sélectionnés pour le génotypage sur la base de cette analyse systématique de la littérature et comprenaient les principaux facteurs de risque génétiques de la MC (incluant les gènes NOD2, NOD1, IL23R, ATG16L1, DGL5 et IL10R1), les variants des gènes impliqués dans les voies de l'inflammation (TNFα, TNFRSF14, TNFRSF9, IL6 et NFKB1, etc.), les variants révélés dans les études d'association ou GWAS dans la MC et la RCH (gènes

PER3, WNT4, ITLN1, DNMT3A et PUS10 etc.), des variants au niveau des loci HLA-DRB1 et HLA-DQA1 dont l'implication dans la susceptibilité aux MICI a été largement démontrée, des variants de gènes codant pour des facteurs impliqués dans les interactions avec les micro-organismes (DEFB1 et DEFB4), des variants de gènes codant pour des protéines qui modulent l'immunité innée (TLR2, TLR4, DECTIN1, CARD9, etc), et enfin des variants génétiques associés à des maladies auto-immunes ou qui en modulent la gravité (PDCD1, TNFSF15, TNFRSF14 etc).

Le génotypage a été réalisé à l'aide d'une plateforme Illumina Bead Express (Illumina, Inc., San Diego, CA, USA) par le Centre National de la Recherche Scientifique (CNRS UMR8199) selon le protocole du fabricant. Deux SNP ont été génotypés à l'aide d'un test Taqman (NOD2 R702W -rs2066844 et G908R - rs2066845).

Le contrôle de qualité du génotypage a été effectué en vérifiant l'équilibre de Hardy-Weinberg (seuil : 0,05) et les fréquences des allèles mineurs telles que rapportées dans la base de données 1000genome. Les variants dont la fréquence de l'allèle mineur était inférieure à 1 % (1 variant) ou dont les données manquaient à plus de 5 % (3 variants) ont été écartés de l'analyse. Les sujets présentant plus de 5 % de génotypes manquants sur les 369 SNP restants ont été supprimés de l'analyse (n= 4). Une petite quantité de données était encore manquante - représentant 0,3 % (n=177 données) de la quantité totale de 57 464 données du tableau des 369 SNP * 156 patients - et a été imputée par tirage dans une distribution multinomiale (de probabilités issues des probabilités observées pour chaque variant).

2.6. Cohorte externe

Soixante patients pédiatriques atteints de MC non compliquée au moment du diagnostic ont été extraits d'une cohorte multicentrique française de MC publiée précédemment (209). Pour ces patients, le génotypage a été réalisé par la société Integragen (Evry, France) à l'aide d'un système Fluidgim Biomark et de la technologie AS-PCR.

Seule une partie des SNP a été génotypée dans cette cohorte externe : les variants sélectionnés à partir de l'échantillon de découverte, selon les méthodes décrites dans le paragraphe « analyses statistiques » ou significatifs en analyses univariées à $p<0,05$. Les données sérologiques n'étaient pas disponibles dans cette cohorte. Cet échantillon a été utilisé pour consolider la sélection des variants génétiques réalisée à partir de la cohorte de découverte (Epimad) avant la construction du modèle final.

2.7. Analyses statistiques

Les variables quantitatives ont été décrites par la médiane et l'intervalle interquartile (Q1-Q3) et les variables qualitatives par la fréquence et le pourcentage. Les variants génétiques ont été transformés en compte du nombre d'allèles mineurs (0, 1 ou 2) selon la pratique usuelle (264).

Pour construire un modèle prédictif combinant les données cliniques, sérologiques et génétiques, nous avons utilisé une analyse en deux étapes adaptée de Zhao et al. (238) : dans la première étape, les variables pertinentes ont été sélectionnées séparément dans chaque groupe de variables (c'est-à-dire cliniques, sérologiques et génétiques) et dans la deuxième étape, toutes les variables sélectionnées en première étape ont été incluses dans un modèle de régression finale. A cela s'ajoute une étape intermédiaire de confirmation de la sélection de variables, réalisée sur l'échantillon de découverte à l'aide de la cohorte externe. La stratégie d'analyse statistique est décrite sur la figure 41. Cette stratégie a été appliquée pour les 3 critères de complications définis précédemment.

SELECTION DES VARIABLES

Pour chaque variable expliquée (critères de complication de la maladie), les variables cliniques et sérologiques ont été sélectionnées à l'aide de régressions logistiques univariées au seuil $p \leq 0,2$.

Les génotypes ont été sélectionnés à l'aide de régressions logistiques Lasso et sélection des variables stables sur des rééchantillonnages des données (*stability selection*) (231,245).

La régression Lasso a été choisie car elle permet de modéliser la relation entre les variants génétiques et le phénotype d'intérêt et effectue la sélection des variables en même temps que l'estimation du modèle multivarié. Comme l'estimation du paramètre de régularisation λ par validation croisée tend à sélectionner un nombre trop élevé de variables, nous avons utilisé la "stability selection" pour sélectionner les variables pertinentes et éviter le sur-ajustement : nous avons sélectionné au hasard 200 échantillons bootstrap de taille n (nombre de patients) par tirage avec remise dans l'ensemble de données originales (245). L'idée est que les variables pertinentes seront plus souvent sélectionnées dans les échantillons répliqués que les variables non pertinentes qui sont susceptibles d'être plus sensibles au rééchantillonnage. Les variables sélectionnées sont celles qui atteignent une probabilité de

sélection de plus de 70 % sur les 200 réplications bootstrap. Ce seuil a été choisi afin de limiter le nombre de faux positifs (variables sélectionnées à tort) et d'obtenir un modèle parcimonieux, selon les travaux présentés dans le précédent chapitre (Partie 2 – chapitre 1).

Les variants génétiques ainsi sélectionnés par Lasso ont ensuite été testés dans l'échantillon groupé de la cohorte de découverte et de la cohorte externe afin de réduire le nombre de variants faussement sélectionnés (i.e. des variants sélectionnés alors qu'ils n'ont pas de lien avec la variable d'intérêt). Ces analyses ont été réalisées à l'aide de modèles de régression logistique univariés ajustés pour l'effet cohorte : les variants atteignant $p \leq 0,05$ ont été pris en compte pour être inclus dans le modèle final.

CONSTRUCTION DU MODELE PREDICTIF

Pour chaque variable expliquée, le modèle prédictif a été construit à partir de la cohorte de découverte en incluant toutes les variables cliniques, sérologiques et génétiques sélectionnées précédemment dans un modèle de d'adaptive Lasso logistique. Le vecteur de poids adaptatif est donné par $\hat{w} = 1/|\hat{\beta}|$ avec $\hat{\beta}$ obtenu à partir d'une régression ridge. Le paramètre optimal de pénalisation Lasso λ a été estimé par validation croisée (5-fold) et maximisation de l'aire sous la courbe ROC (AUC).

Les modèles suivants ont été testés : i) modèle incluant les variables génétiques uniquement, ii) modèles incluant variables génétiques et variables cliniques, et iii) modèle incluant variables génétiques, variables cliniques et sérologiques. Pour chaque modèle, un score a été dérivé de l'équation du modèle et le seuil de classification des patients dans les groupes à haut risque et à faible risque de complication a été déterminé à l'aide de la statistique J de Youden.

PERFORMANCE DU MODELE SELECTIONNE ET VALIDATION INTERNE :

La performance discriminante des modèles a été évaluée par :

- La courbe ROC et l'aire sous la courbe ROC,
- La sensibilité,
- La spécificité,
- La valeur prédictive positive,
- La valeur prédictive négative.

Ces caractéristiques sont présentées avec leurs intervalles de confiance à 95 % associés.

La calibration du modèle a été évaluée par le test de Hosmer-Lemeshow pour la qualité de l'ajustement, calculé en utilisant les déciles et représenté graphiquement par le graphique de calibration des probabilités observées par rapport aux probabilités prédites du modèle logistique en utilisant des splines naturels avec six degrés de liberté.

Le facteur de réduction (« shrinkage factor ») a également été estimé (265,266). Ce facteur mesure la réduction estimée des coefficients de régression permettant d'améliorer la prédiction pour les futurs patients. Cette idée est complémentaire du sur-ajustement. En effet, lors d'analyses de régression, une estimation fonctionne généralement moins bien sur un nouvel ensemble de données que sur l'ensemble de données utilisé pour l'ajustement. Ce facteur permet de corriger les estimations afin d'améliorer la prédiction.

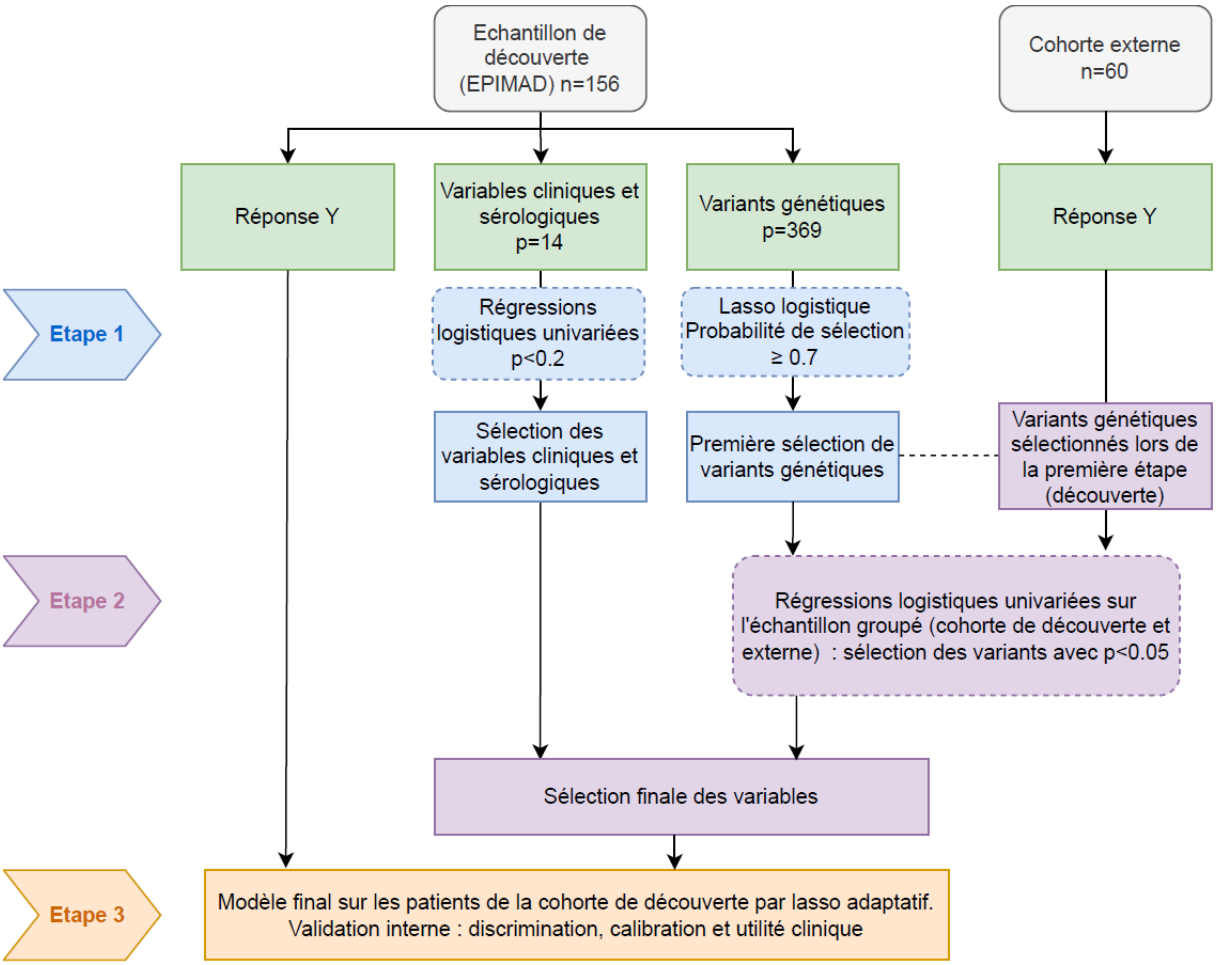
La validation interne des modèles développés a été réalisée en calculant : i) les performances apparentes, ii) les performances corrigées du biais d'optimisme. Ceci a été réalisé pour l'ensemble des performances prédictives citées ci-dessus.

Les performances prédictives apparentes sont les performances calculées directement sur l'ensemble de données originales (c'est-à-dire que l'ensemble de données original est utilisé à la fois comme échantillon d'apprentissage et comme échantillon test). Les performances prédictives ont été corrigées du biais d'optimisme à l'aide de 1 000 échantillons bootstrap de taille n tirés avec remise à partir de l'ensemble de données originales. Le facteur de réduction a également été estimé à partir des échantillons bootstrap.

Enfin, l'utilité clinique du modèle a été évaluée à l'aide d'une analyse de la courbe de décision, représentant le bénéfice net standardisé en fonction des seuils de probabilité et comparant "PREDICT-EPIMAD" aux stratégies "traiter tout le monde" et à la stratégie "ne traiter personne" (267,268). Le bénéfice net est calculé comme suit : (proportion de vrais positifs) - (proportion de faux positifs) * pt/(1-pt), où pt est le seuil décisionnel. Le bénéfice net a été standardisé par le taux d'événements observé dans notre échantillon, de sorte que le bénéfice standardisé maximal soit égal à un. Comme recommandé, la courbe de décision n'a été tracée que pour une gamme raisonnable de probabilités seuils. La zone des probabilités seuils a été déterminée après discussion avec 6 gastro-entérologues experts. Plus précisément les questions suivantes leur ont été posées : « En sachant que le risque de complications à 5 ans est de 35 % et en intégrant votre perception des risques liés au traitement ainsi que des

risques liés à la complication de la maladie : i) combien de patients seriez-vous prêts à traiter pour éviter une complication chez un patient?, ii) si on vous donne pour chaque patient la probabilité estimée que sa maladie se complique : en dessous de quel seuil, sans hésiter, ne traiteriez-vous pas le patient? Au-dessus de quel seuil, sans hésiter, traiteriez-vous le patient? »

Figure 41. Représentation graphique de la stratégie d'analyse statistique.



Les données ont été analysées à l'aide des logiciels SAS V.9.4 (SAS, SAS Institute Inc., Cary, NC, USA) et R 3.4.3. La régression Lasso a été réalisée à l'aide du package R glmnet.

Cette étude a obtenu l'accord du CPP Nord-Ouest IV (IRB : 2007 /32-MS1).

3. Résultats

3.1. Caractéristiques des patients au moment du diagnostic

Au total, 156 patients (garçons : 54 %) atteints de MC à début pédiatrique et présentant un phénotype inflammatoire (B1) au moment du diagnostic ont été inclus, avec un suivi médian de 10,4 ans (intervalle interquartile (IQR) : [7,2 - 14,9]). Les caractéristiques des patients au moment du diagnostic sont présentées dans la table 1. L'âge médian au moment du diagnostic était de 14,3 ans (IQR : [11,9 - 16,0]). La principale localisation de la maladie au moment du diagnostic était iléo-colique (n = 109, 70 %). Soixante pour cent (n = 93) et 19 % (n = 29) des patients ont été exposés à un traitement immunosuppresseur et à un anti-TNF au cours des 5 années suivant le diagnostic de la MC, respectivement (Table 20).

3.2. Evolution compliquée de la maladie

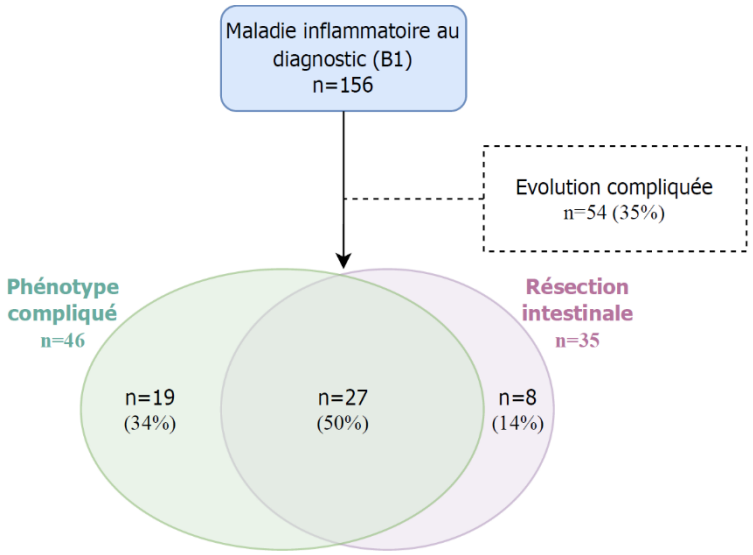
Parmi les 156 enfants atteints d'une maladie inflammatoire (B1) au moment du diagnostic, 22 % (n = 35) avaient subi une résection intestinale et 29 % (n = 46) avaient évolué vers un phénotype compliqué (B2, n = 32 (70 %), B3, n = 14 (30 %)) dans les cinq années suivant le diagnostic. Le critère de complication principal (c'est-à-dire le critère composite : résection intestinale et/ou phénotype compliqué dans les 5 ans) était observé chez 35 % des enfants (n = 54) (Table 20, Figure 42).

Table 20. Caractéristiques cliniques de la cohorte de découverte Epimad (n = 156).

Variable	n (%)
Durée de suivi (médiane, IQR)	10,4 [7,2 – 14,9]
Sexe masculin	85 (54,5 %)
Caractéristiques cliniques au diagnostic	
Age (médiane, IQR)	14,3 [11,9 – 16,0]
Antécédents familiaux de MICI	24 (15,4 %)
Localisation au diagnostic	
Iléale (L1)	17 (11,0%)
Colique (L2)	29 (18,7 %)
Ileo-colique (L3)	109 (70,3 %)
Atteinte digestive haute (L4)	60 (38,5 %)
Lésions ano-périnéales	18 (11,5 %)
Symptômes extra-digestifs	39 (25,0 %)
Traitements dans les 5 ans suivant le diagnostic	
Corticoïdes systémiques	126 (80,8 %)
5-ASA systémiques	133 (85,3 %)
Immunosuppresseurs	93 (59,6 %)
Anti-TNF	29 (18,6 %)
Statut sérologique à l'inclusion	
ASCA IgA +	111 (71,1 %)
ASCA IgG +	135 (86,5 %)
pANCA +	28 (17,9 %)
Anti-CBir1 +	80 (51,3 %)
Anti-Fla2 +	43 (27,6 %)
Anti-FlaX +	44 (28,2 %)
Anti-OmpC +	23 (14,7 %)
Evolution à 5 ans	
Résection intestinale et/ou phénotype compliqué	54 (34,6 %)
Résection intestinale	35 (22,4 %)
Phénotype compliqué (B2/B3)	46 (29,5 %)

IQR : intervalle inter-quartile

Figure 42. Complication de la maladie dans les 5 ans suivant le diagnostic



3.3. Sélection des variables cliniques et sérologiques

Ces variables ont été sélectionnées au seuil de signification p ≤ 0,2. Les analyses univariées des variables cliniques et sérologiques sont présentées dans la table 21. Pour le critère composite, la localisation iléale ou iléo-colique de la maladie au moment du diagnostic (OR [IC 95%] : 1,87 [0,74 ; 4,71], p=0,187), la positivité aux pANCA (OR : 0,18 [0,05 ; 0,63], p=0,007), la positivité aux ASCA-IgG (OR : 2,50 [0,80 ; 7,85], p=0,116) et la positivité aux anti-OmpC (OR : 1,92 [0,78 ; 4,70], p=0,154) ont été sélectionnées. Ainsi, la localisation iléale de la maladie était associée à une évolution péjorative de la maladie. Les variables sérologiques étaient également associées à une évolution péjorative à l'exception des pANCA dont la positivité était associée à une évolution favorable de la maladie.

Table 21. Analyses univariées des variables cliniques et sérologiques (n = 156 patients). Les variables surlignées en gris sont celles atteignant p ≤ 0.2 et sélectionnées pour la construction des modèles prédictifs.

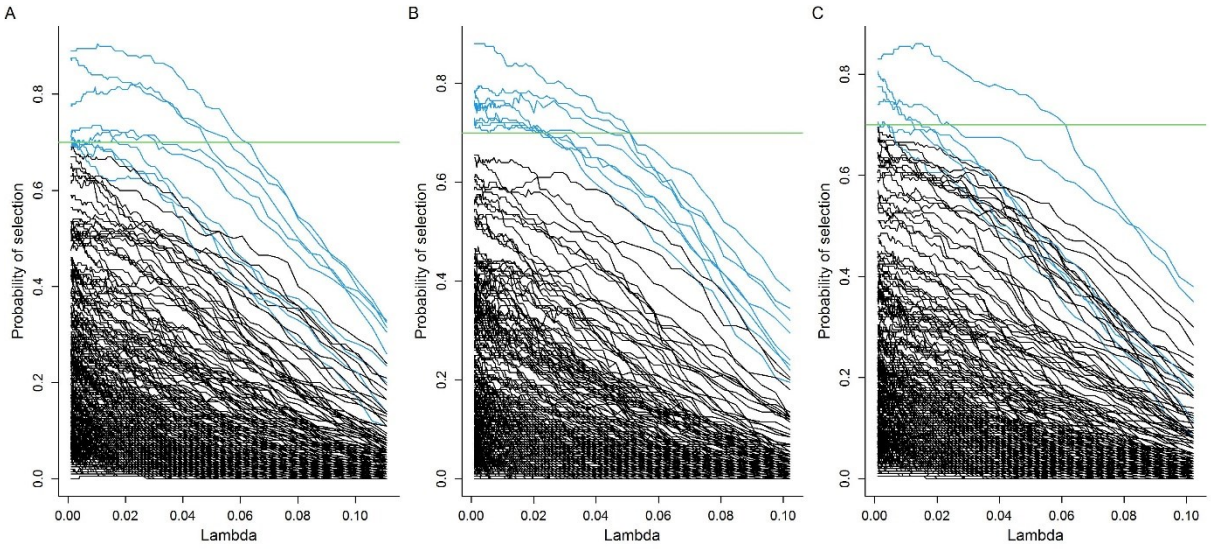
	Critère principal: resection intestinale ou phenotype compliqué à 5 ans		Résection intestinale à 5 ans		Phénotype compliqué à 5 ans	
	OR (IC 95%)	p-value	OR (IC 95%)	p-value	OR (IC 95%)	p-value
Variables cliniques au diagnostic						
Sexe masculin	1,07 [0,55; 2,07]	0,845	0,85 [0,40; 1,81]	0,680	1,45 [0,72; 2,91]	0,302
Age au diagnostic†	1,07 [0,94; 1,21]	0,319	1,05 [0,90; 1,21]	0,540	1,07 [0,93; 1,22]	0,333
Symptômes extra-digestifs	1,08 [0,50; 2,30]	0,846	0,86 [0,35; 2,09]	0,740	1,08 [0,49; 2,39]	0,839
Antécédents familiaux de MICI	1,43 [0,59; 3,47]	0,431	1,53 [0,58; 4,05]	0,392	1,24 [0,49; 3,13]	0,654
Localisation L1 ou L3	1,87 [0,74; 4,71]	0,184	1,15 [0,43; 3,08]	0,787	1,78 [0,67; 4,72]	0,244
Atteinte digestive haute (L4)	0,71 [0,36; 1,42]	0,339	0,67 [0,30; 1,50]	0,333	0,91 [0,45; 1,86]	0,803
Lésions ano-périnéales	0,94 [0,33; 2,65]	0,903	0,66 [0,18; 2,43]	0,535	0,91 [0,30; 2,78]	0,866
Variables sérologiques à l'inclusion						
ASCA-IgA +	1,25 [0,59; 2,62]	0,558	1,22 [0,52; 2,87]	0,643	1,42 [0,65; 3,13]	0,380
ASCA-IgG +	2,50 [0,80; 7,85]	0,116	3,07 [0,68; 13,89]	0,145	1,92 [0,61; 6,05]	0,267
pANCA +	0,18 [0,05; 0,63]	0,007	0,22 [0,05; 0,98]	0,048	0,24 [0,07; 0,83]	0,024
Anti-CBir1 +	0,82 [0,43; 1,60]	0,569	0,75 [0,35; 1,59]	0,455	1,05 [0,53; 2,09]	0,885
Anti-Fla2 +	0,65 [0,30; 1,41]	0,279	0,67 [0,34; 1,31]	0,240	0,89 [0,41; 1,96]	0,789
Anti-FlaX +	0,73 [0,34; 1,54]	0,405	0,70 [0,29; 1,69]	0,426	1,00 [0,47; 2,16]	0,992
Anti-OmpC +	1,92 [0,78; 4,70]	0,154	1,27 [0,46; 3,50]	0,650	2,07 [0,83; 5,14]	0,116

†Les ORs sont donnés pour une augmentation d'un an d'âge. OR : odds ratio; IC 95% : intervalle de confiance à 95%.

3.4. Sélection des variants génétiques

Neuf variants ont été sélectionnés pour le critère principal par lasso et stability selection (Figure 43). Parmi les variants sélectionnés, six (dans les gènes NFKB1, IKZF1, UBE2D1, IATPR, TNFSF11 et IKZF3) ont été confirmés en utilisant l'échantillon groupé des cohortes de découverte et externe (Table 22).

Figure 43. Chemins de stabilité pour la sélection des variants génétiques par Lasso avec stability selection au seuil de 0,7. Cette figure représente les probabilités de sélection sur 200 réplications bootstrap pour chaque variable en fonction du paramètre de pénalisation λ du Lasso. Chaque courbe est associée à une variable. Les variables atteignant le seuil de 0,7 (ligne verte horizontale) ont été sélectionnées. Ces variables sélectionnées sont représentées en bleu. A) Critère composite à 5 ans B) Critère chirurgie à 5 ans C) Critère phénotype compliqué à 5 ans.



Pour les critères secondaires, neuf et six SNP ont été sélectionnés par régression Lasso pour la résection intestinale et le phénotype compliqué à 5 ans, respectivement. Quatre SNP (dans les gènes IHPK1, HMGB1, NOD2 et ORMDL3) ont été confirmés pour la résection intestinale en utilisant l'échantillon groupé des cohortes de découverte et externe et 5 (dans les gènes TLR5, PRDM1, IKZF1, KLF6 et AKT1) pour le phénotype compliqué (Table 22).

Table 22. Variants génétiques retenus pour être inclus dans le modèle multivarié pour chaque critère de complication. Les allèles à risque sont mis en évidence en gras.

Chromosome	Gène	SNP	Allèle mineur dbSNP (CEU)	Allèle majeur dbSNP (CEU)	Fréquence de l'allèle mineur dans l'échantillon Epimad		
					Tous (n = 156)	Absence de complication	Complication
Critère composite : résection intestinale ou phénotype compliqué à 5 ans							
chr4	NFKB1	rs230530	G	**A**	0.47	0.52	0.36
chr7	IKZF1	rs1456896	**C**	T	0.27	0.22	0.36
chr10	UBE2D1	rs1819658	T	**C**	0.18	0.22	0.11
chr10	IATPR	rs2755996	**T**	C	0.10	0.07	0.16
chr13	TNFSF11	rs2062305	**G**	A	0.49	0.43	0.60
chr17	IKZF3	rs907091	**T**	C	0.50	0.44	0.61
Résection intestinale à 5 ans							
chr3	IHPK1	rs9872864	G	**A**	0.51	0.55	0.37
chr13	HMGB1	rs1045411	T	**C**	0.23	0.27	0.10
chr16	NOD2	rs2066845	**C**	G	0.07	0.05	0.14
chr17	ORMDL3	rs8076131	G	**A**	0.49	0.54	0.33
Phénotype compliqué à 5 ans							
chr1	TLR5	rs851192	G	**C**	0.39	0.42	0.30
chr6	PRDM1	rs548234	C	**T**	0.31	0.34	0.26
chr7	IKZF1	rs1456896	**C**	T	0.27	0.23	0.35
chr10	KLF6	rs6601764	C	**T**	0.44	0.50	0.32
chr14	AKT1	rs2494731	**C**	G	0.31	0.28	0.40

3.5. Construction du modèle prédictif et validation interne

Pour chaque critère, les variables sélectionnées ont été introduites dans des modèles de régression logistique de type adaptive Lasso. Les performances discriminantes des modèles sont présentées dans la Table 23. Les coefficients des modèles et les seuils de décision sont présentés en table supplémentaire de l'article (Annexe 1).

Sur l'ensemble des modèles, les AUC variaient entre 0,77 et 0,84 avec des intervalles de confiance qui se chevauchaient (AUC corrigées entre 0,72 et 0,80).

Nous avons considéré que le meilleur modèle prédictif était celui qui atteignait l'AUC la plus élevée avec le plus petit nombre de variables. Le meilleur modèle prédictif (appelé PREDICT-EPIMAD dans la suite) était ainsi celui qui prédisait le critère composite et incluait la localisation iléale, 6 SNP (gènes NFKB1, IKZF1, UBE2D1, IATPR, TNFSF11, et IKZF3), et la positivité des pANCA. Ce modèle atteignait une AUC de 0,84 [IC à 95 % : 0,78 ; 0,90] dans l'échantillon de découverte. Les performances corrigées du biais d'optimisme étaient une AUC

de 0,80 [0,73 ; 0,85] (Figure 44, panel A), une sensibilité de 0,79 [0,67 ; 0,88], une spécificité de 0,74 [0,62 ; 0,83], une VPP de 0,61 [0,48 ; 0,73], et une VPN de 0,87 [0,80 ; 0,93]. L'introduction des pANCA dans le modèle améliorait les performances prédictives. La spécificité augmentait de 0,67 [0,48 ; 0,86] à 0,74 [0,62 ; 0,83] et la VPP de 0,56 [0,42 ; 0,73] à 0,61 [0,48 ; 0,73], alors que la sensibilité et la VPN restaient stables. L'ajout de la localisation et des pANCA permettait ainsi de réduire le nombre de faux positifs, c'est-à-dire de réduire le nombre de patients qui seraient traités à tort (Figure 45, Table 23).

Figure 44. Résultats de la validation interne du modèle basé sur 6-SNP prédisant le critère composite (PREDICT-EPIMAD). A) Courbe ROC de PREDICT-EPIMAD (ligne noire) comparée au modèle incluant uniquement des variables cliniques (ligne bleue). L'AUC corrigée correspond à l'AUC corrigée du biais d'optimisme à l'aide de 1 000 échantillons bootstrap. B) Probabilité prédite de complications : rectangle rouge : patients à "haut risque" selon PREDICT-EPIMAD, rectangle bleu : patients à "faible risque" selon PREDICT-EPIMAD. Les points rouges et bleus représentent respectivement la présence ou l'absence de complications observées à cinq ans.

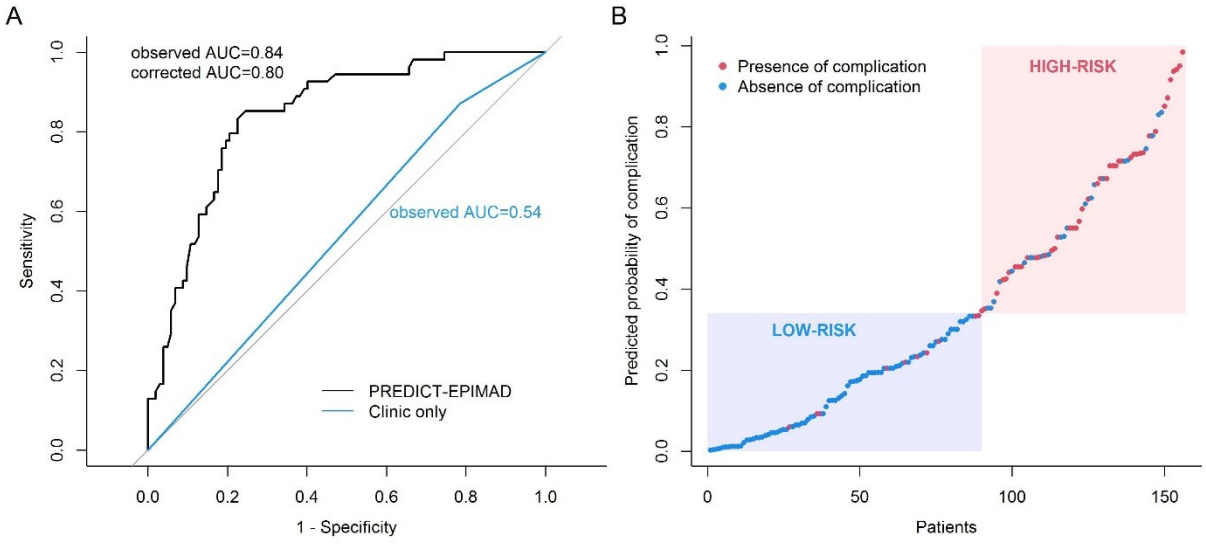


Table 23. Performances discriminantes issues de la validation interne des modèles incluant les SNP et les données cliniques et sérologiques (n = 156 patients Epimad). Les corrections pour le biais d'optimisme ont été effectuées à l'aide de 1 000 échantillons bootstrap. Le meilleur modèle prédictif (PREDICT-EPIMAD) est surligné en gris.

Critère	Variables	AUC (cohorte d'apprentissage)	Performances corrigées du biais d'optimisme (validation interne)				
			AUC (IC 95%)	Sensibilité (IC 95%)	Spécificité (IC 95%)	VPP (IC 95%)	VPN (IC 95%)
Critère principal: resection intestinale et/ou complication du phenotype (B2/B3) dans les 5 ans	6 SNPs	0,81 [0,74; 0,88]	0,77 [0,70; 0,83]	0,79 [0,58; 0,93]	0,67 [0,48; 0,86]	0,56 [0,42; 0,73]	0,86 [0,76; 0,95]
	6 SNPs, localisation iléale	0,82 [0,75; 0,89]	0,78 [0,71; 0,84]	0,81 [0,68; 0,91]	0,68 [0,55; 0,79]	0,57 [0,45; 0,68]	0,87 [0,80; 0,94]
	6 SNPs, localisation iléale, pANCA	0,84 [0,77; 0,90]	0,80 [0,73; 0,85]	0,79 [0,67; 0,88]	0,74 [0,62; 0,83]	0,61 [0,48; 0,73]	0,87 [0,80; 0,93]
	6 SNPs, localisation iléale, pANCA, ASCA IgG	0,84 [0,78; 0,90]	0,79 [0,73; 0,85]	0,77 [0,66; 0,90]	0,74 [0,56; 0,84]	0,62 [0,47; 0,74]	0,86 [0,79; 0,94]
	6 SNPs, localisation iléale, pANCA, ASCA IgG, anti-OmpC	0,84 [0,78; 0,90]	0,79 [0,72; 0,85]	0,79 [0,68; 0,90]	0,71 [0,56; 0,81]	0,59 [0,47; 0,70]	0,87 [0,80; 0,94]
Résection intestinale à 5 ans	4 SNPs	0,80 [0,71; 0,87]	0,76 [0,67; 0,83]	0,71 [0,52; 0,92]	0,71 [0,44; 0,86]	0,42 [0,26; 0,60]	0,90 [0,83; 0,96]
	4 SNPs, pANCA	0,81 [0,73; 0,88]	0,76 [0,69; 0,83]	0,80 [0,54; 0,96]	0,62 [0,42; 0,86]	0,37 [0,22; 0,61]	0,92 [0,85; 0,98]
	4 SNPs, pANCA, ASCA IgG	0,81 [0,74; 0,88]	0,77 [0,70; 0,83]	0,80 [0,55; 0,95]	0,62 [0,40; 0,86]	0,38 [0,23; 0,60]	0,92 [0,85; 0,98]
Complication (B2/B3) à 5 ans	5 SNPs	0,77 [0,68; 0,84]	0,72 [0,64; 0,80]	0,75 [0,53; 0,90]	0,62 [0,44; 0,81]	0,45 [0,32; 0,61]	0,86 [0,78; 0,93]
	5 SNPs, pANCA	0,80 [0,72; 0,88]	0,76 [0,68; 0,83]	0,72 [0,56; 0,87]	0,72 [0,56; 0,84]	0,52 [0,37; 0,67]	0,86 [0,79; 0,93]
	5 SNPs, pANCA, anti-OmpC	0,80 [0,72; 0,88]	0,75 [0,67; 0,82]	0,71 [0,56; 0,84]	0,71 [0,55; 0,82]	0,51 [0,37; 0,65]	0,85 [0,78; 0,91]

AUC : Aire sous la courbe ROC
IC : intervalle de confiance
VPP : valeur prédictive positive
VPN : valeur prédictive négative

Figure 45. Statut réel et classement des patients issu de la prédiction selon le modèle A) Génétique seule B) génétique+localisation+pANCA (PREDICT-EPIMAD). VP : vrais positifs, FP : faux positifs, VN : Vrais négatifs, FN : Faux négatifs.

A

Génétique seule		Statut réel	
		Complication	Absence de complication
Prédiction	Complication	VP **43**	FP 30
	Absence de complication	11 FN	**72** VN

B

Génétique + Localisation + pAnca		Statut réel	
		Complication	Absence de complication
Prédiction	Complication	VP **45**	FP 23
	Absence de complication	9 FN	**79** VN

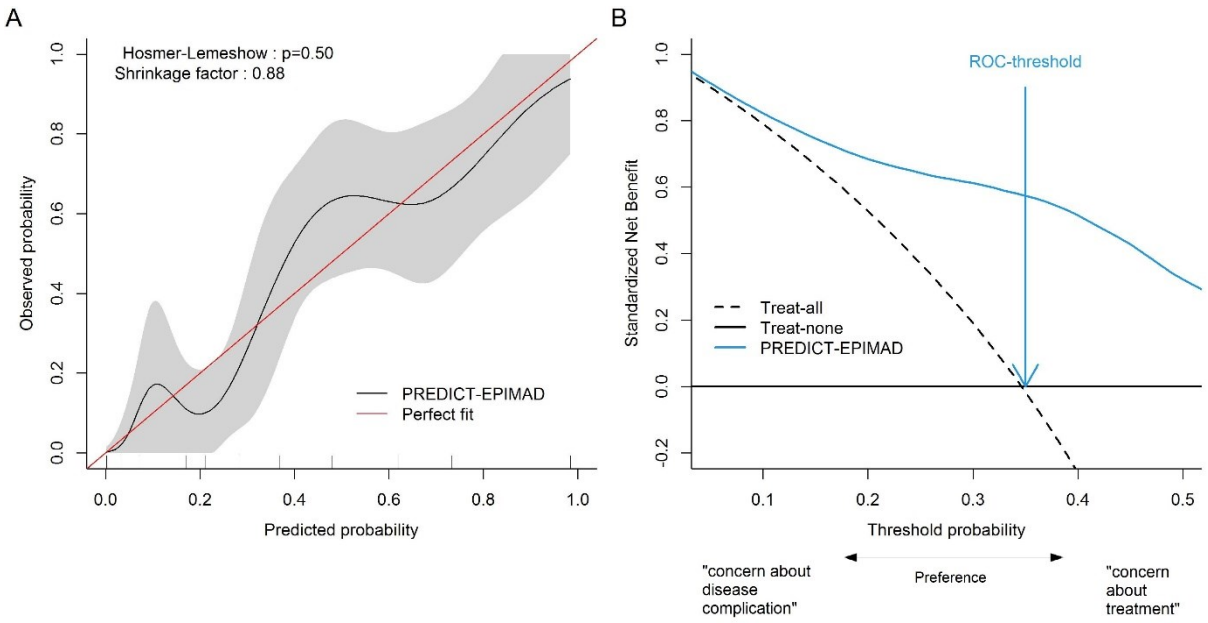
La représentation graphique des probabilités prédites selon la classification dans les groupes "faible risque" et "haut risque" à l'aide de PREDICT-EPIMAD par rapport à la présence ou à l'absence de complications observées illustre la bonne précision de la classification (Figure 44, panel B).

Le graphique de calibration montrait une bonne calibration (Figure 46, Panel A). Le test de Hosmer-Lemeshow ne montrait aucun écart significatif par rapport à un bon ajustement (p = 0,50). Le facteur de réduction du modèle calculé par bootstrap était de 0,88.

Les modèles pour les deux autres critères montraient également une bonne discrimination et une bonne calibration avec une AUC corrigée du biais d'optimisme de 0,76 à la fois pour la résection intestinale et le phénotype compliqué (Table 23). Le test de Hosmer-Lemeshow confirmait également une bonne calibration, avec des p-values de 0,70 et 0,93 pour la résection intestinale et le phénotype compliqué, respectivement.

L'analyse de la courbe de décision a confirmé l'utilité clinique de PREDICT-EPIMAD pour des probabilités seuils comprises entre 0,05 et 0,5, le bénéfice net de PREDICT-EPIMAD étant supérieur au bénéfice des stratégies "traiter tout le monde " et "ne traiter personne " (Figure 45, Panel B).

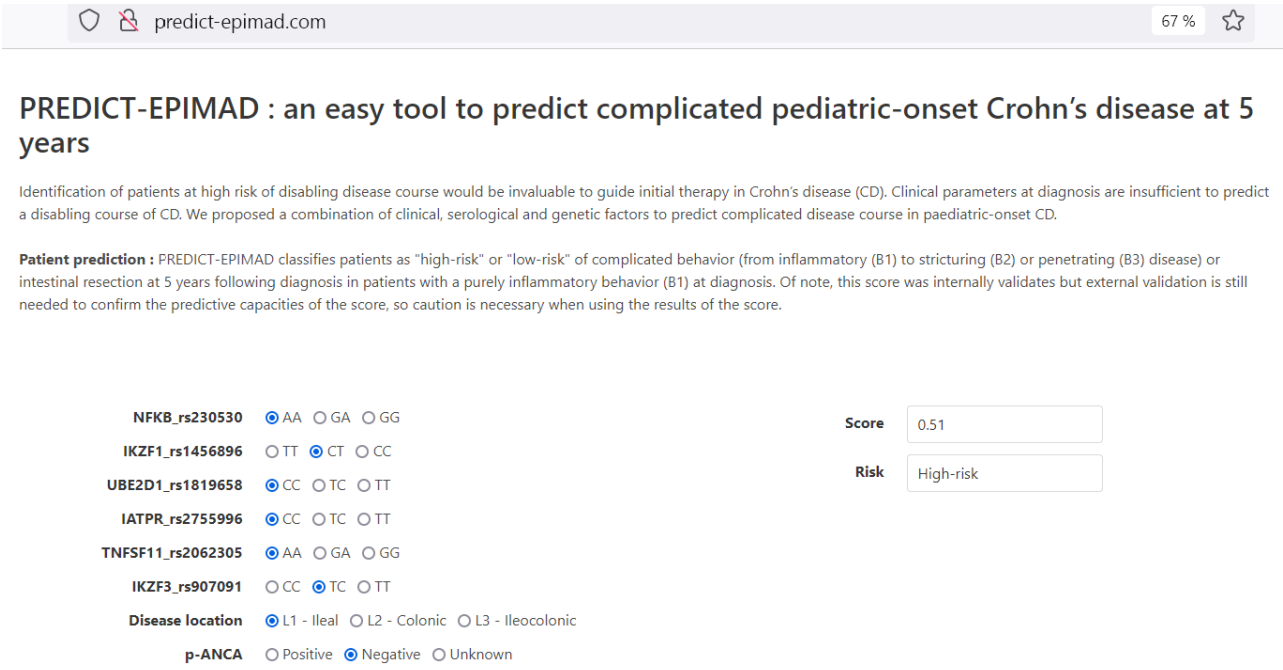
Figure 46. Calibration et utilité clinique de la validation interne de la signature basée sur le 6-SNP prédisant le résultat composite (PREDICT-EPIMAD). B) Graphique de calibration montrant les probabilités observées par rapport aux probabilités prédites : la ligne noire représente la courbe de calibration, avec l'intervalle de confiance en gris, et la ligne rouge la courbe d'un modèle parfaitement calibré. B) Courbe de décision montrant le bénéfice net standardisé en fonction des probabilités seuils. Le bénéfice net est calculé comme suit : (proportion de vrais positifs) - (proportion de faux positifs) * pt/(1-pt), où pt est la probabilité seuil. L'axe horizontal peut être interprété comme l'axe des "préférences" : plus la probabilité seuil est à gauche, plus la "préoccupation concernant les complications de la maladie" est grande ; plus elle est à droite, plus la "préoccupation concernant le traitement" est grande. L'interprétation de la courbe de décision est que la stratégie présentant le bénéfice le plus élevé sur l'ensemble des probabilités a la valeur clinique la plus élevée.



3.6. Utilisation

Le score est mis à disposition gratuitement, l'équation est donnée en matériel supplémentaire de l'article et un site web permet de le calculer facilement (Figure 47).

Figure 47. Exemple de calcul du score PREDICT-EPIMAD à partir du site web predict-epimad.com.



4. Discussion

RESUME DES PRINCIPAUX RESULTATS

A partir d'une cohorte de patients atteints de MC à début pédiatrique issue du registre Epimad, nous avons construit un modèle basé sur la combinaison de facteurs cliniques, sérologiques et génétiques afin de prédire le risque individuel à cinq ans de résection intestinale et/ou de complications sténosantes ou pénétrantes. Le meilleur modèle prédictif (PREDICT-EPIMAD) incluait la localisation au moment du diagnostic (la localisation iléale augmentant le risque), la positivité des pANCA (facteur protecteur) et 6 variants génétiques. La validation interne a montré une bonne discrimination et une bonne calibration du modèle ainsi que son utilité clinique, confirmée la courbe de décision comparant PREDICT-EPIMAD aux stratégies consistant à ne traiter personne d'une part et à traiter l'ensemble des patients d'autre part.

COMPARAISON AUX DONNEES DE LA LITTERATURE

FACTEURS CLINIQUES :

Les premières études sur la prédiction du risque d'évolution plus agressive de la maladie se sont focalisées sur des variables cliniques (205,258,260). Cependant, les paramètres cliniques au moment du diagnostic se sont avérés insuffisants pour prédire une évolution invalidante, en particulier dans le cas de la MC pédiatrique (258). Cela a été confirmé dans notre étude, dans laquelle seule la localisation iléale de la maladie a été retenue dans le modèle et ne permettait pas, à elle seule, de prédire l'évolution de la maladie. Les atteintes ano-périnéales au moment du diagnostic n'étaient pas été associées au pronostic de la maladie dans notre étude. Ceci est possiblement lié à un manque de puissance statistique puisque notre population est relativement faible avec seulement 18 patients présentant des lésions ano-périnéales au moment du diagnostic. Cependant, la présence de lésions ano-périnéales n'était pas associée au risque de complication dans la cohorte RISK (218).

SEROLOGIE ET RISQUE DE CHIRURGIE :

Il a également été démontré que la présence et l'ampleur des réponses immunitaires aux antigènes microbiens étaient associés à l'évolution de la maladie chez les enfants atteints de MC (214,259–261,269).

Une méta-analyse récente a retenu une association significative entre la positivité des ASCA et la survenue d'une intervention chirurgicale (270). Dans l'étude de Siegel et al., la quantité de pANCA était significativement et inversement associée à la complication de la maladie (définie par un phénotype compliqué B2/B3 ou une résection intestinale) dans le modèle multivarié, mais l'association à la résection intestinale seule n'était pas étudiée (214). Les pANCA n'étaient pas étudiés dans l'étude de Gupta et al (260). Dans la méta-analyse de Ricciuto et al., ainsi que l'étude de Amre et al, les pANCA n'étaient pas retenus comme associés à la survenue d'une résection intestinale (261,270).

SEROLOGIE ET COMPLICATIONS STENOSANTES ET/OU FISTULISANTES :

La littérature sur la sérologie antimicrobienne et l'évolution vers une MC compliquée sténosante ou fistulisante chez les enfants est plus difficile à interpréter. Plusieurs études ont rapporté une association significative entre la positivité des ASCA et le risque de complications pénétrantes ou sténosantes (214,261). Plus récemment, l'étude RISK a identifié un risque accru de complications B3 en cas de positivité des ASCA-IgA (218). Enfin, un récent consensus

international a établi que la positivité des ASCA prédisait l'évolution vers des complications pénétrantes et pouvait prédire des complications sténosantes, mais n'a pas retenu le statut des pANCA comme prédicteur d'un phénotype compliqué et/ou d'une intervention chirurgicale dans la MC pédiatrique en raison de données contradictoires dans la littérature actuelle (270).

Dans notre étude, nous avons observé une association significative entre la positivité des pANCA et le critère composite (OR : 0,18 [0,05 ; 0,63], p = 0,007), ainsi qu'entre la positivité des pANCA et la résection intestinale ou le phénotype compliqué, analysés individuellement (OR : 0,22 [0,05 ; 0,98], p = 0,048 et OR = 0,24 [0,07 ; 0,83], p = 0,024, respectivement). Quatre-vingt-dix pour cent des patients ayant une positivité pANCA n'ont pas eu de complication de leur maladie (résection intestinale ou phénotype compliqué). La positivité des pANCA était ainsi un bon prédicteur de l'absence de sévérité au cours de l'évolution de la maladie et augmentait ainsi la spécificité du score proposé. La positivité des ASCA-IgG n'était pas statistiquement significative (OR : 2.50 [0.80 ; 7.85], p = 0.116 pour le critère composite). Il est à noter que les variables sérologiques n'étaient pas associées dans notre étude à la localisation de la maladie. En effet, respectivement 6 %, 24 % et 18 % des patients étaient positifs aux pANCA chez les patients L1, L2 et L3 (p = 0,34). Pour les ASCA-IgG, respectivement 82 %, 79 % et 89 % des patients de L1, L2 et L3 étaient positifs (p=0,31).

GENETIQUE :

NOD2/CARD15 est la variable génétique la plus étudiée chez les enfants. Dans une récente méta-analyse, la présence de NOD2/CARD15 était associée à un risque accru de chirurgie et de complication B2 (270). Cette dernière association n'est toutefois pas claire, car la plupart de ces études ne tenaient pas compte de la localisation de la maladie et présentaient une hétérogénéité substantielle. D'autres études se sont concentrées sur d'autres variants de susceptibilité. Dans l'étude de Jakobsen et al, aucun des 41 variants de susceptibilité étudiés n'était associé à la chirurgie (262). Cependant, il est important de mentionner que les variants associés à la progression de la maladie peuvent être différents des variants de susceptibilité (219). En particulier, les scores de risque génétique combinant tous les loci de susceptibilité ne se sont pas révélés prédictifs du pronostic de la maladie, confirmant la nécessité d'intégrer des variants distincts des variants de susceptibilité dans les modèles (218), même si certains

variants de susceptibilité ont été associés à la progression de la maladie dans la littérature comme par exemple IL12B associé au risque de chirurgie (220).

Dans notre étude, le variant NOD2 G908R était associé aux trois critères en analyse univariée (données non présentées) mais NOD2 a été retenu uniquement dans le modèle multivarié pour la résection intestinale. Enfin, notre modèle incluait des variants considérés comme des gènes de susceptibilité à la MC mais aussi et surtout des variants qui pourraient être associés à une inflammation plus sévère ou réfractaire (par exemple, NFKB1, TNFSF11, TLR5, ou PRDM1).

Plus récemment, Biasci et al. ont proposé une signature transcriptomique des cellules des lymphocytes CD8 permettant d'identifier les patients adultes présentant une évolution sévère de la maladie (PredictSURE IBD) (224). Ils ont identifié un sous-groupe de patients présentant une maladie significativement plus agressive, avec un besoin plus précoce de traitement et plus d'escalades du traitement au fil du temps. Cependant, ces résultats n'étaient pas reproductibles dans une cohorte de patients pédiatriques atteints de MICI (271).

Enfin, les prédictions les plus solides proviendront probablement de modèles qui intègrent plusieurs variables, plutôt qu'un seul facteur. Nos hypothèses étaient les suivantes : i) la génétique joue un rôle plus important dans les maladies à début pédiatrique, tandis que l'exposome peut avoir un impact plus important chez les patients plus âgés, et ii) des variants différents des variants de susceptibilité pourraient prédire le pronostic.

COMPARAISON AUX MODELES EXISTANTS

Trois études se sont concentrées spécifiquement sur les cas pédiatriques, en incluant plusieurs types de variables et en fournissant les performances prédictives des modèles. Premièrement, PROSPECT est un outil permettant de prédire et de stratifier le risque pour les patients adultes et pédiatriques (214). Ce modèle permet d'identifier les patients dont la maladie se complique et inclut la localisation de la maladie, les marqueurs sérologiques (ASCA, CBir1, et ANCA), et la mutation rs2066847 de NOD2. Deuxièmement, la cohorte pédiatrique RISK a proposé un modèle incluant l'âge au diagnostic, l'origine ethnique, la localisation iléale, les anticorps antimicrobiens (ASCA-IgA et CBir1) ainsi qu'une signature génétique de la matrice extracellulaire iléale permettant de prédire le risque de maladie compliquée (B2 ou B3 analysés séparément)(218). Contrairement à notre étude, ces études n'ont pas analysé de

variants autres que les variants de susceptibilité (seulement NOD2 dans PROSPECT, score de RISK génétique dans RISK). Enfin, une nouvelle étude basée sur des patients de l'étude RISK a identifié une signature de protéines dans le sang permettant de prédire les complications pénétrantes et sténosantes (222).

Les modèles de ces études démontraient des performances prédictives considérées comme modérées ou bonnes, avec un c-index de Harrell de 0,75 pour les enfants dans PROSPECT, une AUC de 0,72 pour le meilleur modèle prédictif de RISK, et une AUC de 0,69 et 0,79 pour B2 et B3, respectivement, dans l'étude d'Ungaro et al, contre 0,80 après validation interne pour PREDICT-EPIMAD. Dans notre étude, la sensibilité était de 0,79 et la spécificité de 0,74 après correction pour le biais d'optimisme contre 0,69 et 0,71, respectivement, dans RISK (données non fournies dans les autres études).

Dans la littérature récente, l'accent a été mis sur l'importance d'obtenir une valeur prédictive négative (VPN) élevée (218,224), c'est-à-dire la probabilité d'absence de complication chez les patients classés comme "à faible risque" et donc non traités. PREDICT-EPIMAD atteignait une VPN élevée de 0,87, contre 0,94 dans l'étude RISK. Il est à noter que la VPN dépend de la proportion d'événements, qui était plus élevée dans notre population (35 % des patients présentant une résection intestinale et/ou un phénotype compliqué à cinq ans) que dans celle de l'étude RISK (9 % présentant un phénotype compliqué après 3 ans de suivi), dans laquelle la prévalence plus faible des complications dans la cohorte peut également, à l'inverse, expliquer, en partie, la VPP plus faible du modèle. En effet, la VPP de notre modèle était de 0,61 contre 0,24 pour le score de prédiction RISK. Dans notre population une stratégie "ne traiter aucun patient" atteindrait une VPN de 0,65 et une stratégie "traiter tous les patients" une VPP de 0,35. Le modèle PREDICT-EPIMAD constitue donc un bon compromis entre les stratégies "Traiter tous les patients" et "Ne traiter aucun patient" en termes de VPN et de VPP. Ceci a été confirmé par l'analyse de la courbe de décision, évaluant l'utilité clinique de PREDICT-EPIMAD.

FORCES ET FAIBLESSES

Le principal atout de notre étude est qu'elle était basée sur un registre en population, ce qui permettait d'éviter un biais de sélection.

Un autre point fort est l'utilisation de méthodes de régression pénalisées qui ont permis d'inclure simultanément un grand nombre de marqueurs génétiques et de produire un modèle parcimonieux qui ne comprend au final qu'un petit nombre de variables afin d'éliminer les variables de bruit et de proposer un outil facile à utiliser. En outre, seuls les variants stables identifiés à l'aide de rééchantillonnages par bootstrap ont été sélectionnés dans la cohorte d'apprentissage. Ces variants ont ensuite été confirmés à l'aide d'une cohorte externe, ce qui a permis de réduire considérablement le risque de faux positifs parmi les variants sélectionnés et de limiter le risque de sur-ajustement. Une validation interne rigoureuse a ensuite été effectuée à l'aide d'échantillons bootstrap afin d'effectuer des corrections pour le biais d'optimisme. Enfin, le nombre final de variables retenu est raisonnable au regard du nombre d'évènements observés.

Un avantage de notre score est que les sérotypes et les génotypes sont stables dans le temps et que ces tests sont faciles à réaliser. Les sérologies ont été effectuées au moment de l'inclusion. Cependant, nous avons considéré que les anticorps étaient stables dans le temps, ce qui semble être une hypothèse raisonnable et bien acceptée (272,273). En ce qui concerne le traitement, les patients étaient peu exposés aux anti-TNF puisqu'ils ont été diagnostiqués entre 1988 et 2004. En effet, 60 % (n=93) et 19 % (n=29) des patients ont été exposés à un traitement immunosuppresseur et à un anti-TNF au cours des 5 années suivant le diagnostic de la MC, respectivement. En particulier, seul un patient a été exposé précocement à un anti-TNF tel que défini dans l'étude RISK, c'est-à-dire dans les 3 mois suivant le diagnostic. Ceci présente un avantage pour le développement d'un score prédictif car les résultats sont susceptibles d'être moins impactés par l'effet du traitement sur l'évolution de la maladie que dans les études récentes. Nous avons choisi d'évaluer le risque de complications à cinq ans, car ce risque survient principalement au cours des premières années suivant le diagnostic (76). Ce seuil de 5 ans est également un horizon raisonnable dans la décision de traiter dans les premiers mois qui suivent le diagnostic.

Notre étude présente également plusieurs limites. La première concerne la taille de l'échantillon de la cohorte de découverte utilisée pour développer le score. Le manque de puissance statistique pourrait expliquer l'absence d'association entre certaines variables cliniques comme par exemple les lésions ano-périnéales et l'évolution de la maladie. Cependant, les études de cohortes de la MC pédiatrique sont rares et seules quelques études

se sont concentrées sur cette population spécifique. Ce manque d'études est notamment pointé dans la discussion de la revue de la littérature menée par Ricciuto et al (270). Deuxièmement, les données sur le statut tabagique des patients n'étaient pas disponibles. L'exposition au tabagisme est cependant difficile à évaluer dans une population pédiatrique. Enfin, nous avons considéré que le meilleur modèle prédictif était celui qui atteignait l'AUC la plus élevée avec le plus petit nombre de variables, mais on peut noter que les intervalles de confiance des AUC et des autres performances prédictives se chevauchaient entre les différents modèles testés.

La limite la plus importante est qu'une validation externe est nécessaire. Nous n'avons pas utilisé la cohorte externe comme cohorte de validation car : i) la taille de l'échantillon était faible, ii) les données sérologiques n'étaient pas disponibles, et iii) les patients provenaient uniquement de centres experts. La validation dans une cohorte externe est rendue difficile par le fait que les cohortes pédiatriques issues de la population générale sont rares et surtout par la fait que les cohortes récentes sont fortement exposées aux anti-TNFs, ce qui modifie l'histoire naturelle de la maladie. L'utilité clinique pourrait ultérieurement être évaluée dans un essai randomisé stratifié sur le résultat du score. Une telle étude est cependant compliquée à mettre en pratique (274).

Conclusions de la seconde partie

Dans cette deuxième partie de la thèse, nous avons tout d'abord comparé, dans le premier chapitre, par simulations de données, plusieurs méthodes pour analyser conjointement des données cliniques et omiques dans l'objectif d'analyser les données du registre EPIMAD. Nous avons comparé plusieurs méthodes, en une ou deux étapes, basées sur une sélection parcimonieuse de variables par des méthodes étendant le Lasso. Ces travaux m'ont permis d'acquérir de nouvelles méthodes statistiques et d'esquisser une réflexion sur le choix optimal de la méthode selon les données à analyser.

Dans le second chapitre, PREDICT-EPIMAD est un score qui combine des facteurs cliniques, sérologiques et génétiques pour prédire l'évolution de la MC inflammatoire pédiatrique vers une évolution compliquée de la maladie. Le score proposé est facile à déterminer, car il est basé sur 6 SNP, les données cliniques et sérologiques habituelles. Un outil web permettant de classer les patients en fonction de leur risque de complications (élevé ou faible) est disponible gratuitement à l'adresse www.predict-epimad.com. Il s'agit d'une première étape pour répondre à un besoin non satisfait, afin d'aider les médecins à proposer des options de traitement aux enfants et à leurs familles, de mieux les informer sur le pronostic et de justifier une intervention précoce pour modifier l'évolution naturelle de la maladie. Cependant, une validation externe est encore nécessaire pour confirmer les capacités prédictives du score. Notre objectif à moyen et long terme est d'introduire l'outil actuel dans la pratique clinique, ce qui permettrait aux médecins de proposer une thérapie intensive précoce appropriée aux patients à haut risque ou, au contraire, une approche progressive pour les patients à faible risque.

Ces résultats (PREDICT-EPIMAD) ont fait l'objet d'une publication dans la revue Inflammatory Bowel Diseases en 2023. J'ai réalisé les analyses statistiques et rédigé l'article. L'article est présenté dans les pages qui suivent (version acceptée). Le matériel supplémentaire est présenté en Annexe du mémoire.

A novel 8-predictors signature to predict complicated disease course in pediatric onset Crohn's disease: a population-based study.

Authors : Hélène Sarter, *MS* [1,2], Guillaume Savoye, *MD, PhD,* [3], Guillemette Marot, PhD, [4,5], Delphine Ley, *MD, PhD* [2,6], Dominique Turck, *MD,* [2,6], Jean-Pierre Hugot, *MD, PhD* [7,8], Francis Vasseur, *MD, PhD* [4], Alain Duhamel, *PhD* [4], Pauline Wils, *MD,* [2,9], Fred Princen, *PhD* [10], Jean-Frédéric Colombel, *MD,* [11], Corinne Gower-Rousseau, *MD, PhD* [1,2,12], Mathurin Fumery, *MD, PhD* [13], and EPIMAD study group*.

1. Lille Hospital and University, Public Health, Epidemiology and Economic Health, Epimad registry, Regional house of clinical research, F-59000 Lille, France

2. Univ. Lille, Inserm, CHU Lille, U1286 - INFINITE - Institute for Translational Research in Inflammation, F-59000 Lille, France

3. Rouen Hospital and University, Gastroenterology Unit, Epimad registry, Rouen, France

4. Univ. Lille, CHU Lille, ULR 2694-METRICS : Evaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France.

5. Inria Lille Nord Europe, Modal, Lille, France

6. Lille University Jeanne de Flandre Children's Hospital and Faculty of Medicine, Division of Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, Lille, France

7. Centre de Recherche sur l'Inflammation, UMR1149 INSERM et Université de Paris, France.

8. Department of Pediatric Gastroenterology, Hôpital Robert Debré, Assistance Publique Hôpitaux de Paris (AP-HP), Paris, France.

9. Gastroenterology Unit, Lille Hospital and University, Lille, France

10. Prometheus Laboratories, San Diego, California

11. Icahn School of Medicine at Mount Sinai, Division of Gastroenterology, New York, United States

12. Research and Public Health Unit, Reims University & Hospital, Robert-Debré Hospital, Reims, France

13. Amiens Hospital and University, Gastroenterology Unit, Epimad Registry, and PeriTox, UMR I-01, Amiens, France

Correspondence to Hélène Sarter, Service de Santé Publique, Epidémiologie, Economie de Santé et Prévention, Registre Epimad, Maison Régionale de la Recherche Clinique, Centre Hospitalier Universitaire Régional, CS 70001, 59037 Lille Cedex, France, helene.sarter@chru-lille.fr or Prof Mathurin Fumery, Gastroenterology Departement, Rond-point du Pr Cabrol, Centre hospitalier Universitaire d'Amiens, 80000, Amiens, France, fumery.mathurin@chu-amiens.fr.

This work was presented in part at the European Crohn's and Colitis Organization (ECCO) meeting held in Vienna in 2018, the Digestive Disease Week (DDW) meeting held in Washington, DC in 2018, the UEG Week meeting in Vienna in 2018, and the JFHOD in Paris in 2019.

Word count: 3,799

DECLARATIONS

Ethics approval and consent to participate : The study protocol was approved by the ANSM under the number 2007-A00468-45 and by the CPP Nord-Ouest IV IRB under the number 2007 /32-MS1. For the external cohort, the study received approval from the French national ethics committee CPP Ile-de-France IV (Hôpital Saint Louis, Paris, France). All participants signed an informed consent form.

Consent for publication : Not applicable.

Availability of data and materials : Data are available on reasonable request at the corresponding author.

Competing interests :

GS has served as speaker for MSD France, Ferring France, Abbvie France, and Vifor France.

JPH received congress fees from MSD and Biogen.

JFC has received research grants from AbbVie, Janssen Pharmaceuticals and Takeda; has received payment for lectures from AbbVie, Amgen, Allergan, Bristol-Myers Squibb Company, Ferring Pharmaceuticals, Shire, Takeda and Tillots; has received consulting fees from AbbVie, Amgen, Arena Pharmaceuticals, Boehringer Ingelheim, Bristol-Myers Squibb Company, Celgene Corporation, Celltrion, Eli Lilly, Enterome, Ferring Pharmaceuticals, Genentech, Gilead, Iterative Scopes, Ipsen, Immunic, lmtbio, Inotrem, Janssen Pharmaceuticals, Landos, LimmaTech Biologics AG, Medimmune, Merck, Novartis, O Mass, Otsuka, Pfizer, Shire, Takeda, Tigenix, Viela bio; and hold stock options in Intestinal Biotech Development.

CGR has served as a speaker for Ferring France & International, Takeda France, Tillotts France, Janssen International, and MSD France.

MF received lecture fees or consultancy fees from MSD, Abbvie, Takeda, Ferring, Gilead, Celgene, Celltrion, Biogen, Janssen, Hospira, and Boehringer.

HS, GM, DL, DT, AD, FV, PW and FP have nothing to declare.

Funding: EPIMAD is organized under an agreement between the Institut National de la Santé et de la Recherche Médicale (INSERM) and Santé Publique France and also received financial support from the François Aupetit Association, Lille, Amiens, and Rouen University Hospitals. This work was supported by the Programme Hospitalier de Recherche Clinique Inter-regional (number PHRC 22-3IR). Serological analyses were financed by Prometheus Laboratories under a scientific collaborative

agreement (number COL11IBD02) between Lille University Hospital, Amiens University Hospital and Prometheus Laboratories. We would like to thank the DigestScience Foundation and Association Robert Debré pour la Recherche Médicale.
Role of the funding source : Prometheus Laboratories financed the serological analyses and reviewed the final manuscript before submission, but the academic authors retained editorial control. All other funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Authors' contributions :

Concept and Study design: GS, DT, FV, JFC, CGR, and MF
Data acquisition: GS, JPH, FV, CGR, PW, FP, and MF
Data management: HS, FV
Statistical analysis: HS, GM, and AD
Interpretation of the data: HS, GS, GM, DL, JPH, FV, PW, CGR, and MF
Drafting of the manuscript: HS, MF
Critical revision of the manuscript: GS, GM, DL, DT, JPH, FV, AD, PW, FP, JFC, and CGR
HS, CGR and MF verified the underlying data.

Acknowledgements :

The authors wish to thank the interviewing practitioners who collected data: N. Guillon, S. Auzou, B. David, H. Pennel, A. Pétillon. The authors thank all patients and all gastroenterologists who participated in this study, the European Charity Fondation DigestScience and Takeda laboratories.

Abbreviations: AUC, area under the curve; IBD, inflammatory bowel disease; CD, Crohn's disease; CI : confidence interval; Lasso, least absolute shrinkage and selection operator; IQR, inter-quartile range; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic; SNP, single nucleotide polymorphism.

Brief summary: The identification of patients at high-risk of a disabling disease course would be invaluable in guiding initial therapy in Crohn's disease. We constructed a score that combines clinical, serological, and genetic factors able to predict the evolution of pediatric-onset inflammatory Crohn's disease to a complicated disease course.

What is already known?

It is crucial to personalize therapeutic strategies in pediatric-onset Crohn's disease (CD) and identify patients requiring early intensive therapy at disease onset.

What is new here?

PREDICT-EPIMAD is a score that combines clinical, serological, and genetic factors to predict the evolution of pediatric-onset inflammatory CD to a complicated disease course.

How can this study help patient care?

The proposed score is easy to determine, as it is based on 6 SNPs and the usual clinical and serological data. It is a direct response to the unmet need, to aid physicians in providing treatment options for children, better inform them of the prognosis, and justify early intervention provided these results can be validated by an independent cohort with similar testing conditions.

Abstract

Background: The identification of patients at high risk of a disabling disease course would be invaluable in guiding initial therapy in Crohn's disease (CD). Our objective was to evaluate a combination of clinical, serological, and genetic factors to predict complicated disease course in pediatric-onset CD.

Methods: Data for pediatric-onset CD patients, diagnosed before 17 years of age between 1988 and 2004 and followed more than five years, were extracted from the population-based Epimad registry. The main outcome was defined by the occurrence of complicated behavior (stricturing or penetrating) and/or intestinal resection within the five years following diagnosis. Lasso logistic regression models were used to build a predictive model based on clinical data at diagnosis, serological data (ASCA, pANCA, anti-OmpC, anti-Cbir1, anti-Fla2, anti-Flax), and 369 candidate single nucleotide polymorphisms (SNPs).

Results: In total, 156 children with an inflammatory (B1) disease at diagnosis were included. Among them, 35% (n=54) progressed to a complicated behavior or an intestinal resection within the five years following diagnosis. The best predictive model (PREDICT-EPIMAD) included the location at diagnosis, pANCA, and 6 SNPs. This model showed good discrimination and good calibration, with an AUC of 0. 80 after correction for optimism bias (sensitivity:79%, specificity:74%, positive predictive value:61%, negative predictive value:87%). Decision curve analysis confirmed the clinical utility of the model.

Conclusions: If these results can be validated in an independent cohort, a combination of clinical, serotypic, and genotypic variables can predict disease progression with high accuracy in this population-based pediatric-onset CD cohort and represents a step towards personalized therapy in children.

Key words: inflammatory bowel disease; Crohn's disease; prognosis; complication; genetics; prediction.

Number of abstract words: 243.

BACKGROUND

Crohn's disease (CD) is a complex chronic, relapsing, destructive inflammatory disorder of the gastrointestinal tract that can lead to intestinal damage and impair the quality of life.[1] Crohn's disease is a heterogeneous disorder, with differences in disease presentation at diagnosis and an unpredictable disease course.[2] Even though incidence of CD is overall plateauing in westernized countries, incidence is still increasing in pediatric populations .[3,4] Pediatric CD has been described as a more severe disease than adult onset CD, with up to one third of children showing complicated disease (stricturing or penetrating) within the five years following diagnosis and undergoing early intestinal resection.[5,6] The more severe pattern of pediatric CD may result in the retardation of growth and puberty and a high level of disability. New therapeutic strategies based on the early intensive use of immunosuppressants and/or anti-TNF have emerged over the last few decades to achieve corticosteroid-free remission and prevent disease progression and irreversible bowel damage.[7,8] Such therapies are however limited by an increased risk of infection, paradoxical manifestations, or cancer.[9–11]

As only certain patients progress to a complicated disease course, it is crucial to personalize therapeutic strategies and identify patients at high-risk of complications requiring early intensive therapy at disease onset. Most existing prediction models have a number of limitations. Notably, they were developed from single centers or referral centers and mostly in the adult population or are based on only a limited number of clinical, serological, and genetic variables. Studies specifically focused on pediatric-onset CD, for which the impact of genetics may be different, are rare.[12–19] Recently, a study showed poor replication of existing studies, including RISK model.[17,20]

We therefore aimed to develop a score based on clinical, serological, and genetic data to predict a complicated disease course in pediatric-onset CD at diagnosis using a well-defined population-based cohort.

METHODS

Study population (Discovery cohort)

The data of patients with a diagnosis of CD made before 17 years of age between January 1988 and December 2004 were extracted from the Epimad registry. Only patients with a non-penetrating, non-stricturing disease inflammatory behavior (B1) at diagnosis and with at least five years of follow-up were considered. The methodology of the Epimad registry has been previously described in detail.[21] Briefly, each gastroenterologist of Northern France practicing in the private or public sector (n = 265) reported all patients first consulting for symptoms compatible with IBD. The final diagnosis of IBD was made by two GE experts according to previously published validated criteria.[21]

Phenotypic data collection

Data were extracted from medical records, collected by interviewer practitioners in standardized questionnaires and reviewed for accuracy and completeness by the responsible investigator (CGR). CD behavior was defined according to the Montreal classification: B1, inflammatory (non-stricturing non-penetrating); B2, stricturing; and B3, penetrating disease. B2 and B3 were pooled and defined as 'complicated behavior'. Surgery was restricted to intestinal resection.

Outcomes

The main outcome was defined by the composite criteria: complicated behavior (B2 or B3) and/or intestinal resection within the five years following diagnosis. Secondary outcomes were i) need for intestinal resection within the five years following diagnosis and ii) complicated behavior at five years (B2 or B3), analyzed separately.

Serological data

Blood samples for sera and plasma were collected at inclusion. Serological determination for anti-*Saccharomyces-cerevisiae* (ASCA) IgA and IgG, perinuclear antineutrophilic cytoplasmic antibody (pANCA), anti-outer membrane protein C precursor (anti-OmpC), and anti-flagellins (anti-CBir1, anti-Fla1, and anti-FlaX) was performed by Prometheus Laboratories (San Diego, CA).

DNA genotyping

The full list of SNPs is provided in Supplementary Table S1. Selection of SNPs was performed as follows. First, pathways of interest for the present study were identified by discussion between geneticist (FV) and expert gastroenterologists. Then, SNPs were identified from a systematic review of the literature carried out by the geneticist (FV) at the time of the design of the study. Second, we conducted a comprehensive search of multiple electronic databases with no language restrictions. One reviewer (FV) independently assessed the title and abstract of studies identified in the primary search for inclusion, and the full text of remaining articles were examined to determine whether they met inclusion criteria. In total, 373 SNPs were selected for genotyping on the basis of a systematic literature review including major CD genetic risk factors (as the *NOD2, NOD1, IL23R, ATG16L1, DGL5,* and *IL10R1* genes), variants in genes involved in inflammation pathways (such as *TNFα, TNFRSF14, TNFRSF9, IL6,* and *NFKB1*), genetic variants disclosed by GWAS in both CD and UC (as the *PER3, WNT4, ITLN1, DNMT3A,* and *PUS10* genes), variants at the HLA-DRB1 and HLA-DQA1 loci extensively reported to be involved in IBD susceptibility, variants in genes encoding factors involved in interactions with microorganisms (as the *DEFB1,* and *DEFB4* genes), variants in genes encoding proteins that modulate innate immunity (as the *TLR2, TLR4, DECTIN1,* and *CARD9* genes), and genetic variants reported to be associated with or that modulate the severity of immune-mediated diseases (as the *PDCD1, TNFSF15,* and *TNFRSF14* genes). Genotyping was performed using an Illumina Bead Express platform (Illumina, Inc., San Diego, CA, USA) by the Centre National de la Recherche Scientifique (CNRS

UMR8199, National Center for Scientific Research) according to the manufacturer's protocol. Two SNPs were genotyped using a Taqman assay (*NOD2* R702W -rs2066844 and G908R - rs2066845). Genotyping quality control was performed by checking the Hardy-Weinberg equilibrium (cutoff : 0.05) and minor allele frequencies as reported in the 1000genome database. Variants with a minor allele frequency < 1% (1 variant) or with >5% missing data (3 variants) were discarded from the analysis. Subjects with more than 5% of missing genotypes on the 369 remaining SNPs were discarded from the analysis (n = 4). A small amount of data was still missing – representing 0.3% (n=177 data) of the total amount of 57 464 data from the 369 SNPs * 156 patients table - and was imputed by sampling from a multinomial distribution with probabilities taken from the observed genotypes.

External cohort

Sixty pediatric-onset patients with non-complicated CD at diagnosis were extracted from a previously published French cohort of CD.[22] For these patients, genotyping was performed by the company Integragen (Evry, France) using a Fluidgim Biomark system and AS-PCR technology. Only SNPs selected after statistical analyses in the discovery cohort were available for this external cohort. Serological data were not available in this cohort. This sample was used to consolidate the selection of genetic variants performed from the discovery cohort (Epimad).

Statistical analysis

Quantitative variables are expressed as the median and interquartile ranges (Q1–Q3) and qualitative variables by frequency and percentage. Variants were transformed into counts of the number of minor alleles (0, 1, or 2).

To build a predictive model combining clinical, serological, and genetic data, we used a two-step analysis : in the first step, relevant variables were selected separately in each group of variables (*i.e.,* clinical, serological, and genetic) and in the second step, all selected variables were included in one final regression model.[23] The statistical strategy is depicted in figure 1.

Selection of variables

For each outcome, clinical and serological variables were selected using univariable logistic regressions at $p \leq 0.2$. Genotypes were selected using lasso logistic regression and stability selection.[24] All selected variants were then tested in the pooled sample of the discovery cohort and the external cohort to reduce the number of false positive SNPs. These analyses were performed using univariable logistic regression models adjusted for study effect: variants achieving $p \leq 0.05$ were considered for inclusion in the final model.

Predictive model building

For each outcome, the predictive model was built from the discovery cohort by including all selected clinical, serological, and genetic variables in an adaptive lasso logistic regression model (Supplementary

Methods). The following models were tested: i) genetics only, ii) genetics and clinical variables, and iii) genetics, clinical, and serological variables. For each model, a scoring system was derived from the equation of the model and the threshold for classification of patients in high- and low-risk groups of complication was determined using Youden's J statistic.

Performance of the selected model and internal validation

Discriminative performance of the models was assessed by the area under the ROC curve (AUC), sensitivity, specificity, positive predictive value, and negative predictive value, presented with the associated 95% confidence intervals. Calibration of the model was assessed by the Hosmer–Lemeshow test for goodness-of-fit, calculated using deciles and represented graphically by the calibration plot of observed versus predicted probabilities of the logistic model using natural splines with six degrees of freedom. Internal validation was done by using bootstrap resampling with 1000 repetitions to estimate the AUC and other discriminative performances corrected for optimism bias and the shrinkage factors. The shrinkage factor is the estimated shrinking in the regression coefficients to improve the prediction in future patients. Finally, clinical usefulness of the model was assessed using decision curve analysis, plotting standardized net benefit against threshold probabilities, and comparing "treat-all" and "PREDICT-EPIMAD" strategies against "treat-none" (Supplementary Methods).[25]

Data were analyzed using SAS V.9.4 (SAS, SAS Institute Inc., Cary, NC, USA) and R 3.4.3 software. More details on the statistical methodology are available in Supplementary Material.

RESULTS

Characteristics of patients from the Epimad discovery cohort at diagnosis

In total, 156 children (male 54%) with an inflammatory (B1) phenotype at diagnosis were included, with a median follow-up of 10. 4 years (Interquartile range (IQR), 7. 2 – 14. 9). The patient characteristics at diagnosis are presented in Table 1. The median age at diagnosis was 14.3 years (IQR, 11.9 – 16.0). The main disease location at diagnosis was ileo-colonic (n = 109, 70%). Sixty percent (n=93) and 19% (n=29) of patients were exposed to immunosuppressive therapy and anti-TNF during the 5 years following CD diagnosis, respectively (Table 1, Supplementary Figure S2). Patients from the external cohort are described in Supplementary Table S2.

Complicated disease course

Among the 156 children with an inflammatory (B1) disease at diagnosis, 22% (n = 35) had intestinal resection and 29% (n = 46) complicated behavior (B2, n = 32 (70%), B3, n = 14 (30 %)) within five years following diagnosis. The main outcome (*i.e.,* composite outcome: intestinal resection and/or complicated behavior within 5 years) was observed in 35% of the children (n = 54) (Table 1).

Selection of clinical and serological variables

These variables were selected at the significance level $p \leq 0.2$. Univariable analyses of clinical and serological variables are presented in Table 2. For the composite outcome, ileal or ileo-colonic location of the disease at diagnosis (Odds Ratio (OR), 95% CI (Confidence Interval) 1.87 [0.74; 4.71], p = 0.187), p-ANCA positivity (0.18 [0.05; 0.63], p = 0.007), ASCA-IgG positivity (2.50 [0.80; 7.85], p = 0·116), and anti-OmpC positivity (1.92 [0.78; 4.70], p = 0.154) were selected.
Variables selected for intestinal resection within five years were p-ANCA positivity (0.22 [0.05; 0.98], p = 0.048) and ASCA-IgG positivity (3.07 [0.68; 13.89], p = 0.145). For complicated behavior within 5 years, p-ANCA positivity (0.24 [0.07; 0.83], p = 0.024) and anti-OmpC positivity (2.07 [0.83; 5.14], p = 0.116) were selected.

Selection of genetic variants

Nine SNPs were selected for the main outcome by lasso and stability selection (Supplementary Figure S1). Among the selected variants, six (in the *NFKB1, IKZF1, UBE2D1, IATPR, TNFSF11*, and *IKZF3* genes) were confirmed using the pooled sample of the discovery and external cohorts (Table 3, Supplementary Table S3). For secondary outcomes, nine and six SNPs were selected by lasso regression for intestinal resection and complicated behavior within five years, respectively. Four SNPs (in the *IHPK1, HMGB1, NOD2*, and *ORMDL3* genes) were confirmed for intestinal resection using the pooled sample of the discovery and external cohorts and five (in the *TLR5, PRDM1, IKZF1, KLF6*, and *AKT1* genes) for complicated behavior (Table 3, Supplementary Table S3).

Predictive model construction and internal validation

All selected variables were introduced into adaptive lasso logistic regression models for each outcome. The discriminative performances of the models are presented in Table 4. Model coefficients and cut-offs to implement the score are presented in Supplementary Table S4. AUCs ranged between 0.77 and 0.84 with overlapping CIs. We considered the best predictive model as the one achieving the highest AUC with the smallest number of variables. The best predictive model (called PREDICT-EPIMAD in the following) was that predicting the main outcome and included ileal location, 6 SNPs (in the *NFKB1, IKZF1, UBE2D1, IATPR, TNFSF11*, and *IKZF3* genes), and pANCA. This model achieved an AUC of 0.84 [95% CI: 0.78; 0.90] in the discovery sample. The performances corrected for optimism bias were an AUC of 0.80 [0.73; 0.85] (Figure 2a), a sensitivity of 0.79 [0.67; 0.88], a specificity of 0.74 [0.62; 0.83], a PPV of 0.61 [0.48; 0.73], and an NPV of 0.87 [0.80; 0.93]. Introduction of pANCA in the model improved the predictive performance. Specificity increased from 0.67 [0.48; 0.86] to 0.74 [0.62; 0.83] and the PPV from 0.56 [0.42; 0.73] to 0.61 [0.48; 0.73], whereas sensitivity and the NPV remained stable. The plot of predicted probabilities according to classification in “Low-risk” and “High-risk” groups using PREDICT-EPIMAD as compared to observed presence or absence of complications shows the good accuracy of the classification (Figure 2b). Calibration plot showed a good calibration (Figure

3a). The Hosmer-Lemeshow test showed no significant departures from a good fit (p = 0.50). The shrinkage factor of the model calculated by bootstrapping was 0.88. Models for secondary outcome also showed good discrimination and calibration with an over-optimism corrected AUC of 0.76 for both intestinal resection and complicated behavior (Table 4). The Hosmer-Lemeshow test also confirmed good calibration, with p-values of 0.70 and 0.93 for intestinal resection and complicated behavior, respectively.

Decision curve analysis confirmed the clinical usefulness of PREDICT-EPIMAD for threshold probabilities between 0.05 and 0.5, the net benefit from PREDICT-EPIMAD being superior to the benefit from both "treat-all" and "treat-none" strategies (Figure 3b).

DISCUSSION

We used a pediatric-onset CD cohort issued from a population-based study to construct a model based on the combination of clinical, serological, and genetic factors to predict the individual five-year risk of intestinal resection and/or stricturing or penetrating complications. The best predictive model included the location at diagnosis, pANCA levels, and 6 SNPs to predict the composite outcome (intestinal resection and/or complicated behavior at five years). Internal validation showed good discrimination and calibration of the models as well as its clinical usefulness, confirmed by decision curve analysis. A web-tool to classify patients as high- or low-risk for complications is freely available at www.predict-epimad.com.

Patients with CD typically present with inflammatory behavior at diagnosis, which progresses to irreversible bowel damage for some, leading to surgery. Early biologic treatment is associated with improved clinical outcomes for both adult and pediatric CD patients.[8,17,26] Not all patients progress to a complicated disease course. Thus, this strategy comes with the risk of potentially over-treating certain patients, with the associated cost and exposure to medications with measurable serious side effects.[9–11] Models to predict those patients that are at the highest risk of complications are needed to incorporate a personalized approach to therapy.

Early studies on risk prediction used clinical markers to find associations with a more aggressive disease evolution. However, clinical parameters at diagnosis are insufficient to predict a disabling course, especially in pediatric CD.[12] This was confirmed in our study, in which only the ileal location of the disease was included in the model and was unable to predict the disease course alone. Of note, perianal CDs at diagnosis were not associated to disease prognosis in our study. A lack of statistical power is likely to be present since our population is relatively small with only 18 patients having perianal CD at diagnosis. The presence and magnitude of immune responses to microbial antigens have also been shown to be associated with the disease course of children with CD.[13–16,27] Contrary to ANCA, the ASCA status and surgery showed a significant association in a recent meta-analysis.[28] The literature on

antimicrobial serology and progression to complicated CD in children is more difficult to interpret. Several studies have reported a significant association between the combination of ASCA status and the risk of penetrating or stricturing complications.[15,16,27,29] More recently, the RISK study identified a clearly increased risk of B3 complications with ASCA-IgA positivity. Finally, a recent international consensus stated that ASCA positivity predicts progression to penetrating complications and may predict stricturing complications but did not retain the ANCA status as a predictor of complicated behavior and/or surgery in pediatric CD because of conflicting data in the current literature.[27] Here, we observed a significant association between pANCA positivity and composite outcome (OR: 0.18 [0.05; 0.63], p = 0.007), as well as between pANCA positivity and intestinal resection or complicated behavior, analyzed individually (OR: 0.22 [0.05; 0.98], p = 0.048 and OR = 0.24 [0.07; 0.83], p = 0.024, respectively). Ninety percent of patient having pANCA positivity did not have a complication of their disease (intestinal resection or complicated behavior) so pANCA positivity is a good predictor of absence of severity during disease course and thus increased the specificity of the proposed score. ASCA-IgG positivity did not reach statistical significance (OR: 2.50 [0.80; 7.85], p = 0.116 for composite outcome). Of note, serologic variables were not associated to disease location. Respectively 6%, 24% and 18% of patients were pANCA positive in L1, L2 and L3 patients respectively (p=0.34). For ASCA-IgG, respectively 82%, 79% and 89% of L1, L2 and L3 patients were positive (p=0.31).

The *NOD2/CARD15* variant is the most studied genetic variable in children. In a recent meta-analysis, the presence of *NOD2/CARD15* was associated with an increased risk of surgery and B2 complication.[28] This last association is still unclear, however, as most of these studies did not adjust for disease location and showed substantial heterogeneity. Other studies have focused on other susceptibility variants in children. In Jakobsen et al,[19] none of the 41 studied susceptibility variants were associated with surgery and genetic scores combining all known susceptibility loci did not have a predictive value in the RISK study,[17] confirming the need to integrate variants distinct from susceptibility variants in genetic analyses. In our study, *NOD2* G908R was associated with the three outcomes in univariable analysis (data not shown) but *NOD2* was retained only in the multivariable model for intestinal resection. Finally, our model includes SNPs considered as susceptibility gene for CD but also and mostly SNPs which could be associated with more severe or refractory inflammation (e.g., NFKB1, TNFSF11, TLR5, or PRDM1). More recently, Biasci et al. used transcriptional profiling of circulating T cells to identify adult patients with a severe disease course (PredictSURE IBD).[30] They identified a subgroup of patients with significantly more aggressive disease, with an earlier need for treatment escalation and more escalations over time. Of note, these results were not replicated in a cohort of pediatric IBD patients.[31]

Finally, the most robust predictions will likely come from models that incorporate multiple variables, as opposed to a single factor. Our hypotheses were that i) genetics plays a greater role in pediatric-onset disease, whereas the exposome may have greater impact in older patients, and ii) variants different from susceptibility variants may predict the prognosis.

Three studies focused specifically on children, including multiple types of variables and providing discriminative performance. First, PROSPECT is a tool to predict and stratify the risk for both adult and pediatric patients.[16] The model identified patients with complicated behavior and included disease location, serological markers (ASCA, CBir1, and ANCA), and the rs2066847 *NOD2* frameshift mutation. Second, the pediatric RISK cohort identified age, race, ileal location, anti-microbial antibodies (ASCA IgA and CBir1), and an ileal extracellular matrix gene signature.[17] Contrary to our study, these studies did not analyze SNPs other than susceptibility variants (only *NOD2* in PROSPECT). Finally, a new study based on RISK patients identified a novel blood protein signature to predict penetrating and stricturing complications.[18] The models of these studies showed moderate to good predictive performance, with a Harrell's c-index of 0.75 for children in PROSPECT, an AUC of 0.72 for the best predictive model of RISK, and an AUC of 0.69 and 0.79 for B2 and B3, respectively, in the study of Ungaro et al, versus 0.80 following internal validation for PREDICT-EPIMAD. In our study, the sensitivity was 0.79 and specificity 0.74 after correction for over-optimism versus 0.69 and 0.71, respectively, in RISK (data not given in the other studies). In the recent literature, more focus has been given to a high negative predictive value,[17,30] *i.e.* the probability of being free from complication for patients classified as "low-risk". PREDICT-EPIMAD achieved a high NPV of 0.87, versus 0.94 in the RISK study. Of note, the NPV depends on the proportion of events, which was higher in our population (35% of patients presenting an intestinal resection and/or complicated behavior at five years) versus that of the RISK study (9% presenting a complicated behavior after 3 years of follow-up), in which the lower prevalence of complications in the cohort may also explain, in part, the lower PPV of the model. Finally, the PPV of our model was 0.61 versus 0.24 for the RISK prediction score. Notably, a "treat-none" strategy would have resulted in a NPV of 0.65 and "treat-all" strategy a PPV of 0.35. Thus, the PREDICT-EPIMAD model is a good trade-off between the "treat-all" and "treat-none" strategies in terms of NPV and PPV. This was confirmed by decision curve analysis, which assessed the clinical utility of PREDICT-EPIMAD.

The major strength of our study was that it was population-based, thus avoiding a selection bias. Another strength was the use of multivariable regularized regression methods that made it possible to simultaneously enter a number of genetic markers and produce a sparse model that included only a few markers to eliminate noisy variables and propose an easy-to-use tool. Moreover, only variants that were stable using bootstrap resampling were selected in discovery cohort. These variants were then confirmed using external cohort, thus highly reducing the risk of false selection of variants and preventing over-fitting. A rigorous internal validation was then performed using bootstrap samples in order to correct for optimism bias. Another advantage of our score is that serotypes and genotypes are stable over time and the assays are easy to perform. Of note, we provide, for the first time, a web-tool that allows clinicians to easily use the PREDICT-EPIMAD score in daily-practice. Concerning treatment, patients were little exposed to anti-TNF and only exposed to immunosuppressants towards the end of the study, as they

were diagnosed between 1988 and 2004. Indeed, 60 % (n=93) and 19 % (n=29) of patients were exposed to immunosuppressive therapy and anti-TNF during the 5 years following CD diagnosis, respectively. In particular, only 1 patient was exposed to early anti-TNF as defined in RISK study, that is in the 3 months following diagnosis. This is an advantage for developing a discriminative score because the role of genetics is likely to be less confounded by the effect of treatment on the disease course than in recent studies. We chose to evaluate the five-year risk of complications, as such risk occurs mostly within the first few years after diagnosis,[5] with the long-term risk model eventually becoming confounded or outdated with evolving therapeutic strategies.

Our study also had several limitations. The first relates to the sample size of the discovery cohort used to derive the score. Lack of power may explain the absence of association between some clinical variables like ano-perineal lesions and disease evolution. However, population-based cohorts of pediatric-onset CD are rare and only a few studies have focused on this specific population. Second, data on smoking were missing. Exposure to smoking is difficult to assess in a pediatric population, with a certain degree of subjectivity of the response, which can lead to inaccurate classification. Ethnic origin was also unavailable since it is forbidden to record this parameter due to stringent ethical considerations in France. Third, we considered the best predictive model as the one achieving the highest AUC with the smallest number of variables but it has to be noticed that CIs of AUC and other predictive performances overlapped between the different models that were tested. Finally, external validation is necessary. We did not use the external cohort as a validation cohort because: i) the sample size was small, ii) the serological data were not available, and iii) the patients only came from expert centers. Validation in an external cohort is made difficult by the fact that pediatric population-based cohorts are rare and that recent cohorts are highly exposed to immunosuppressants and anti-TNF, thus modifying the natural history of the disease. Clinical utility may also be further evaluated using a randomized biomarker-stratified trial.[32]

CONCLUSIONS : In conclusion, PREDICT-EPIMAD is a score that combines clinical, serological, and genetic factors to predict the evolution of pediatric-onset inflammatory CD to a complicated disease course. The proposed score is easy to determine, as it is based on 6 SNPs and the usual clinical and serological data. A web-tool to classify patients as high- or low-risk for complications is freely available at www.predict-epimad.com, which is a first step to address an the unmet need, to aid physicians in providing treatment options for children and their families, better inform them of the prognosis, and justify early intervention to change the natural progression of the disease. Of note, external validation is still needed to confirm the predictive capacities of the score. Our medium- to long-term goal is to bring the current tool to clinical practice, which would allow physicians to propose the appropriate early intensive therapy to high-risk patients or, alternatively, a step-up approach for low-risk patients.

References

1 Torres J, Mehandru S, Colombel J-F, Peyrin-Biroulet L. Crohn's disease. *Lancet Lond Engl* 2017; **389**: 1741–55.

2 Torres J, Colombel J-F. Genetics and phenotypes in inflammatory bowel disease. *Lancet Lond Engl* 2016; **387**: 98–100.

3 Ghione S, Sarter H, Fumery M, *et al.* Dramatic Increase in Incidence of Ulcerative Colitis and Crohn's Disease (1988-2011): A Population-Based Study of French Adolescents. *Am J Gastroenterol* 2018; **113**: 265–72.

4 Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* 2021; **18**: 56–66.

5 Duricova D, Burisch J, Jess T, Gower-Rousseau C, Lakatos PL, ECCO-EpiCom. Age-related differences in presentation and course of inflammatory bowel disease: an update on the population-based literature. *J Crohns Colitis* 2014; **8**: 1351–61.

6 Herzog D, Fournier N, Buehr P, *et al.* Prevalence of intestinal complications in inflammatory bowel disease: a comparison between paediatric-onset and adult-onset patients. *Eur J Gastroenterol Hepatol* 2017; **29**: 926–31.

7 Colombel J-F, Narula N, Peyrin-Biroulet L. Management Strategies to Improve Outcomes of Patients With Inflammatory Bowel Diseases. *Gastroenterology* 2017; **152**: 351–361.e5.

8 Khanna R, Bressler B, Levesque BG, *et al.* Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* 2015; **386**: 1825–34.

9 Bae JM, Lee HH, Lee B-I, *et al.* Incidence of psoriasiform diseases secondary to tumour necrosis factor antagonists in patients with inflammatory bowel disease: a nationwide population-based cohort study. *Aliment Pharmacol Ther* 2018; **48**: 196–205.

10 Kirchgesner J, Lemaitre M, Carrat F, Zureik M, Carbonnel F, Dray-Spira R. Risk of Serious and Opportunistic Infections Associated With Treatment of Inflammatory Bowel Diseases. *Gastroenterology* 2018; **155**: 337–346.e10.

11 Lichtenstein GR, Rutgeerts P, Sandborn WJ, *et al.* A pooled analysis of infections, malignancy, and mortality in infliximab- and immunomodulator-treated adult patients with inflammatory bowel disease. *Am J Gastroenterol* 2012; **107**: 1051–63.

12 Savoye G, Salleron J, Gower-Rousseau C, *et al.* Clinical predictors at diagnosis of disabling pediatric Crohn's disease. *Inflamm Bowel Dis* 2012; **18**: 2072–8.

13 Dubinsky MC, Lin Y-C, Dutridge D, *et al.* Serum Immune Responses Predict Rapid Disease Progression among Children with Crohn's Disease: Immune Responses Predict Disease Progression. *Am J Gastroenterol* 2006; **101**: 360–7.

14 Gupta N, Cohen SA, Bostrom AG, *et al.* Risk factors for initial surgery in pediatric patients with Crohn's disease. *Gastroenterology* 2006; **130**: 1069–77.

15 Amre DK, Lu S-E, Costea F, Seidman EG. Utility of serological markers in predicting the early occurrence of complications and surgery in pediatric Crohn's disease patients. *Am J Gastroenterol* 2006; **101**: 645–52.

16 Siegel CA, Horton H, Siegel LS, *et al.* A validated web-based tool to display individualised Crohn's disease predicted outcomes based on clinical, serologic and genetic variables. *Aliment Pharmacol Ther* 2016; **43**: 262–71.

17 Kugathasan S, Denson LA, Walters TD, *et al.* Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet Lond Engl* 2017; **389**: 1710–8.

18 Ungaro RC, Hu L, Ji J, *et al.* Machine learning identifies novel blood protein predictors of penetrating and stricturing complications in newly diagnosed paediatric Crohn's disease. Aliment Pharmacol Ther. 2021 Jan;53(2):281-290.

19 Jakobsen C, Cleynen I, Andersen PS, *et al.* Genetic susceptibility and genotype-phenotype association in 588 Danish children with inflammatory bowel disease. *J Crohns Colitis* 2014; **8**: 678–85.

20 Atia O, Kang B, Orlansky-Meyer E, *et al.* Existing prediction models of disease course in pediatric Crohn's disease are poorly replicated in a prospective inception cohort. *J Crohns Colitis* 2022; published online Jan 10. DOI:10.1093/ecco-jcc/jjac005.

21 Gower-Rousseau C, Salomez JL, Dupas JL, *et al.* Incidence of inflammatory bowel disease in northern France (1988-1990). *Gut* 1994; **35**: 1433–8.

22 Jung C, Colombel J-F, Lemann M, *et al.* Genotype/phenotype analyses for 53 Crohn's disease associated genetic polymorphisms. *PloS One* 2012; **7**: e52223.

23 Zhao Q, Shi X, Xie Y, Huang J, Shia B, Ma S. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Brief Bioinform* 2015; **16**: 291–303.

24 Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Ser B Stat Methodol* 2010; **72**: 417–73.

25 Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; **352**: i6.

26 Ungaro RC, Aggarwal S, Topaloglu O, Lee W-J, Clark R, Colombel J-F. Systematic review and meta-analysis: efficacy and safety of early biologic treatment in adult and paediatric patients with Crohn's disease. *Aliment Pharmacol Ther* 2020; **51**: 831–42.

27 Dubinsky MC, Kugathasan S, Mei L, *et al.* Increased immune reactivity predicts aggressive complicating Crohn's disease in children. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc* 2008; **6**: 1105–11.

28 Ricciuto A, Aardoom M, Meyer EO, *et al.* Predicting Outcomes in Pediatric Crohn's Disease for Management Optimization: Systematic Review and Consensus Statements From the Pediatric Inflammatory Bowel Disease-Ahead Program. Gastroenterology. 2021 Jan;160(1):403-436.e26.

29 Aloi M, Viola F, D'Arcangelo G, *et al.* Disease course and efficacy of medical therapy in stricturing paediatric Crohn's disease. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver* 2013; **45**: 464–8.

30 Biasci D, Lee JC, Noor NM, *et al.* A blood-based prognostic biomarker in IBD. *Gut* 2019; **68**: 1386–95.

31 Gasparetto M, Payne F, Nayak K, *et al.* Transcription and DNA Methylation Patterns of Blood-Derived CD8+ T Cells Are Associated With Age and Inflammatory Bowel Disease But Do Not Predict Prognosis. Gastroenterology. 2021 Jan;160(1):232-244.e7.

32 Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst* 2010; **102**: 152–60.

Tables

Table 1. Clinical and serological characteristics of patients from the Epimad discovery cohort (n = 156).

Variable	
Follow-up in years (median, IQR)	10.4 [7.2 – 14.9]
Male gender n (%)	85 (54.5 %)
Clinical data at diagnosis	
Age (median, IQR)	14.3 [11.9 – 16.0]
Familial history of IBD n (%)	24 (15.4 %)
Location at diagnosis n (%)	
Ileal (L1)	17 (11.0%)
Colonic (L2)	29 (18.7 %)
Ileo-colonic (L3)	109 (70.3 %)
Upper digestive disease (L4) n (%)	60 (38.5 %)
Ano-perineal lesions n (%)	18 (11.5 %)
Extra-intestinal symptoms n (%)	39 (25.0 %)
Treatments during the 5 years following diagnosis n (%)	
Systemic steroids	126 (80.8 %)
Systemic 5ASA	133 (85.3 %)
Immunosupressants	93 (59.6 %)
Anti-TNF	29 (18.6 %)
Serological status at inclusion n (%)	
ASCA IgA	111 (71.1 %)
ASCA IgG	135 (86.5 %)
pANCA	28 (17.9 %)
Anti-CBir1	80 (51.3 %)
Anti-Fla2	43 (27.6 %)
Anti-FlaX	44 (28.2 %)
Anti-OmpC	23 (14.7 %)
Clinical outcomes at 5 years	
Composite outcome: intestinal resection and/or complicated behavior	54 (34.6 %)
Intestinal resection	35 (22.4 %)
Complicated behavior (B2/B3)	46 (29.5 %)

IQR : Interquartile range

Table 2. Univariable analyses of clinical and serological variables (n = 156 patients). Variables highlighted in grey are those with p ≤ 0.2, which were selected for predictive models.

	Main outcome: intestinal resection or complicated behavior at 5 years		**Intestinal resection at 5 years**		**Complicated behavior at 5 years**	
	OR (95% CI)	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Clinical variables at diagnosis						
Male Gender	1.07 [0.55; 2.07]	0.845	0.85 [0.40; 1.81]	0.680	1.45 [0.72; 2.91]	0.302
Age at diagnosis†	1.07 [0.94; 1.21]	0.319	1.05 [0.90; 1.21]	0.540	1.07 [0.93; 1.22]	0.333
Extra-intestinal manifestations	1.08 [0.50; 2.30]	0.846	0.86 [0.35; 2.09]	0.740	1.08 [0.49; 2.39]	0.839
Family history of IBD	1.43 [0.59; 3.47]	0.431	1.53 [0.58; 4.05]	0.392	1.24 [0.49; 3.13]	0.654
Ileal (L1) or ileocolonic disease (L3)	1.87 [0.74; 4.71]	0.184	1.15 [0.43; 3.08]	0.787	1.78 [0.67; 4.72]	0.244
Upper gastrointestinal disease (L4)	0.71 [0.36; 1.42]	0.339	0.67 [0.30; 1.50]	0.333	0.91 [0.45; 1.86]	0.803
Anoperineal disease	0.94 [0.33; 2.65]	0.903	0.66 [0.18; 2.43]	0.535	0.91 [0.30; 2.78]	0.866
Serological variables at inclusion						
ASCA-IgA positive	1.25 [0.59; 2.62]	0.558	1.22 [0.52; 2.87]	0.643	1.42 [0.65; 3.13]	0.380
ASCA-IgG positive	2.50 [0.80; 7.85]	0·116	3.07 [0.68; 13.89]	0.145	1.92 [0.61; 6.05]	0.267
pANCA positive	0.18 [0.05; 0.63]	0.007	0.22 [0.05; 0.98]	0.048	0.24 [0.07; 0.83]	0.024
Anti-CBir1 positive	0.82 [0.43; 1.60]	0.569	0.75 [0.35; 1.59]	0.455	1.05 [0.53; 2.09]	0.885
Anti-Fla2 positive	0.65 [0.30; 1.41]	0.279	0.67 [0.34; 1.31]	0.240	0.89 [0.41; 1.96]	0.789
Anti-FlaX positive	0.73 [0.34; 1.54]	0.405	0.70 [0.29; 1.69]	0.426	1.00 [0.47; 2.16]	0.992
Anti-OmpC positive	1.92 [0.78; 4.70]	0.154	1.27 [0.46; 3.50]	0.650	2.07 [0.83; 5.14]	0.116

†ORs are presented for each one-year increase in age. OR : odds ratio; 95% CI : 95% confidence interval.

Table 3. SNPs included in the final multivariable model for each outcome. Risk alleles are shown in bold.

Chromosome	candidate gene	SNP	Minor Allele dbSNP (CEU)	Major allele dbSNP (CEU)	Minor allele frequency in discovery sample		
					All (n = 156)	Absence of complication	Complication
Main outcome: intestinal resection and/or complicated behavior at 5 years							
chr4	NFKB1	rs230530	G	**A**	0.47	0.52	0.36
chr7	IKZF1	rs1456896	**C**	T	0.27	0.22	0.36
chr10	UBE2D1	rs1819658	T	**C**	0.18	0.22	0.11
chr10	IATPR	rs2755996	**T**	C	0.10	0.07	0.16
chr13	TNFSF11	rs2062305	**G**	A	0.49	0.43	0.60
chr17	IKZF3	rs907091	**T**	C	0.50	0.44	0.61
Secondary outcome: intestinal resection at 5 years							
chr3	IHPK1	rs9872864	G	**A**	0.51	0.55	0.37
chr13	HMGB1	rs1045411	T	**C**	0.23	0.27	0.10
chr16	NOD2	rs2066845	**C**	G	0.07	0.05	0.14
chr17	ORMDL3	rs8076131	G	**A**	0.49	0.54	0.33
Secondary outcome: complicated behavior at 5 years							
chr1	TLR5	rs851192	G	**C**	0.39	0.42	0.30
chr6	PRDM1	rs548234	C	**T**	0.31	0.34	0.26
chr7	IKZF1	rs1456896	**C**	T	0.27	0.23	0.35
chr10	KLF6	rs6601764	C	**T**	0.44	0.50	0.32
chr14	AKT1	rs2494731	**C**	G	0.31	0.28	0.40

Table 4. Discriminative performance from internal validation of the models including SNPs and clinical and serological data (n = 156 Epimad patients). Corrections for optimism bias were performed using 1,000 bootstrap samples. The best predictive model (PREDICT-EPIMAD) is highlighted in grey.

Outcome	Variables	AUC in discovery cohort	Performances corrected for optimism bias (internal validation)				
			AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Main outcome: intestinal resection and/or complicated behavior at 5 years	6 SNPs	0.81 [0.74; 0.88]	0.77 [0.70; 0.83]	0.79 [0.58; 0.93]	0.67 [0.48; 0.86]	0.56 [0.42; 0.73]	0.86 [0.76; 0.95]
	6 SNPs, ileal location	0.82 [0.75; 0.89]	0.78 [0.71; 0.84]	0.81 [0.68; 0.91]	0.68 [0.55; 0.79]	0.57 [0.45; 0.68]	0.87 [0.80; 0.94]
	6 SNPs, ileal location, pANCA	0.84 [0.77; 0.90]	0.80 [0.73; 0.85]	0.79 [0.67; 0.88]	0.74 [0.62; 0.83]	0.61 [0.48; 0.73]	0.87 [0.80; 0.93]
	6 SNPs, ileal location, pANCA, ASCA IgG	0.84 [0.78; 0.90]	0.79 [0.73; 0.85]	0.77 [0.66; 0.90]	0.74 [0.56; 0.84]	0.62 [0.47; 0.74]	0.86 [0.79; 0.94]
	6 SNPs, ileal location, pANCA, ASCA IgG, anti-OmpC	0.84 [0.78; 0.90]	0.79 [0.72; 0.85]	0.79 [0.68; 0.90]	0.71 [0.56; 0.81]	0.59 [0.47; 0.70]	0.87 [0.80; 0.94]
Intestinal resection at 5 years	4 SNPs	0.80 [0.71; 0.87]	0.76 [0.67; 0.83]	0.71 [0.52; 0.92]	0.71 [0.44; 0.86]	0.42 [0.26; 0.60]	0.90 [0.83; 0.96]
	4 SNPs, pANCA	0.81 [0.73; 0.88]	0.76 [0.69; 0.83]	0.80 [0.54; 0.96]	0.62 [0.42; 0.86]	0.37 [0.22; 0.61]	0.92 [0.85; 0.98]
	4 SNPs, pANCA, ASCA IgG	0.81 [0.74; 0.88]	0.77 [0.70; 0.83]	0.80 [0.55; 0.95]	0.62 [0.40; 0.86]	0.38 [0.23; 0.60]	0.92 [0.85; 0.98]
Complicated behavior at 5 years	5 SNPs	0.77 [0.68; 0.84]	0.72 [0.64; 0.80]	0.75 [0.53; 0.90]	0.62 [0.44; 0.81]	0.45 [0.32; 0.61]	0.86 [0.78; 0.93]
	5 SNPs, pANCA	0.80 [0.72; 0.88]	0.76 [0.68; 0.83]	0.72 [0.56; 0.87]	0.72 [0.56; 0.84]	0.52 [0.37; 0.67]	0.86 [0.79; 0.93]
	5 SNPs, pANCA, anti-OmpC	0.80 [0.72; 0.88]	0.75 [0.67; 0.82]	0.71 [0.56; 0.84]	0.71 [0.55; 0.82]	0.51 [0.37; 0.65]	0.85 [0.78; 0.91]

AUC : Area under the ROC curve

Figures

Figure 1. Flow chart of the statistical analysis strategy.

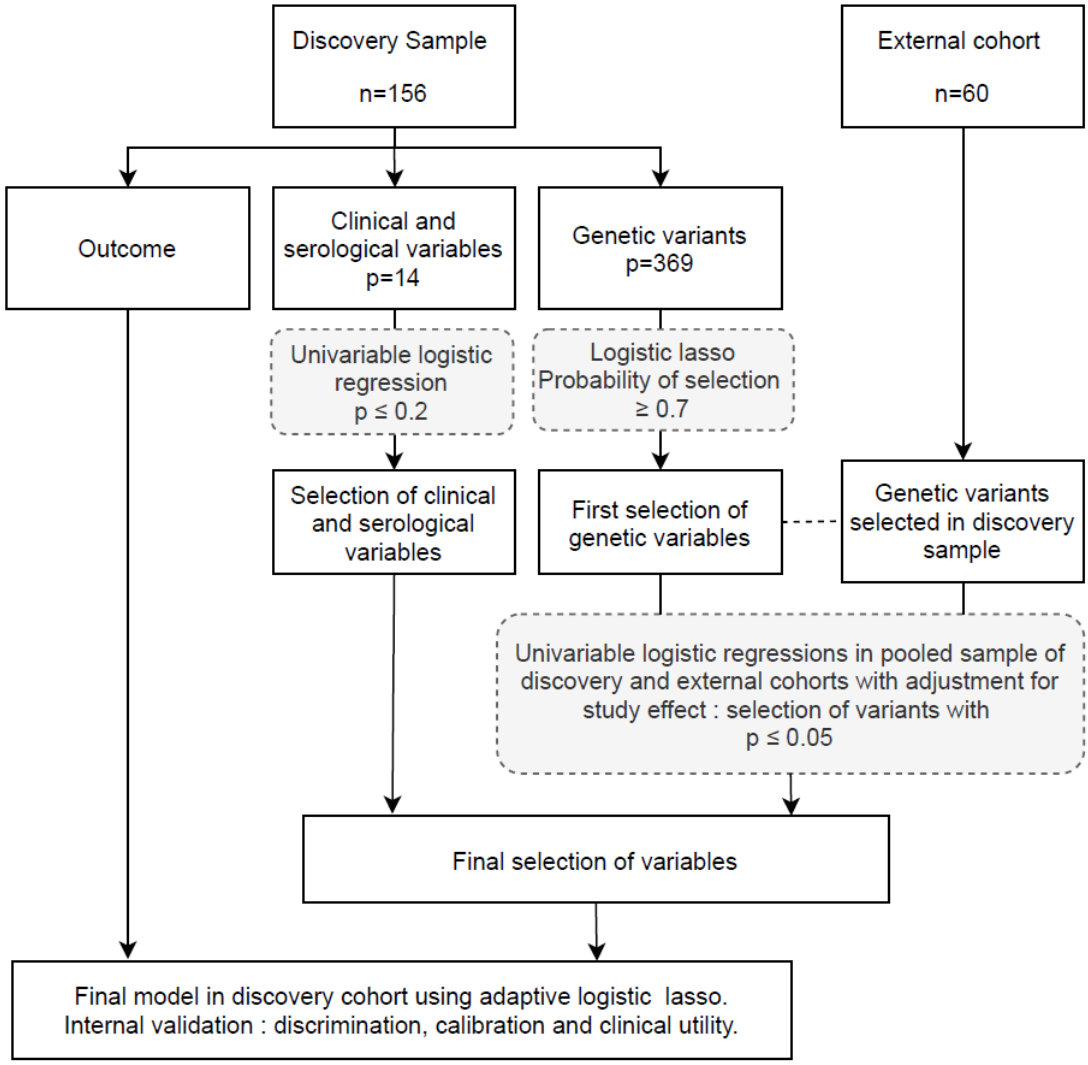


Figure 2. Discrimination from internal validation of the 6-SNP-based signature predicting the composite outcome (PREDICT-EPIMAD). a) Receiver operating characteristic curve (ROC) of PREDICT-EPIMAD (black line) compared to model including clinical variables only (green line). Corrected AUC refers to AUC corrected for optimism bias using 1,000 bootstrap samples. b) Predicted probability of complications: red rectangle: "High-risk" patients according to PREDICT-EPIMAD, blue rectangle: "Low-risk" patients according to PREDICT-EPIMAD. The red and blue points represent the observed presence or absence of complications at five years, respectively.

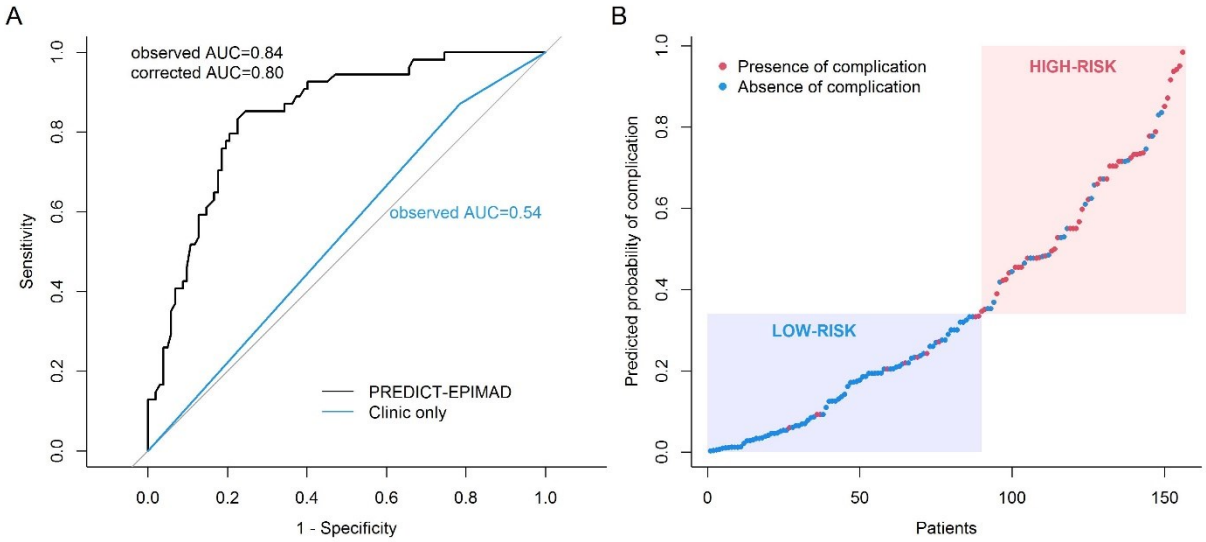
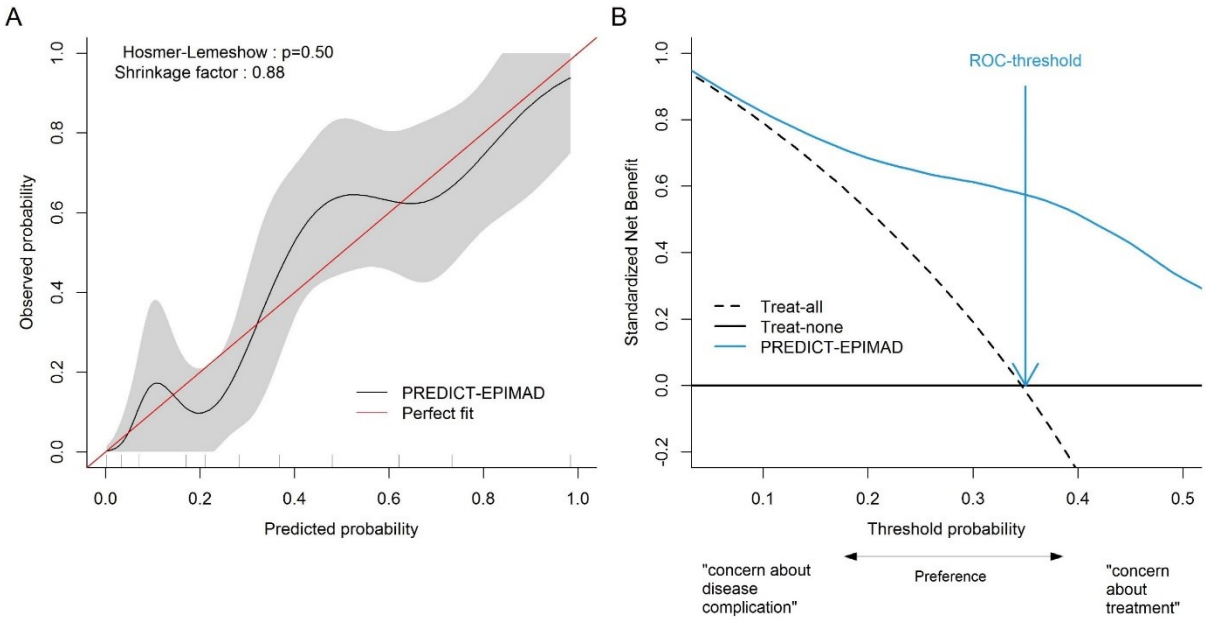


Figure 3. Calibration and clinical utility from internal validation of the 6-SNP-based signature predicting the composite outcome (PREDICT-EPIMAD). a) Calibration plot showing the observed versus predicted probabilities: the black line represents the calibration curve, with the confidence interval in grey, and the red line the curve for a perfectly calibrated model. b) Decision curve showing the standardized net benefit according to threshold probabilities. The net benefit is calculated as (proportion of true positives) – (proportion of false positives)*$p_t/(1-p_t)$, where p_t is the threshold probability. The horizontal axis can be interpreted as the "preference" axis: the farther the threshold probability is to the left, the greater the "concern about disease complication"; the more on the right, more "concern about treatment". The interpretation of the decision curve is that the strategy with the highest benefit over the range of probabilities has the highest clinical value.



Discussion générale, perspectives et conclusions

Discussion générale

1. Résumés des principaux résultats

Ce travail a examiné divers aspects des maladies inflammatoires chroniques de l'intestin dans la vraie vie, de l'épidémiologie descriptive en population générale à l'épidémiologie analytique et la médecine personnalisée.

Dans une **première partie**, les travaux mettent en avant le principal objectif des registres en population générale, à savoir le calcul de l'incidence et l'analyse de son évolution temporelle selon l'âge et le sexe. Ainsi, l'analyse de 30 ans de données d'incidence a montré que ces incidences augmentent régulièrement et plus particulièrement chez les enfants et jeunes adultes. L'incidence de la RCH est également en augmentation, et plus particulièrement chez les femmes. Pour la première fois, la prévalence a été estimée et pourrait atteindre 0,6 % de la population en 2030, avec, de plus, un vieillissement de la population atteinte de MICI. Ces résultats montrent que le plateau d'incidence observé dans certains pays occidentaux n'est pas encore atteint en France. Dans un second chapitre, nous nous sommes intéressés à l'impact des MICI sur la vie sociale et professionnelle des patients ayant débuté leur maladie dans l'enfance : ces patients ont au final un niveau d'études plus élevé et un taux de chômage plus faible que ceux de la population générale de même âge et de même sexe. Cependant, les patients mentionnent tout de même de nombreuses difficultés dans leur parcours scolaire ainsi qu'un impact sur leur choix d'études et de profession. De manière intéressante, ils sont plus nombreux à travailler dans le secteur de la santé et dans la fonction publique. Ces deux derniers résultats n'ont jamais été montrés, à notre connaissance.

Dans la **seconde partie**, des travaux méthodologiques ont été présentés avant la construction d'un nouveau score permettant d'estimer un risque de complications à 5 ans. Les travaux méthodologiques ont comparé différentes méthodes d'intégration de données cliniques et omiques. PREDICT-EPIMAD, un score simple d'utilisation, reposant sur 8 variables cliniques, génétiques et sérologiques a ensuite été proposé pour estimer le risque de complications à 5 ans de la MC à début pédiatrique.

Les résultats ont été discutés à la fin de chacune des études présentées. La discussion qui suit est donc axée sur une discussion plus générale sur la place des registres épidémiologiques, sur les apports des données omiques, sur la portée des résultats ainsi que sur les perspectives.

2. Place des registres épidémiologiques par rapport aux autres types d'études et de bases de données

Les principales forces d'un registre épidémiologique, qui en font un outil unique, résident dans l'exhaustivité du recueil des cas en population générale, le contrôle de la qualité des données recueillies et dans l'expertise des dossiers. En cela, les registres s'opposent d'une part aux données recueillies dans des centres experts ou dans le cadre d'essais randomisés et d'autre part aux données issues de bases de données administratives. Si les différences avec les premières sont assez évidentes, les différences avec les secondes le sont moins et les bases de données administratives sont souvent utilisées aux mêmes fins que les données de registres épidémiologiques.

En tant que bases de données observationnelles, les registres s'opposent tout d'abord aux essais randomisés. Cependant, ces deux types d'études sont des formes de recherche complémentaires permettant de vérifier que les résultats des essais cliniques se concrétisent bien par des effets tangibles au niveau de la population (275). La principale force des essais réside dans leur validité interne liée à la randomisation qui permet de s'assurer que les effets observés sont bien liés à l'exposition. Cependant, la validité externe (généralisation) des résultats des essais n'est pas assurée. En effet, même si un traitement a été prouvé efficace dans un essai randomisé, les études observationnelles peuvent être utiles pour mettre en évidence : une sur- ou sous-utilisation des traitements malgré des recommandations d'usage ; le bénéfice et les effets secondaires des traitements dans la vraie vie qui peuvent différer des essais randomisés. Ainsi, le registre Epimad a récemment publié plusieurs résultats sur l'utilisation des traitements en population à partir des données de la cohorte pédiatrique appelée « Inspired » : i) En parallèle de l'augmentation de l'utilisation d'immunosuppresseurs et d'anti-TNF, un risque diminué à la fois de résections intestinales et de complications sténosantes chez les patients atteints de MC et une baisse importante du risque de colectomie chez les patients atteints de RCH débutant dans l'enfance ont été observés (89,90) ; ii) Le taux d'échec à 5 ans des anti-TNF est d'environ 60 % dans la MC et 70 % dans la RCH pédiatriques.

La perte de réponse représente environ les deux tiers des échecs, tant pour la MC que pour la RCH (167).

L'exhaustivité est une force mais n'est pas toujours primordiale. Par exemple, pour la création du score PREDICT-EPIMAD, le caractère populationnel n'est pas obligatoire, mais est un plus par rapport à une étude réalisée dans un centre expert car un plus large éventail de patients sera représenté, et non pas seulement des cas graves ayant des caractéristiques spécifiques. En effet, dans le registre Epimad, près de 80 % des cas sont diagnostiqués par des gastroentérologues libéraux. Pour l'étude sur le niveau d'étude et l'insertion professionnelle, en revanche, la représentativité de l'ensemble de la population des MICI est particulièrement importante mais l'échantillon peut être soumis au biais de non-réponse.

Enfin, et c'est le cœur du travail d'un registre épidémiologique en population générale, l'exhaustivité est indispensable pour l'étude des incidences et des prévalences. Une méta-analyse récente des études d'incidences en Océanie permet d'illustrer l'apport des données en population générale par rapport aux données hospitalières. L'incidence totale des MICI basée sur des études épidémiologiques en population générale était estimée à 27,2 /10^5 versus 11,3 à partir de données hospitalières (40). De nombreuses études d'incidences sont basées sur des bases de données médico-administratives, notamment dans les pays nordiques qui disposent depuis plusieurs décennies de bases couvrant la quasi-totalité de la population. Si ces bases de données peuvent être exhaustives, elles souffrent cependant de l'absence d'expertise des dossiers. L'extraction des cas de MICI est basée sur des algorithmes permettant le repérage des cas basé sur le codage diagnostic des hospitalisations et/ou la prise de traitements. En Norvège une étude a étudié l'impact de la définition des cas sur l'estimation des incidences (68). Le résultat est illustré sur la Figure 48 présentant les évolutions temporelles d'incidence de MC et de RCH pour 3 définitions de cas. On voit que, non seulement la valeur de l'incidence est modifiée selon la définition des cas mais également que les tendances temporelles sont différentes pouvant mener à des conclusions différentes - stabilité ou décroissance - selon la définition utilisée (Figure 48). Dans cette étude, 15 % des cas étaient enregistrés avec à la fois un diagnostic de MC et de RCH. Sans retour au dossier médical, il est difficile de distinguer pour ces dossiers les cas correspondant à des colites inclassées, des erreurs de codage ou des changements de diagnostics au cours du temps. Au

Danemark, deux études d'incidences ont été publiées en 2022 et 2023 avec des définitions différentes (49,171) mais résultaient en des estimations et des tendances similaires.

Dans tous les cas, l'utilisation de ces bases de données nécessite des validations préalables. En Suède, une étude a démontré des valeurs prédictives positives élevées (VPP > 90 %) pour le repérage des cas à partir de deux registres (Swedish National Patient Register (NPR) et Swedish Quality Register SWIBREG) (276). Cependant cette étude ne donnait pas d'estimation des valeurs de sensibilité et spécificité qui sont importantes pour l'utilisation à des fins d'études épidémiologiques des incidences et des prévalences. Au Canada, plus de 5 000 algorithmes ont été testés (277). Le meilleur algorithme pour identifier les patients âgés de 18 à 64 ans au moment du diagnostic était cinq contacts avec un médecin ou hospitalisations en 4 ans (sensibilité : 76,8 % ; spécificité : 96,2 % ; valeur prédictive positive : 81,4 % ; valeur prédictive négative : 95,0 %). On note cependant que 23 % des cas ne sont pas repérés par cet algorithme et que 19 % des patients identifiés comme MICI n'en sont pas.

Les études sur la mortalité et le risque de cancer nécessitent également des données issues de la population générale. Le registre Epimad a ainsi permis de montrer que les patients atteints de MICI débutant dans l'enfance ont un risque accru à la fois de cancer (SIR : 2,7) et de mortalité (SMR : 1,7), notamment pour le cancer colorectal (166).

Les données des registres sont de plus en plus concurrencées par les données des bases de données médico-administratives. En France, l'accès aux données du SNDS (système national des données de santé) est facilité. Depuis 2016, le SNDS est alimenté par :

- La base de données « SNIIRAM » (Système national d'information inter-régimes de l'assurance maladie), qui contient les informations sur toutes les dépenses de l'assurance maladie ;
- La base de données « PMSI » (Programme de médicalisation des systèmes d'information), qui compile les données d'analyse de l'activité des établissements de santé ;
- La base de données du CépiDc (Centre d'épidémiologie sur les causes médicales de décès), qui recense les informations relatives aux causes de décès.

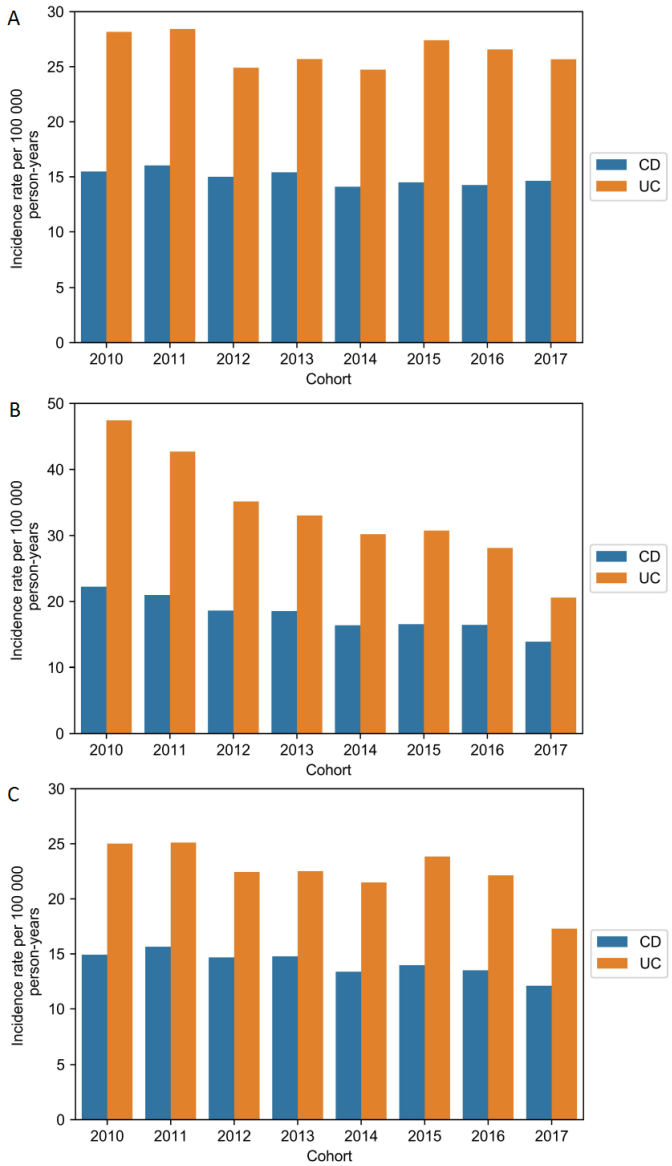
Dans le cadre des MICI ces données ont été utilisées, notamment, pour des analyses en vie réelle et pour reproduire des résultats d'essais cliniques (278,279). Les principaux avantages

de ces données sont qu'elles ne nécessitent pas un recueil spécifique et qu'elles couvrent l'ensemble du territoire et près de 95 % de la population.

Ces données issues des bases médico-administratives manquent cependant de données cliniques sur les caractéristiques des maladies (localisation, phénotype, activité de la maladie etc.). Ainsi, de nombreuses études d'incidences, notamment dans les pays nordiques, ne s'accompagnent pas d'une description des caractéristiques cliniques des patients. En France, le projet ALGO-IBD a pour but de créer des algorithmes permettant d'identifier les caractéristiques principales et l'activité clinique des MICI en utilisant l'intégration de trois sources de données réelles : le Système National des Données de Santé (SNDS), l'entrepôt de données de santé de l'Assistance Publique-Hôpitaux de Paris (EDS AP-HP), et une cohorte prospective de patients atteints de MICI suivis dans des centres de l'AP-HP (SUVIMIC).

Les registres souffrent aussi de certaines faiblesses liées à l'absence de données de suivi et au manque de données associées au contact du patient comme des prélèvements. De plus, les registres sont des outils coûteux et difficiles à mettre en place, limitant ainsi leur nombre (n=53, toutes pathologies confondues). Ainsi, le recueil des données du registre Epimad nécessite 5 ETP (équivalent temps plein) d'infirmières-enquêtrices ou ARC. Ces enquêtrices recueillent les données dans l'ensemble des hôpitaux publics et privés et dans l'ensemble des cabinets des gastroentérologues de la zone, soit quelques 250 lieux à visiter plusieurs fois par an. A cela s'ajoutent un contrôle et un recueil complémentaire éventuel à partir des données issues des PMSI des DIM nécessitant l'analyse et le tri de longues listes de patients. De ce fait, contacts avec les patients pour effectuer par exemple des prélèvements ou le remplissage de questionnaires de qualité de vie ne sont généralement envisagés que pour des études très ponctuelles avec une demande d'autorisation aux instances légales (CPP, ANSM, CNIL).

Figure 48 : Impact de la définition de cas sur l'estimation de l'incidence. A) Au moins une visite pour MICI et au moins deux prescriptions ou deux visites. Les patients ayant reçu un traitement dans les 60 jours précédant la première visite sont exclus. B) Au moins deux enregistrements (pas de données de prescription) C) Au moins deux enregistrements. Les patients ayant reçu un traitement dans les 60 jours précédant la première visite sont exclus. Tiré de Lirhus et al, Clinical Epidemiology, 2021.



3. Apports des données omiques

L'analyse des données issues de patients du registre EPIMAD a montré que l'addition de données génétiques permettait d'améliorer la prédiction de la complication de la maladie de Crohn à début pédiatrique. Par le passé, les modèles cliniques n'avaient pas fait leur preuve pour prédire la complication de la maladie, même si un certains nombres de facteurs avaient été identifiés comme associés à la complication de la maladie.

D'une manière générale, dans tous les domaines de la santé, les données omiques font l'objet d'un intérêt grandissant du fait de l'amélioration des techniques qui rendent les résultats plus fiables et permettent d'explorer simultanément un grand nombre de variables. La réduction des coûts des expériences à haut débit est aussi un argument fort pour les utiliser de plus en plus. Ces données semblent être le graal et la solution à des questions non résolues en médecine de précision (280). Cependant les méthodes d'analyses sont complexes et ont fait l'objet d'une littérature abondante ces dernières années, notamment pour l'analyse de jeux de données multi-omiques. Une revue des outils mathématiques utilisés pour l'intégration de données omiques en 2016 est publiée dans Bersanelli et al. (281). Deux bonnes comparaisons de méthodes plus récentes sont publiées dans Pierre-Jean et al. (282) et Cantini et al. (283). L'analyse multi-omiques est plus souvent réalisée dans un objectif de compréhension de la biologie moléculaire et cellulaire que dans un objectif prédictif. Les derniers travaux se placent d'ailleurs à l'échelle de la cellule unique ou incluent une information spatiale (284). Les quelques publications s'intéressant à l'intégration de données cliniques et omiques dans un objectif de construction de score sont citées dans cette thèse.

L'apport des données omiques dans un score clinique n'est pas toujours évident. Dans certains domaines les données cliniques ont fait leur preuve et l'ajout de données omiques ne permet pas toujours de surpasser les modèles basés sur des facteurs cliniques forts. Par exemple, une étude comparant des mesures d'expression de gènes à des marqueurs pronostiques conventionnels pour prédire les métastases à distance chez les patientes atteintes de cancer du sein a montré que les données d'expression de gènes n'amélioraient pas significativement les performances par rapport aux marqueurs cliniques comme le Nottingham Prognostic Index (285). Cependant, dans le domaine des MICI il n'existe pas, à

l'heure actuelle, de prédicteur clinique suffisamment fort pour prédire la complication de la maladie ou la réponse au traitement.

Un point de vigilance dans la construction de score utilisant des données omiques concerne la prise en compte des facteurs d'ajustement cliniques. En 2009, Bøvelstad et al. ont comparé 7 méthodes (notamment PLS, RCP, ridge, Lasso) d'intégration de données cliniques et omiques sur 3 jeux de données dans le but de comparer les performances prédictives du modèle clinique, du modèle génomique et du modèle clinico-génomique (286). Ils ont conclu que la combinaison des covariables cliniques traditionnelles avec des données omiques peut conduire à de meilleures prédictions que celles obtenues en utilisant les données séparément. En 2019, Volkman et al. ont publié un plaidoyer pour la prise en compte des données cliniques lors de l'analyse de données omiques dans un cadre prédictif (287). A partir de deux jeux de données, ils ont montré que sur le premier jeu de données les données omiques n'apportaient pas de valeur prédictive ajoutée par rapport aux données cliniques et que sur le second les données omiques avaient une valeur ajoutée mais cependant uniquement pour les cas dans lesquels peu d'information clinique était inclue. Au final, l'inclusion de données cliniques permettait de sélectionner un plus faible nombre de prédicteurs omiques. Il est donc essentiel lors de la construction d'un score de réaliser un travail interdisciplinaire pour savoir quelles variables cliniques sont faciles à récolter sans trop de valeurs manquantes, quelles variables omiques sont facilement mesurables en routine. Il est également important de s'interroger sur la valeur ajoutée des données omiques par rapport à des variables cliniques, par exemple pour savoir si elles permettent de prendre en compte l'influence de l'environnement, d'une hygiène de vie ou de comorbidités. L'âge biologique mesuré à partir de données épigénomiques, par exemple données de méthylation (288) pourrait ainsi remplacer l'âge civil, notamment chez les personnes âgées, par ailleurs largement exclues des essais cliniques en raison de comorbidités. Récemment, Jardillier et al. ont montré, à partir de 16 jeux de données, que les profils tumoraux contenaient plus d'information pronostique que les variables cliniques telles que le stade pour de nombreux sous-types de cancers (289).

Dans ce contexte, il semble important de valider statistiquement l'apport des données omiques par rapport aux données cliniques, mais ceci est compliqué car le risque de sur-ajustement est élevé lors de l'analyse de données omiques. Boulesteix et al. ont proposé différentes méthodes de validation pour confirmer l'apport des données omiques (234). Notre

étude souffre de l'absence d'une validation externe, indispensable à son utilisation en pratique. On peut cependant noter que le nombre de données omiques candidates était relativement limité (~400), que les variants étudiés avaient été choisis a priori sur la base d'une revue de la littérature et que nous avons par ailleurs veillé à ne sélectionner qu'un petit nombre de variants afin de limiter le risque de sur-ajustement. Ainsi, le rapport du nombre de variables par rapport au nombre d'évènements est proche des recommandations (290).

L'analyse de données multi-omiques est plus avancée dans le domaine du cancer (5666 citations dans Pubmed pour la recherche « multi-omics » AND « cancer ») que dans les MICI (90 citations). Il s'agit cependant un domaine en plein essor allant de pair avec l'intérêt grandissant pour la médecine de précision dans les MICI. La complexité de ces maladies rend la tâche compliquée mais la justifie également car les données cliniques sont insuffisantes. La métabolomique et la protéomique pourraient permettre de dépasser les modèles cliniques et d'aboutir à des outils permettant de prédire la réponse au traitement. La recherche en ce sens a déjà porté ses fruits avec l'identification de marqueurs métagénomiques fécaux, métabolomiques et protéomiques dans le sérum qui prédisent une réponse aux thérapies anti-cytokines ou anti-intégrines dans les MICI (291). Le modèle utilisant uniquement les caractéristiques cliniques et les marqueurs inflammatoires sériques tels que la CRP n'était pas performant pour prédire la rémission à la semaine 14, avec une AUC de 0,62. En revanche, l'ajout des caractéristiques métagénomiques, métabolomiques ou protéomiques a considérablement augmenté la valeur prédictive avec des AUC de 0,85, 0,77 et 0,81, respectivement. Dans un sous-groupe de 21 participants atteints de MICI ayant des profils protéomiques, métabolomiques et métagénomiques disponibles, l'AUC du modèle avec des marqueurs cliniques et multi-omiques atteignait 0,96.

4. Portée des résultats

4.1. Pour la recherche

L'analyse des tendances temporelles des incidences des MICI permet d'orienter les pistes de recherche, notamment des facteurs environnementaux à leur origine puisque les changements dans les évolutions des incidences se produisent parallèlement aux changements dans l'environnement ou le mode de vie. Elle fournit également des indications sur les populations les plus touchées. Ainsi, nous avons montré que l'augmentation de l'incidence de la MC était plus marquée chez l'enfant et le jeune adulte ainsi que chez la femme dans la RCH. Ces augmentations suggèrent que non seulement les facteurs environnementaux potentiellement responsables des MICI persistent mais également qu'ils augmentent en nombre ou en intensité d'exposition.

L'augmentation de l'incidence des MICI chez les enfants et jeunes adultes est un point important pour la recherche des facteurs causaux. Au diagnostic de leur MICI, les enfants présentent plus fréquemment un phénotype purement inflammatoire que les patients plus âgés (83 % des enfants versus 73 %, 40 % et 67 % des 17-39 ans, 40-59 ans et 60 ans et plus, respectivement). Ce résultat appelle à i) poursuivre la recherche de facteurs environnementaux à l'origine de cette augmentation afin de mettre en place une prévention primaire ; ii) à développer des outils de médecine de précision spécifiquement chez l'enfant, permettant de cibler les patients nécessitant un traitement précoce afin d'éviter le développement de complications chez ces jeunes patients présentant une maladie inflammatoire au diagnostic. C'est ce que nous avons cherché à faire dans le chapitre 2.

L'analyse de la prévalence a montré que la prévalence augmentait fortement chez les personnes âgées avec quasiment un triplement de la prévalence chez les plus de 60 ans entre 2010 et 2030. Or, la prise en charge thérapeutique de cette population est souvent sous-optimisée en raison des comorbidités, des poly-médications, des effets secondaires plus sévères mais également en l'absence de données sur la sécurité et l'efficacité des traitements dans cette population âgée de plus de 60 ans. Une étude menée à partir du registre Epimad a montré que les patients âgés de plus de 60 ans étaient largement exposés aux 5-ASA et aux corticoïdes systémiques mais que très peu d'entre eux recevaient des traitements biologiques (292) : les probabilités cumulées de recours aux 5-ASA à 5 ans étaient de 77 % dans la MC et

78 % dans la RCH, de 45 % dans la MC et 34 % dans la RCH pour l'exposition aux corticoïdes, de 18 % dans la MC et 10 % dans la RCH pour l'exposition aux immunosuppresseurs et de 5 % dans la MC pour l'exposition aux anti-TNF (seuls 4 patients atteints de RCH avaient reçu de l'infliximab). A titre de comparaison, chez l'enfant les probabilités cumulées à 5 ans était de 50 % et 16 % pour les immunosuppresseurs et les anti-TNF, respectivement (293). Ceci souligne l'importance de s'intéresser davantage à cette population dans la recherche des facteurs de maladie agressive mais également dans les essais cliniques desquels les personnes âgées sont exclues. L'objectif est de déterminer des schémas de traitement optimaux pour cette population de patients de plus de 60 ans souffrant de MICI (294). Ce sujet du traitement des personnes âgées est listé dans les besoins non satisfaits dans les MICI publiée par Revés et al (92).

Dans cet article sur les besoins non satisfaits dans la recherche sur les MICI, la nécessité d'une meilleure stratification des patients selon leur risque d'évolution péjorative de leur maladie est cité en premier (92). PREDICT-EPIMAD répond à ce besoin non satisfait pour la MC à début pédiatrique, mais ce score nécessite une validation externe.

La restauration de la qualité de vie fait également partie des besoins non satisfaits dans les MICI. C'est un objectif à la fois des essais cliniques (critère de jugement) mais également une cible à atteindre avec les traitements donnés aux patients. Nos travaux montrent que les patients dont la MICI a débuté dans l'enfance, s'insèrent bien dans la vie active malgré des difficultés importantes rencontrées dans le déroulement de leurs études ou le choix de leur profession. Ils étaient près de 80 % à avoir une qualité de vie « normale » ou « élevée » soulignant les capacités d'adaptation dont les adolescents ont su faire preuve pour arriver à ce résultat à l'âge adulte. Cependant, la qualité de vie était associée au niveau d'études obtenu. A titre de comparaison, dans l'étude BIRD, ils étaient 47 % à présenter une qualité de vie normale et aucun ne présentait une qualité de vie élevée (175).

Enfin, la nécessité d'analyse conjointe de données cliniques et génétiques m'a amenée à m'interroger sur les méthodes statistiques optimales à utiliser. Les premiers travaux, présentés dans ce mémoire, ont permis de dégager quelques premières pistes pour l'analyse de ces données en grand nombre mais ces résultats doivent être enrichis de nouvelles analyses. Ces travaux méthodologiques ont permis d'orienter le choix de méthodes d'analyse pour la création du score PREDICT-EPIMAD et la poursuite de ces travaux pourrait permettre

de produire des recommandations concrètes pour l'analyse de ces données. Depuis nos travaux, de nouvelles méthodes ont été proposées et/ou comparées, mais ces travaux portaient essentiellement sur les analyses multi-omiques sans données cliniques. Par exemple, une comparaison récente de 13 méthodes non-supervisées d'intégration de données multi-omiques a été mise en œuvre et a conclu que MoCluster était l'une des meilleures méthodes pour effectuer la classification tout en révélant des biomarqueurs candidats associés aux sous-groupes (282). Pierre-Jean et al. ont également récemment proposé une nouvelle méthode PIntMF (Penalized Integrative Matrix Factorization), méthode reposant sur un modèle de factorisation de matrice incluant des contraintes Lasso pour la sélection de variables (295). Ces méthodes pourraient être intéressantes à tester en combinaison avec des données cliniques.

4.2. Pour la prise de décision partagée

Le concept de prise de décision partagée entre le médecin et son patient a gagné beaucoup d'intérêt dans la pratique clinique mais aussi dans la littérature, en parallèle de l'augmentation de l'arsenal thérapeutique au cours de ces dernières décennies. Les différentes options de traitements présentent des avantages et des inconvénients, et les préférences des patients peuvent également différer. Si les patients sont associés à la prise de décision médicale, leurs adhésion, observance et satisfaction peuvent être améliorées, ce qui pourrait conduire à de meilleurs résultats thérapeutiques (296). La prise en considération de l'impact de la maladie et de son traitement sur la qualité de vie fait désormais partie de la prise en charge standard (88). Plusieurs études ont montré que les patients souhaitaient être impliqués dans la prise de décision thérapeutique (297,298).

Il a été montré, dans une vaste étude multicentrique française du GETAID (Groupe d'Étude Thérapeutique des Affections Inflammatoires du Tube Digestif), que 75 % des patients avaient des inquiétudes au sujet des traitements (299). Les préoccupations principales étaient liées au risque d'événements indésirables et à l'efficacité des traitements. D'autres préoccupations concernaient leur mode d'administration. L'amélioration de la qualité de l'information et la mise en œuvre de processus de prise de décision partagée sont donc importants chez les patients atteints de MICI.

En ce sens, la mise à disposition d'un outil simple permettant d'estimer le risque de complication de la maladie, pourra aider le gastroentérologue et le patient dans la prise de décision en fournissant un élément supplémentaire, à analyser en regard d'autres éléments disponibles (qualité de vie, préférence du patient, effets secondaires…). En particulier chez l'enfant, l'impact de la maladie implique également les parents. Une étude a montré que les parents se montraient plus inquiets des effets secondaires des traitements que des symptômes de la maladie (80 % versus 73 %) (300).

4.3. Pour la santé publique

Les résultats de cette thèse ont aussi une portée en santé publique.

Tout d'abord, les résultats d'incidence et de prévalence sont utiles dans un cadre de planification des soins. Ils permettent de fournir des données pour l'anticipation des besoins, en l'occurrence croissants. Un point important est l'anticipation d'une augmentation du nombre de cas particulièrement marquée dans la population âgée, dont la prise en charge doit être adaptée.

Les résultats de l'étude sur le niveau d'études et l'insertion professionnelle des patients débutant leur maladie dans l'enfance sont positifs et démontrent une réelle capacité d'adaptation des enfants. Cependant, les patients mentionnant des difficultés pendant leurs études avaient un niveau d'éducation plus bas, soulignant la nécessité de traiter les enfants et les adolescents atteints de MICI avec une approche holistique, médicalement, pour minimiser le nombre de rechutes, et socialement, pour leur permettre de suivre un programme scolaire aussi normal que possible.

Enfin, l'utilisation d'un score prédictif de complication de la maladie permettrait de mieux traiter les patients, d'éviter de traiter des patients à tort permettant une réduction des couts des traitements et de leurs effets secondaires.

5. Perspectives

PRINCIPALES PERSPECTIVES DU REGISTRE EPIMAD

Deux principales perspectives d'envergure sont à noter. Premièrement, le registre Epimad déposera à l'automne un dossier de demande de chainage des données du registre aux

données du SNDS. Cet enrichissement des données permettra d'obtenir pour un vaste échantillon de patients à la fois des données phénotypiques bien caractérisées et des données de suivi concernant les traitements, les hospitalisations, les chirurgies, permettant de mener des études sur l'histoire naturelle des MICI et notamment d'étudier l'efficacité des traitement ou l'impact des délais d'initiation des traitements par biothérapie sur l'histoire naturelle des MICI. Des recueils complémentaires de données phénotypiques au cours du suivi pourraient également être réalisés car ces données ne sont pas disponibles dans les bases de données du SNDS.

Par ailleurs, une biobanque de patients atteints de MICI à début pédiatrique (moins de 20 ans au moment du diagnostic) est en cours de mise en place. Des échantillons de sang et de selles seront collectés au diagnostic et à 1 an de suivi, et permettront d'identifier des profils métagénomiques, transcriptomiques et métabolomiques particuliers, associés à l'évolution de la maladie et à la réponse au traitement.

PATIENTS REPORTED OUTCOMES

Concernant les outils de mesure de qualité de vie et de handicap fonctionnel, nos travaux se poursuivent par :

- Une étude sur la qualité de vie utilisant le questionnaire EQ-5D-DL (301) :

Le rôle croissant des études d'évaluation médico-économique des actions de soins nécessite la connaissance de la qualité de vie réelle des patients, mesurée avec des questionnaires standards permettant le calcul d'années de vie pondérées par la qualité (Quality Adjusted Life Years, ou QALY). Pour ce faire, le questionnaire générique EQ-5D est la norme mais les données sur utilisation dans les MICI sont limitées. Nous avons participé à une enquête prospective et transversale auprès de patients français (n=2841) atteints de MICI âgés de 18 ans ou plus au moment de l'étude, provenant de trois bases de données : le registre EPIMAD, la base de données SUVIMIC (dossier médical informatisé du service de gastro-entérologie de l'Hôpital Saint Antoine) et une enquête en ligne de l'AFA (association de patients atteints de MICI). L'objectif de cette étude était d'estimer les valeurs d'utilité (score de qualité de vie permettant de pondéré les années de vie pour calculer les QALY) pour les patients atteints de MICI sur un large échantillon de patients, en analysant la relation avec des indices spécifiques de qualité de vie (SIBDQ) et l'activité de la maladie, et à identifier les principaux facteurs

influençant la qualité de vie des patients. L'article, dont je suis premier auteur, a été soumis à Quality of Life research.

- L'analyse de la sensibilité au changement de l'IBD-DI :

Comme évoqué en introduction du chapitre 2, nous avions précédemment étudié les caractéristiques psychométriques du questionnaire de handicap fonctionnel à partir d'un échantillon de 200 patients du registre Epimad (174). Cette étude incluait la validité factorielle, la consistance externe, la consistance interne ainsi que la reproductibilité inter- et intra-observateur. Ces mêmes 200 patients ont été recontactés entre novembre 2019 et juin 2021 afin d'étudier la sensibilité au changement de l'IBD-DI. L'analyse des données est en cours et mènera à la rédaction d'un article dont je serai premier auteur. Les patients seront classés en trois groupes en fonction de l'activité de leur maladie soient : patients stables, en amélioration, en détérioration. La sensibilité au changement sera évaluée en observant l'évolution des scores IBD-DI des patients dans ces 3 groupes. La différence minimale cliniquement pertinente, utile pour l'utilisation de ce score dans des essais thérapeutiques, sera calculée à partir du changement dans l'activité de la maladie (utilisation de courbes ROC à partir des 3 groupes de patients précédemment définis), ainsi que par rapport à un effet de taille de 0,5 (302).

FACTEURS ENVIRONNEMENTAUX

Enfin, le registre Epimad est impliqué depuis de nombreuses années dans la recherche sur les facteurs environnementaux.

Premièrement, le projet HEROIC a été initié par Corinne Gower-Rousseau. Il s'agit d'un consortium multidisciplinaire impliquant cliniciens, épidémiologistes, historiens, sociologues, mathématiciens et toxicologues dont l'objectif est l'identification de nouveaux facteurs environnementaux de la MC dans le but d'ouvrir de nouvelles pistes de prévention. Le projet HEROIC a débuté à plusieurs niveaux depuis 2018 :

- Analyse de la répartition spatio-temporelle de l'incidence des MICI dans le nord-ouest de la France : détermination de clusters de sur et de sous incidence de la MC et de la RCH.
- Etude ANR CROPS (CROhn disease and Pollution of Soils). Cette étude avait pour objectif de caractériser les clusters de sur- et de sous-incidence à partir de données

environnementales de bases de données existantes, de données historiques et de prélèvements de sols. Cette étude a fait l'objet d'un financement de l'ANR PRTS en 2020 et de 3 thèses d'Université dont les résultats définitifs seront connus et publiés à l'automne 2024.

- L'étude DENTACROHN qui devrait débuter à l'automne 2024 : Cette étude pilote a pour objectif d'étudier l'exposition environnementale dans les 10-15 années précédant le diagnostic de MC par une approche originale de mesure de l'exposome dentaire. Vingt patients atteints de MC à début pédiatrique (entre 5 et 17 ans au moment du diagnostic) et 20 témoins appariés sur l'âge et le sexe et ne présentant pas de MC seront inclus. L'analyse des biomarqueurs dans les dents (exposome dentaire) est une approche unique et non invasive pour évaluer rétrospectivement les expositions et retracer ces expositions dans le temps. Au cours du développement de la dent (dès la grossesse pour les dents de lait), des composés organiques et des métaux circulants dans le corps sont capturés par la dent en formation dans des strates de dentine. L'analyse dans les dents des expositions à différents composés (métaux toxiques, nutriments essentiels, polluants organiques etc.), rendue possible par des méthodes d'amplification, permet ainsi de retracer l'exposition pendant les 10-15 premières années de vie (303,304). A titre d'exemple, ce type d'analyse a suggéré que des niveaux plus élevés de métaux toxiques (Plomb) ainsi que des niveaux plus faibles d'éléments essentiels (Zinc et Manganèse) au cours de fenêtres de développement spécifiques augmentaient le risque et la gravité des troubles du spectre autistique (305). A notre connaissance, il n'existe qu'une seule étude de très petite taille sur les MICI (7 patients atteints de MC, 5 atteints de RCH et 16 témoins) et l'exposome dentaire (154). Cette étude a révélé une différence dans les niveaux de Plomb, Zinc, Cuivre et Chrome pendant la vie intra-utérine et le début de la vie postnatale chez les individus qui ont développé ensuite une MICI.

A terme, les résultats de ces différentes études préliminaires pourraient mener à une étude interventionnelle de type cas-témoin.

Deuxièmement, le registre Epimad est fortement impliqué dans l'étude Mikinautes, portée par le Pr Jean-Pierre Hugot (APHP) et le Dr Corinne Gower-Rousseau et financée par l'afa Crohn RCH. Mikinautes est un projet multipartenaire impliquant le GETAID pédiatrique,

EPIMAD et l'afa Crohn RCH. Comme cela a été décrit en introduction, l'alimentation pourrait jouer un rôle dans le déclenchement des MICI. Cependant, peu de travaux se sont intéressés au lien entre alimentation et rechutes. L'objectif de l'étude Mikinautes est d'identifier les facteurs de risque alimentaires, présents en période de rémission et associés à l'apparition des rechutes par le suivi prospectif d'une e-cohorte d'adultes et d'adolescents atteints de MICI. Le registre Epimad est impliqué dans cette étude à plusieurs degrés : i) inclusion de patients ; ii) implication dans le comité exécutif de l'étude et ; iii) réalisation des analyses statistiques. Les premiers résultats devraient être connus et publiés en fin d'année 2024.

Conclusions

Cette thèse s'est intéressée à différents aspects des MICI, de l'incidence aux complications de la maladie et à son impact sur la vie des patients. On retiendra que l'incidence continue d'augmenter, particulièrement chez l'enfant et le jeune adulte. La prévalence augmente également et cette augmentation est particulièrement marquée chez les personnes plus âgées. Ceci n'est pas sans conséquence sur la prise en charge des patients : i) les MICI sont souvent plus agressives chez l'enfant et ont par ailleurs un impact sur le long terme tout au long de la vie ; ii) la prise en charge des personnes âgées est spécifique (comorbidités, poly-médications etc) et cette population a fait l'objet de moins d'études dans les essais cliniques, du fait de l'exclusion fréquente dans ces essais des patients de plus de 60 ans.

La population pédiatrique nous a particulièrement intéressés dans cette thèse. Tout d'abord, nous nous sommes intéressés à l'impact de la maladie sur la poursuite des études et l'insertion professionnelle des patients atteints de MICI pendant l'enfance ou l'adolescence. Malgré un impact certain de la maladie sur le déroulement des études et le choix de la profession, les patients atteints de MICI dans l'enfance font preuve de capacités d'adaptation telles qu'au final leur niveau d'étude est supérieur à celui de la population générale du même âge et leur taux de chômage plus faible. Un résultat nouveau de ce travail est que les patients développent un attrait tout particulier pour le domaine de la santé et du soin du fait de leur parcours médical. Enfin ils travaillent plus fréquemment dans la fonction publique dont l'attrait réside sûrement dans la sécurité de l'emploi et la possibilité d'arrêts maladie. Ensuite, nous avons développé un score de prédiction de la maladie de Crohn compliquée incluant 8 variables cliniques, sérologiques et génétiques chez l'enfant qui pourrait permettre, une fois validé, de mieux cibler les patients nécessitant un traitement précoce.

L'ensemble de ces résultats montre une partie de l'éventail des possibilités d'études observationnelles et interventionnelles grâce aux registres épidémiologiques en population générale.

Afin de pouvoir analyser les données du registre EPIMAD, nous nous sommes également intéressés aux méthodes d'intégration de données cliniques et omiques et avons amorcé des travaux méthodologiques en ce sens. L'intégration des données omiques est un domaine

relativement récent qui a connu un essor grâce à l'arrivée des technologies de séquençage haut débit. Ce domaine est en constante évolution, avec de nouvelles méthodes de séquençage et de méthodes pour l'analyse et l'intégration de ces données. Malgré le grand nombre de méthodes disponibles, l'analyse de ces données demeure un sujet de recherche d'actualité. Les défis liés à la grande dimension et à l'hétérogénéité des données risquent de s'intensifier avec l'avènement de technologies encore plus performantes et l'augmentation continue de la masse de données. Les données de protéomique et de métabolomique notamment font l'objet d'un intérêt grandissant pour la médecine de précision et posent de nouveaux défis du fait de leur complexité et de leur variabilité.

En conclusion, les MICI demeurent une question de santé publique importante et leur poids ne fait qu'augmenter par leur fréquence, l'âge jeune des patients qui vont vivre toute leur vie avec leur MICI et le coût des nouvelles thérapeutiques mises sur le marché (biothérapies). L'évolution temporelle des incidences de MICI en fonction du genre et de l'âge diagnostique des patients est un facteur à prendre en compte pour la compréhension de la pathogénèse des MICI. Le poids des MICI augmentant, nous devons préparer notre système de santé et continuer à travailler dans la recherche de méthodes de stratification des patients selon leur risque de complication et selon leur réponse au traitement.

Références

1. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. Lancet. 2017;389(10080):1741–55.
2. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. 2017;389(10080):1756–70.
3. Silverberg MS, Satsangi J, Ahmad T, Arnott IDR, Bernstein CN, Brant SR, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol. 2005;19 Suppl A:5A–36A.
4. Levine A, Griffiths A, Markowitz J, Wilson DC, Turner D, Russell RK, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. Inflamm Bowel Dis. 2011;17(6):1314–21.
5. Solberg IC, Vatn MH, Høie O, Stray N, Sauar J, Jahnsen J, et al. Clinical course in Crohn's disease: results of a Norwegian population-based ten-year follow-up study. Clin Gastroenterol Hepatol. 2007;5(12):1430–8.
6. Solberg IC, Lygren I, Jahnsen J, Aadland E, Høie O, Cvancarova M, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study). Scand J Gastroenterol. 2009;44(4):431–40.
7. Pariente B, Cosnes J, Danese S, Sandborn WJ, Lewin M, Fletcher JG, et al. Development of the Crohn's disease digestive damage score, the Lémann score. Inflamm Bowel Dis. 2011;17(6):1415–22.
8. Pariente B, Mary JY, Danese S, Chowers Y, De Cruz P, D'Haens G, et al. Development of the Lémann index to assess digestive tract damage in patients with Crohn's disease. Gastroenterology. 2015;148(1):52–63.e3.
9. Pariente B, Torres J, Burisch J, Arebi N, Barberio B, Duricova D, et al. Validation and Update of the Lémann Index to Measure Cumulative Structural Bowel Damage in Crohn's Disease. Gastroenterology. 2021;161(3):853–864.e13.
10. Gilletta C, Lewin M, Bourrier A, Nion-Larmurier I, Rajca S, Beaugerie L, et al. Changes in the Lémann Index Values During the First Years of Crohn's Disease. Clin Gastroenterol Hepatol. 2015;13(9):1633–1640.e3.
11. Cosnes J, Cattan S, Blain A, Beaugerie L, Carbonnel F, Parc R, et al. Long-term evolution of disease behavior of Crohn's disease. Inflamm Bowel Dis. 2002;8(4):244–50.
12. Louis E, Collard A, Oger AF, Degroote E, Aboul Nasr El Yafi FA, Belaiche J. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. Gut. 2001;49(6):777–82.
13. Peyrin-Biroulet L, Loftus EV, Colombel JF, Sandborn WJ. The natural history of adult Crohn's disease in population-based cohorts. Am J Gastroenterol. 2010;105(2):289–97.
14. Thia KT, Sandborn WJ, Harmsen WS, Zinsmeister AR, Loftus EV. Risk Factors Associated With Progression to Intestinal Complications of Crohn's Disease in a Population-Based Cohort. Gastroenterology. 2010;139(4):1147–55.
15. Torres J, Billioud V, Sachar DB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis as a progressive disease: the forgotten evidence. Inflamm Bowel Dis. 2012;18(7):1356–63.
16. Fumery M, Singh S, Dulai PS, Gower-Rousseau C, Peyrin-Biroulet L, Sandborn WJ. Natural History of Adult Ulcerative Colitis in Population-based Cohorts: A Systematic Review. Clin Gastroenterol Hepatol. 2018;16(3):343–356.e3.

17. Cleveland NK, Torres J, Rubin DT. What Does Disease Progression Look Like in Ulcerative Colitis, and How Might It Be Prevented? Gastroenterology. 2022;162(5):1396–408.

18. Le Berre C, Ananthakrishnan AN, Danese S, Singh S, Peyrin-Biroulet L. Ulcerative Colitis and Crohn's Disease Have Similar Burden and Goals for Treatment. Clin Gastroenterol Hepatol. 2020;18(1):14–23.

19. Roda G, Narula N, Pinotti R, Skamnelos A, Katsanos KH, Ungaro R, et al. Systematic review with meta-analysis: proximal disease extension in limited ulcerative colitis. Aliment Pharmacol Ther. 2017;45(12):1481–92.

20. Laurain PA, Guillo L, D'Amico F, Netter P, Danese S, Baumann C, et al. Incidence of and Risk Factors for Colorectal Strictures in Ulcerative Colitis: A Multicenter Study. Clin Gastroenterol Hepatol. 2021;19(9):1899–1905.e1.

21. Duricova D, Pedersen N, Elkjaer M, Gamborg M, Munkholm P, Jess T. Overall and cause-specific mortality in Crohn's disease: a meta-analysis of population-based studies. Inflamm Bowel Dis. 2010;16(2):347–53.

22. Jess T, Gamborg M, Munkholm P, Sørensen TIA. Overall and cause-specific mortality in ulcerative colitis: meta-analysis of population-based inception cohort studies. Am J Gastroenterol. 2007;102(3):609–17.

23. Ekbom A, Helmick C, Zack M, Adami HO. Ulcerative colitis and colorectal cancer. A population-based study. N Engl J Med. 1990;323(18):1228–33.

24. Gros B, Kaplan GG. Ulcerative Colitis in Adults: A Review. JAMA. 2023;330(10):951–65.

25. Lutgens MWMD, van Oijen MGH, van der Heijden GJMG, Vleggaar FP, Siersema PD, Oldenburg B. Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies. Inflamm Bowel Dis. 2013;19(4):789–99.

26. Munkholm P. Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. Aliment Pharmacol Ther. 2003;18 Suppl 2:1–5.

27. Ngu JH, Gearry RB, Wright AJ, Stedman CAM. Inflammatory bowel disease is associated with poor outcomes of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol. 2011;9(12):1092–7.

28. White H. A Discussion on "Ulcerative Colitis." Introductroy Address. Proc R Soc Med. 1909;2(Med Sect):79–82.

29. Crohn BB, Ginzburg L, Oppenheimer GD. Regional Ileitis: A Pathologic and Clinical Entity. JAMA. 1932;99:1323–9.

30. Allchin WH. A Discussion on "Ulcerative Colitis.": Introductory Address. Proc R Soc Med. 1909;2(Med Sect):59–75.

31. Lockhart-Mummery HE, Morson BC. Crohn's disease (regional enteritis) of the large intestine and its distinction from ulcerative colitis. Gut. 1960;1(2):87–105.

32. Price AB. Overlap in the spectrum of non-specific inflammatory bowel disease--'colitis indeterminate'. J Clin Pathol. 1978;31(6):567–77.

33. Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. Nat Rev Gastroenterol Hepatol. 2021;18(1):56–66.

34. GBD 2017 Inflammatory Bowel Disease Collaborators. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol. 2020;5(1):17–30.

35. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2017;390(10114):2769–78.

36. Torabi M, Bernstein CN, Yu BN, Wickramasinghe L, Blanchard JF, Singh H. Geographical Variation and Factors Associated With Inflammatory Bowel Disease in a Central Canadian Province. Inflamm Bowel Dis. 2020;26(4):581–90.

37. Forss A, Clements M, Bergman D, Roelstraete B, Kaplan GG, Myrelid P, et al. A nationwide cohort study of the incidence of inflammatory bowel disease in Sweden from 1990 to 2014. Aliment Pharmacol Ther. 2022;55(6):691–9.

38. Shivashankar R, Tremaine WJ, Harmsen WS, Loftus EV. Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. Clin Gastroenterol Hepatol. 2017;15(6):857–63.

39. Coward S, Benchimol EI, Kuenzig ME, Windsor JW, Bernstein CN, Bitton A, et al. The 2023 Impact of Inflammatory Bowel Disease in Canada: Epidemiology of IBD. J Can Assoc Gastroenterol. 2023;6(Suppl 2):S9–15.

40. Forbes AJ, Frampton CMA, Day AS, Kaplan GG, Gearry RB. The Epidemiology of Inflammatory Bowel Disease in Oceania: A Systematic Review and Meta-Analysis of Incidence and Prevalence. Inflamm Bowel Dis. 2023;izad295.

41. Zhao M, Gönczi L, Lakatos PL, Burisch J. The burden of inflammatory bowel disease in Europe in 2020. J Crohns Colitis. 2021;15(9):1573–87.

42. Burisch J, Jess T, Martinato M, Lakatos PL, ECCO -EpiCom. The burden of inflammatory bowel disease in Europe. J Crohns Colitis. 2013;7(4):322–37.

43. de Groof EJ, Rossen NGM, van Rhijn BD, Karregat EPM, Boonstra K, Hageman I, et al. Burden of disease and increasing prevalence of inflammatory bowel disease in a population-based cohort in the Netherlands. Eur J Gastroenterol Hepatol. 2016;28(9):1065–72.

44. Hammer T, Nielsen KR, Munkholm P, Burisch J, Lynge E. The Faroese IBD Study: Incidence of Inflammatory Bowel Diseases Across 54 Years of Population-based Data. J Crohns Colitis. 2016;10(8):934–42.

45. Nielsen KR, Midjord J, Nymand Lophaven S, Langholz E, Hammer T, Burisch J. The Incidence and Prevalence of Inflammatory Bowel Disease Continues to Increase in the Faroe Islands - A Cohort Study from 1960 to 2020. J Crohns Colitis. 2024;18(2):308–19.

46. Burisch J, Pedersen N, Čuković-Čavka S, Brinar M, Kaimakliotis I, Duricova D, et al. East-West gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. Gut. 2014;63(4):588–97.

47. Shivananda S, Lennard-Jones J, Logan R, Fear N, Price A, Carpenter L, et al. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). Gut. 1996;39(5):690–7.

48. Chaaro-Benallal D, Guerra-Veloz MF, Argüelles-Arias F, Benítez JM, Perea-Amarillo R, Iglesias E, et al. Evolution of the incidence of inflammatory bowel disease in Southern Spain. Rev Esp Enfermedades Dig. 2017;109(11):757–60.

49. Agrawal M, Christensen HS, Bøgsted M, Colombel JF, Jess T, Allin KH. The Rising Burden of Inflammatory Bowel Disease in Denmark Over Two Decades: A Nationwide Cohort Study. Gastroenterology. 2022;163(6):1547–1554.e5.

50. Chouraki V, Savoye G, Dauchet L, Vernier-Massouille G, Dupas JL, Merle V, et al. The changing pattern of Crohn's disease incidence in northern France: a continuing increase in the 10- to 19-year-old age bracket (1988-2007). Aliment Pharmacol Ther. 2011;33(10):1133–42.

51. Molinié F, Gower-Rousseau C, Yzet T, Merle V, Grandbastien B, Marti R, et al. Opposite evolution in incidence of Crohn's disease and ulcerative colitis in Northern France (1988-1999). Gut. 2004;53(6):843–8.

52. Auvin S, Molinié F, Gower-Rousseau C, Brazier F, Merle V, Grandbastien B, et al. Incidence, clinical presentation and location at diagnosis of pediatric inflammatory bowel disease: a prospective population-based study in northern France (1988-1999). J Pediatr Gastroenterol Nutr. 2005;41(1):49–55.

53. Gower-Rousseau C, Salomez JL, Dupas JL, Marti R, Nuttens MC, Votte A, et al. Incidence of inflammatory bowel disease in northern France (1988-1990). Gut. 1994;35(10):1433–8.

54. Ghione S, Sarter H, Fumery M, Armengol-Debeir L, Savoye G, Ley D, et al. Dramatic increase in incidence of ulcerative colitis and Crohn's disease (1988-2011): a population-based study of french adolescents. Am J Gastroenterol. 2018;113(2):265–72.

55. Gower-Rousseau C, Leroyer A, Génin, Michaël, Savoye G, Sarter H, Pariente B, et al. Épidémiologie descriptive et évolution dans le temps et l'espace de l'incidence des maladies inflammatoires chroniques intestinales dans le nord-ouest de la France (1988-2014). Bull Epidémiologique Hebd. 2019;13:228–36.

56. Latour P, Louis E, Belaiche J. Incidence of inflammatory bowel disease in the area of Liège: a 3 years prospective study (1993-1996). Acta Gastro-Enterol Belg. 1998;61(4):410–3.

57. Nerich V, Monnet E, Etienne A, Louafi S, Ramée C, Rican S, et al. Geographical variations of inflammatory bowel disease in France: a study based on national health insurance data. Inflamm Bowel Dis. 2006;12(3):218–26.

58. Genin M, Fumery M, Occelli F, Savoye G, Pariente B, Dauchet L, et al. Fine-scale geographical distribution and ecological risk factors for Crohn's disease in France (2007-2014). Aliment Pharmacol Ther. 2020;51(1):139–48.

59. Genin M, Duhamel A, Preda C, Fumery M, Savoye G, Peyrin-Biroulet L, et al. Space-time clusters of Crohn's disease in northern France. J Public Health. 2013;21(6):497–504.

60. Ng SC, Kaplan GG, Tang W, Banerjee R, Adigopula B, Underwood FE, et al. Population Density and Risk of Inflammatory Bowel Disease: A Prospective Population-Based Study in 13 Countries or Regions in Asia-Pacific. Am J Gastroenterol. 2019;114(1):107–15.

61. Kotze PG, Underwood FE, Damião AOMC, Ferraz JGP, Saad-Hossne R, Toro M, et al. Progression of Inflammatory Bowel Diseases Throughout Latin America and the Caribbean: A Systematic Review. Clin Gastroenterol Hepatol. 2020;18(2):304–12.

62. Parente JML, Coy CSR, Campelo V, Parente MPPD, Costa LA, da Silva RM, et al. Inflammatory bowel disease in an underdeveloped region of Northeastern Brazil. World J Gastroenterol. 2015;21(4):1197–206.

63. Wei SC, Lin MH, Tung CC, Weng MT, Kuo JS, Shieh MJ, et al. A nationwide population-based study of the inflammatory bowel diseases between 1998 and 2008 in Taiwan. BMC Gastroenterol. 2013;13:166.

64. Lewis JD, Parlett LE, Jonsson Funk ML, Brensinger C, Pate V, Wu Q, et al. Incidence, Prevalence, and Racial and Ethnic Distribution of Inflammatory Bowel Disease in the United States. Gastroenterology. 2023;165(5):1197–1205.e2.

65. Weisman MH, Oleg Stens null, Seok Kim H, Hou JK, Miller FW, Dillon CF. Inflammatory Bowel Disease Prevalence: Surveillance data from the U.S. National Health and Nutrition Examination Survey. Prev Med Rep. 2023;33:102173.

66. Coward S, Clement F, Benchimol EI, Bernstein CN, Avina-Zubieta JA, Bitton A, et al. Past and future burden of inflammatory bowel diseases based on modeling of population-based data. Gastroenterology. 2019;156(5):1345–1353.e4.

67. Jones GR, Lyons M, Plevris N, Jenkinson PW, Bisset C, Burgess C, et al. IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. Gut. 2019;68(11):1953–60.

68. Lirhus SS, Høivik ML, Moum B, Anisdahl K, Melberg HO. Incidence and Prevalence of Inflammatory Bowel Disease in Norway and the Impact of Different Case Definitions: A Nationwide Registry Study. Clin Epidemiol. 2021;13:287–94.

69. Larsen L, Karachalia Sandri A, Fallingborg J, Jacobsen BA, Jacobsen HA, Bøgsted M, et al. Has the Incidence of Inflammatory Bowel Disease Peaked? Evidence From the Population-Based NorDIBD Cohort 1978-2020. Am J Gastroenterol. 2023;118(3):501–10.

70. Ruel J, Ruane D, Mehandru S, Gower-Rousseau C, Colombel JF. IBD across the age spectrum: is it the same disease? Nat Rev Gastroenterol Hepatol. 2014;11(2):88–98.

71. Kuenzig ME, Fung SG, Marderfeld L, Mak JWY, Kaplan GG, Ng SC, et al. Twenty-first century trends in the global epidemiology of pediatric-onset inflammatory bowel disease: systematic review. Gastroenterology. 2022;162(4):1147–1159.e4.

72. Benchimol EI, Bernstein CN, Bitton A, Carroll MW, Singh H, Otley AR, et al. Trends in Epidemiology of Pediatric Inflammatory Bowel Disease in Canada: Distributed Network Analysis of Multiple Population-Based Provincial Health Administrative Databases. Am J Gastroenterol. 2017;112(7):1120–34.

73. Kanof ME, Lake AM, Bayless TM. Decreased height velocity in children and adolescents before the diagnosis of Crohn's disease. Gastroenterology. 1988;95(6):1523–7.

74. Ley D, Duhamel A, Behal H, Vasseur F, Sarter H, Michaud L, et al. Growth Pattern in Paediatric Crohn Disease Is Related to Inflammatory Status. J Pediatr Gastroenterol Nutr. 2016;63(6):637–43.

75. Fumery M, Duricova D, Gower-Rousseau C, Annese V, Peyrin-Biroulet L, Lakatos PL. Review article: the natural history of paediatric-onset ulcerative colitis in population-based studies. Aliment Pharmacol Ther. 2016;43(3):346–55.

76. Duricova D, Fumery M, Annese V, Lakatos PL, Peyrin-Biroulet L, Gower-Rousseau C. The natural history of Crohn's disease in children: a review of population-based studies. Eur J Gastroenterol Hepatol. 2017;29(2):125–34.

77. Knowles SR, Graff LA, Wilding H, Hewitt C, Keefer L, Mikocka-Walus A. Quality of Life in Inflammatory Bowel Disease: A Systematic Review and Meta-analyses-Part I. Inflamm Bowel Dis. 2018;24(4):742–51.

78. Knowles SR, Keefer L, Wilding H, Hewitt C, Graff LA, Mikocka-Walus A. Quality of Life in Inflammatory Bowel Disease: A Systematic Review and Meta-analyses-Part II. Inflamm Bowel Dis. 2018;24(5):966–76.

79. Peyrin-Biroulet L, Reinisch W, Colombel JF, Mantzaris GJ, Kornbluth A, Diamond R, et al. Clinical disease activity, C-reactive protein normalisation and mucosal healing in Crohn's disease in the SONIC trial. Gut. 2014;63(1):88–95.

80. Colombel JF, Keir ME, Scherl A, Zhao R, de Hertogh G, Faubion WA, et al. Discrepancies between patient-reported outcomes, and endoscopic and histological appearance in UC. Gut. 2017;66(12):2063–8.

81. Bossuyt P, Baert F, D'Heygere F, Nakad A, Reenaers C, Fontaine F, et al. Early Mucosal Healing Predicts Favorable Outcomes in Patients With Moderate to Severe Ulcerative Colitis Treated With Golimumab: Data From the Real-life BE-SMART Cohort. Inflamm Bowel Dis. 2019;25(1):156–62.

82. Colombel JF, Rutgeerts P, Reinisch W, Esser D, Wang Y, Lang Y, et al. Early mucosal healing with infliximab is associated with improved long-term clinical outcomes in ulcerative colitis. Gastroenterology. 2011;141(4):1194–201.

83. Colombel JF, Sandborn WJ, Reinisch W, Mantzaris GJ, Kornbluth A, Rachmilewitz D, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. N Engl J Med. 2010;362(15):1383–95.

84. D'Haens G, Baert F, van Assche G, Caenepeel P, Vergauwe P, Tuynman H, et al. Early combined immunosuppression or conventional management in patients with newly diagnosed Crohn's disease: an open randomised trial. Lancet. 2008;371(9613):660–7.

85. Danese S, Fiorino G, Peyrin-Biroulet L. Early intervention in Crohn's disease: towards disease modification trials. Gut. 2017;66(12):2179–87.

86. Revés J, Mascarenhas A, José Temido M, Morão B, Neto Nascimento C, Rita Franco A, et al. Early intervention with biologic therapy in Crohn´s disease: how early is early? J Crohns Colitis. 2023;17(11):1752–60.

87. Bouguen G, Levesque BG, Feagan BG, Kavanaugh A, Peyrin-Biroulet L, Colombel JF, et al. Treat to target: a proposed new paradigm for the management of Crohn's disease. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2015;13(6):1042–1050.e2.

88. Turner D, Ricciuto A, Lewis A, D'Amico F, Dhaliwal J, Griffiths AM, et al. STRIDE-II: An Update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) Initiative of the International Organization for the Study of IBD (IOIBD): Determining Therapeutic Goals for Treat-to-Target strategies in IBD. Gastroenterology. 2021;160(5):1570–83.

89. Ley D, Leroyer A, Dupont C, Sarter H, Bertrand V, Spyckerelle C, et al. New Therapeutic Strategies Have Changed the Natural History of Pediatric Crohn's Disease: A Two-Decade Population-Based Study. Clin Gastroenterol Hepatol. 2022;20(11):2588–2597.e1.

90. Ley D, Leroyer A, Dupont C, Sarter H, Bertrand V, Spyckerelle C, et al. New Therapeutic Strategies Are Associated With a Significant Decrease in Colectomy Rate in Pediatric Ulcerative Colitis. Am J Gastroenterol. 2023;118(11):1997.

91. Agrawal M, Ebert Ac, Poulsen G, Ungaro Rc, Faye As, Jess T, et al. Early Ileocecal Resection for Crohn's Disease Is Associated With Improved Long-term Outcomes Compared With Anti-Tumor Necrosis Factor Therapy: A Population-Based Cohort Study. Gastroenterology. 2023;165(4).

92. Revés J, Ungaro RC, Torres J. Unmet needs in inflammatory bowel disease. Curr Res Pharmacol Drug Discov. 2021;2(100070).

93. Vieujean S, Louis E. Precision medicine and drug optimization in adult inflammatory bowel disease patients. Ther Adv Gastroenterol. 2023;16:17562848231173332.

94. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. Nat Rev Immunol. 2008;8(6):458–66.

95. Moller FT, Andersen V, Wohlfahrt J, Jess T. Familial Risk of Inflammatory Bowel Disease: A Population-Based Cohort Study 1977–2011. Am J Gastroenterol. 2015;110(4):564–71.

96. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sørensen TI, Binder V. Familial occurrence of inflammatory bowel disease. N Engl J Med. 1991;324(2):84–8.

97. Baron S, Turck D, Leplat C, Merle V, Gower-Rousseau C, Marti R, et al. Environmental risk factors in paediatric inflammatory bowel diseases: a population based case control study. Gut. 2005;54(3):357–63.

98. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature. 2001;411(6837):599–603.

99. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001;411(6837):603–6.

100. Lesage S, Zouali H, Cézard JP, Colombel JF, Belaiche J, Almer S, et al. CARD15/NOD2 Mutational Analysis and Genotype-Phenotype Correlation in 612 Patients with Inflammatory Bowel Disease. Am J Hum Genet. 2002;70(4):845–57.

101. Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JPA. Differential Effects of NOD2 Variants on Crohn's Disease Risk and Phenotype in Diverse Populations: A Metaanalysis. Am J Gastroenterol. 2004;99(12):2393–404.

102. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017;49(2):256–61.

103. El Hadad J, Schreiner P, Vavricka SR, Greuter T. The Genetics of Inflammatory Bowel Disease. Mol Diagn Ther. 2024;28(1):27–35.

104. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119–24.

105. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, et al. A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. Science. 2006;314(5804):1461–3.

106. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet. 2007;39(5):596–604.

107. Kaser A, Lee AH, Franke A, Glickman JN, Zeissig S, Tilg H, et al. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. Cell. 2008;134(5):743–56.

108. Tysk C, Lindberg E, Järnerot G, Flodérus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. Gut. 1988;29(7):990–6.

109. Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser AL, Barnich N, et al. High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. Gastroenterology. 2004;127(2):412–21.

110. Sokol H, Seksik P, Rigottier-Gois L, Lay C, Lepage P, Podglajen I, et al. Specificities of the fecal microbiota in inflammatory bowel disease. Inflamm Bowel Dis. 2006;12(2):106–11.

111. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut. 2006;55(2):205–11.

112. Benchimol EI, Mack DR, Guttmann A, Nguyen GC, To T, Mojaverian N, et al. Inflammatory bowel disease in immigrants to Canada and their children: a population-based cohort study. Am J Gastroenterol. 2015;110(4):553–63.

113. Foster A, Jacobson K. Changing incidence of inflammatory bowel disease: environmental influences and lessons learnt from the South asian population. Front Pediatr. 2013;1:34.

114. Hammer T, Lophaven SN, Nielsen KR, von Euler-Chelpin M, Weihe P, Munkholm P, et al. Inflammatory bowel diseases in Faroese-born Danish residents and their offspring: further evidence of the dominant role of environmental factors in IBD development. Aliment Pharmacol Ther. 2017;45(8):1107–14.

115. Koutroubakis IE, Vlachonikolis IG. Appendectomy and the development of ulcerative colitis: results of a metaanalysis of published case-control studies. Am J Gastroenterol. 2000;95(1):171–6.

116. Andersson RE, Olaison G, Tysk C, Ekbom A. Appendectomy and protection against ulcerative colitis. N Engl J Med. 2001;344(11):808–14.

117. Sahami S, Kooij IA, Meijer SL, Van den Brink GR, Buskens CJ, Te Velde AA. The Link between the Appendix and Ulcerative Colitis: Clinical Relevance and Potential Immunological Mechanisms. Am J Gastroenterol. 2016;111(2):163–9.

118. Gardenbroek TJ, Pinkney TD, Sahami S, Morton DG, Buskens CJ, Ponsioen CY, et al. The ACCURE-trial: the effect of appendectomy on the clinical course of ulcerative colitis, a randomised international

multicenter trial (NTR2883) and the ACCURE-UK trial: a randomised external pilot trial (ISRCTN56523019). BMC Surg. 2015;15:30.

119. Harnoy Y, Bouhnik Y, Gault N, Maggiori L, Sulpice L, Cazals-Hatem D, et al. Effect of appendicectomy on colonic inflammation and neoplasia in experimental ulcerative colitis. Br J Surg. 2016;103(11):1530–8.

120. Welsh S, Sam Z, Seenan JP, Nicholson GA. The Role of Appendicectomy in Ulcerative Colitis: Systematic Review and Meta-Analysis. Inflamm Bowel Dis. 2023;29(4):633–46.

121. Piovani D, Danese S, Peyrin-Biroulet L, Nikolopoulos GK, Lytras T, Bonovas S. Environmental Risk Factors for Inflammatory Bowel Diseases: An Umbrella Review of Meta-analyses. Gastroenterology. 2019;157(3):647–659.e4.

122. Khalili H, Higuchi LM, Ananthakrishnan AN, Richter JM, Feskanich D, Fuchs CS, et al. Oral contraceptives, reproductive factors and risk of inflammatory bowel disease. Gut. 2013;62(8):1153–9.

123. Pasvol TJ, Bloom S, Segal AW, Rait G, Horsfall L. Use of contraceptives and risk of inflammatory bowel disease: a nested case-control study. Aliment Pharmacol Ther. 2022;55(3):318–26.

124. Shaw SY, Blanchard JF, Bernstein CN. Association between the use of antibiotics and new diagnoses of Crohn's disease and ulcerative colitis. Am J Gastroenterol. 2011;106(12):2133–42.

125. Mak JWY, Yang S, Stanley A, Lin X, Morrison M, Ching JYL, et al. Childhood antibiotics as a risk factor for Crohn's disease: The ENIGMA International Cohort Study. JGH Open. 2022;6(6):369–77.

126. Mark-Christensen A, Lange A, Erichsen R, Frøslev T, Esen BÖ, Sørensen HT, et al. Early-Life Exposure to Antibiotics and Risk for Crohn's Disease: A Nationwide Danish Birth Cohort Study. Inflamm Bowel Dis. 2022;28(3):415–22.

127. Hildebrand H, Malmborg P, Askling J, Ekbom A, Montgomery SM. Early-life exposures associated with antibiotic use and risk of subsequent Crohn's disease. Scand J Gastroenterol. 2008;43(8):961–6.

128. Agrawal M, Poulsen G, Colombel JF, Allin KH, Jess T. Maternal antibiotic exposure during pregnancy and risk of IBD in offspring: a population-based cohort study. Gut. 2023;72(4):804–5.

129. Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: a meta-analysis. Mayo Clin Proc. 2006;81(11):1462–71.

130. Higuchi LM, Khalili H, Chan AT, Richter JM, Bousvaros A, Fuchs CS. A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. Am J Gastroenterol. 2012;107(9):1399–406.

131. Xu L, Lochhead P, Ko Y, Claggett B, Leong RW, Ananthakrishnan AN. Systematic review with meta-analysis: breastfeeding and the risk of Crohn's disease and ulcerative colitis. Aliment Pharmacol Ther. 2017;46(9):780–9.

132. Khalili H, Chan SSM, Lochhead P, Ananthakrishnan AN, Hart AR, Chan AT. The role of diet in the aetiopathogenesis of inflammatory bowel disease. Nat Rev Gastroenterol Hepatol. 2018;15(9):525–35.

133. Meyer A, Dong C, Casagrande C, Chan SSM, Huybrechts I, Nicolas G, et al. Food Processing and Risk of Crohn's Disease and Ulcerative Colitis: A European Prospective Cohort Study. Clin Gastroenterol Hepatol. 2023;21(6):1607–1616.e6.

134. Narula N, Chang NH, Mohammad D, Wong ECL, Ananthakrishnan AN, Chan SSM, et al. Food Processing and Risk of Inflammatory Bowel Disease: A Systematic Review and Meta-Analysis. Clin Gastroenterol Hepatol. 2023;21(10):2483–2495.e1.

135. Chassaing B, Compher C, Bonhomme B, Liu Q, Tian Y, Walters W, et al. Randomized Controlled-Feeding Study of Dietary Emulsifier Carboxymethylcellulose Reveals Detrimental Impacts on the Gut Microbiota and Metabolome. Gastroenterology. 2022;162(3):743–56.

136. Khalili H, Håkansson N, Chan SS, Chen Y, Lochhead P, Ludvigsson JF, et al. Adherence to a Mediterranean diet is associated with a lower risk of later-onset Crohn's disease: results from two large prospective cohort studies. Gut. 2020;69(9):1637–44.

137. Turpin W, Dong M, Sasson G, Raygoza Garay JA, Espin-Garcia O, Lee SH, et al. Mediterranean-Like Dietary Pattern Associations With Gut Microbiome Composition and Subclinical Gastrointestinal Inflammation. Gastroenterology. 2022;163(3):685–98.

138. Jantchou P, Morois S, Clavel-Chapelon F, Boutron-Ruault MC, Carbonnel F. Animal protein intake and risk of inflammatory bowel disease: The E3N prospective study. Am J Gastroenterol. 2010;105(10):2195–201.

139. Talebi S, Zeraattalab-Motlagh S, Rahimlou M, Naeini F, Ranjbar M, Talebi A, et al. The Association between Total Protein, Animal Protein, and Animal Protein Sources with Risk of Inflammatory Bowel Diseases: A Systematic Review and Meta-Analysis of Cohort Studies. Adv Nutr. 2023;14(4):752–61.

140. Ananthakrishnan AN, Khalili H, Konijeti GG, Higuchi LM, de Silva P, Korzenik JR, et al. A prospective study of long-term intake of dietary fiber and risk of Crohn's disease and ulcerative colitis. Gastroenterology. 2013;145(5):970–7.

141. Andersen V, Chan S, Luben R, Khaw KT, Olsen A, Tjonneland A, et al. Fibre intake and the development of inflammatory bowel disease: A European prospective multi-centre cohort study (EPIC-IBD). J Crohns Colitis. 2018;12(2):129–36.

142. Casey K, Lopes EW, Niccum B, Burke K, Ananthakrishnan AN, Lochhead P, et al. Alcohol consumption and risk of inflammatory bowel disease among three prospective US cohorts. Aliment Pharmacol Ther. 2022;55(2):225–33.

143. Ananthakrishnan AN, Khalili H, Song M, Higuchi LM, Richter JM, Chan AT. Zinc intake and risk of Crohn's disease and ulcerative colitis: a prospective cohort study. Int J Epidemiol. 2015;44(6):1995–2005.

144. Ananthakrishnan AN, Khalili H, Higuchi LM, Bao Y, Korzenik JR, Giovannucci EL, et al. Higher predicted vitamin D status is associated with reduced risk of Crohn's disease. Gastroenterology. 2012;142(3):482–9.

145. Lu Y, Zamora-Ros R, Chan S, Cross AJ, Ward H, Jakszyn P, et al. Dietary Polyphenols in the Aetiology of Crohn's Disease and Ulcerative Colitis-A Multicenter European Prospective Cohort Study (EPIC). Inflamm Bowel Dis. 2017;23(12):2072–82.

146. Khalili H, Ananthakrishnan AN, Konijeti GG, Liao X, Higuchi LM, Fuchs CS, et al. Physical activity and risk of inflammatory bowel disease: prospective study from the Nurses' Health Study cohorts. BMJ. 2013;347:f6633.

147. Sun Y, Yuan S, Chen X, Sun J, Kalla R, Yu L, et al. The Contribution of Genetic Risk and Lifestyle Factors in the Development of Adult-Onset Inflammatory Bowel Disease: A Prospective Cohort Study. Am J Gastroenterol. 2023;118(3):511–22.

148. Lopes EW, Chan SSM, Song M, Ludvigsson JF, Håkansson N, Lochhead P, et al. Lifestyle factors for the prevention of inflammatory bowel disease. Gut. 2022;72(1093–1100).

149. Tenailleau QM, Lanier C, Gower-Rousseau C, Cuny D, Deram A, Occelli F. Crohn's disease and environmental contamination: Current challenges and perspectives in exposure evaluation. Environ Pollut. 2020;263(Pt B):114599.

150. Elten M, Benchimol EI, Fell DB, Kuenzig ME, Smith G, Chen H, et al. Ambient air pollution and the risk of pediatric-onset inflammatory bowel disease: A population-based cohort study. Environ Int. 2020;138:105676.

151. Elten M, Benchimol EI, Fell DB, Kuenzig ME, Smith G, Kaplan GG, et al. Residential Greenspace in Childhood Reduces Risk of Pediatric Inflammatory Bowel Disease: A Population-Based Cohort Study. Am J Gastroenterol. 2021;116(2):347.

152. Zhang Z, Chen L, Qian Z (Min), Li H, Cai M, Wang X, et al. Residential green and blue space associated with lower risk of adult-onset inflammatory bowel disease: Findings from a large prospective cohort study. Environ Int. 2022;160:107084.

153. Soon IS, Molodecky NA, Rabi DM, Ghali WA, Barkema HW, Kaplan GG. The relationship between urban environment and the inflammatory bowel diseases: a systematic review and meta-analysis. BMC Gastroenterol. 2012;12(1):51.

154. Nair N, Austin C, Curtin P, Gouveia C, Arora M, Torres J, et al. Association Between Early-life Exposures and Inflammatory Bowel Diseases, Based on Analyses of Deciduous Teeth. Gastroenterology. 2020;159(1):383–5.

155. Lochhead P, Khalili H, Ananthakrishnan AN, Burke KE, Richter JM, Sun Q, et al. Plasma concentrations of perfluoroalkyl substances and risk of inflammatory bowel diseases in women: A nested case control analysis in the Nurses' Health Study cohorts. Environ Res. 2022;207:112222.

156. Steenland K, Kugathasan S, Barr DB. PFOA and ulcerative colitis. Environ Res. 2018;165:317–21.

157. Steenland K, Zhao L, Winquist A. A cohort incidence study of workers exposed to perfluorooctanoic acid (PFOA). Occup Environ Med. 2015;72(5):373–80.

158. Steenland K, Zhao L, Winquist A, Parks C. Ulcerative Colitis and Perfluorooctanoic Acid (PFOA) in a Highly Exposed Population of Community Residents and Workers in the Mid-Ohio Valley. Environ Health Perspect. 2013;121(8):900–5.

159. Xu Y, Li Y, Scott K, Lindh CH, Jakobsson K, Fletcher T, et al. Inflammatory bowel disease and biomarkers of gut inflammation and permeability in a community with high exposure to perfluoroalkyl substances through drinking water. Environ Res. 2020;181:108923.

160. Chen D, Parks CG, Hofmann JN, Beane Freeman LE, Sandler DP. Pesticide use and inflammatory bowel disease in licensed pesticide applicators and spouses in the Agricultural Health Study. Environ Res. 2024;249:118464.

161. Agrawal M, Hansen AV, Colombel JF, Jess T, Allin KH. Association between early life exposure to agriculture, biodiversity, and green space and risk of inflammatory bowel disease: a population-based cohort study. EClinicalMedicine. 2024;70:102514.

162. Règlement sanitaire international (2005) [Internet]. OMS; 2005 [cited 2024 Feb 23]. Available from: https://www.who.int/fr/publications-detail/9789241580496

163. Revision of the European Standard Population - Report of Eurostat's task force - 2013 edition [Internet]. [cited 2024 Jan 31]. Available from: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-13-028

164. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. Stat Med. 1997;16(7):791–801.

165. Mortreux P, Leroyer A, Dupont C, Ley D, Bertrand V, Spyckerelle C, et al. Natural History of Anal Ulcerations in Pediatric-Onset Crohn's Disease: Long-Term Follow-Up of a Population-Based Study. Am J Gastroenterol. 2023;118(9):1671.

166. Dupont-Lucas C, Leroyer A, Ley D, Spyckerelle C, Bertrand V, Turck D, et al. Increased Risk of Cancer and Mortality in a Large French Population-Based Paediatric-Onset Inflammatory Bowel Disease Retrospective Cohort. J Crohns Colitis. 2023;17(4):524–34.

167. Fumery M, Dupont C, Ley D, Savoye G, Bertrand V, Guillon N, et al. Long-term effectiveness and safety of anti-TNF in pediatric-onset inflammatory bowel diseases: A population-based study. Dig Liver Dis. 2024;56(1):21–8.

168. Burisch J, Bergemalm D, Halfvarson J, Domislovic V, Krznaric Z, Goldis A, et al. The use of 5-aminosalicylate for patients with Crohn's disease in a prospective European inception cohort with 5 years follow-up - an Epi-IBD study. United Eur Gastroenterol J. 2020;8(8):949–60.

169. Burisch J, Vardi H, Schwartz D, Friger M, Kiudelis G, Kupčinskas J, et al. Health-care costs of inflammatory bowel disease in a pan-European, community-based, inception cohort during 5 years of follow-up: a population-based study. Lancet Gastroenterol Hepatol. 2020;5(5):454–64.

170. Kaplan GG. The global burden of IBD: from 2015 to 2025. Nat Rev Gastroenterol Hepatol. 2015;12(12):720–7.

171. Dorn-Rasmussen M, Lo B, Zhao M, Kaplan GG, Malham M, Wewer V, et al. The incidence and prevalence of paediatric- and adult-onset inflammatory bowel disease in Denmark during a 37-year period: a nationwide cohort study (1980-2017). J Crohns Colitis. 2023;17(2):259–68.

172. Le Berre C, Peyrin-Biroulet L, SPIRIT-IOIBD study group. Selecting End Points for Disease-Modification Trials in Inflammatory Bowel Disease: the SPIRIT Consensus From the IOIBD. Gastroenterology. 2021;160(5):1452–1460.e21.

173. Peyrin-Biroulet L, Cieza A, Sandborn WJ, Coenen M, Chowers Y, Hibi T, et al. Development of the first disability index for inflammatory bowel disease based on the international classification of functioning, disability and health. Gut. 2012;61(2):241–7.

174. Gower-Rousseau C, Sarter H, Savoye G, Tavernier N, Fumery M, Sandborn WJ, et al. Validation of the Inflammatory Bowel Disease Disability Index in a population-based cohort. Gut. 2017;66(4):588–96.

175. Williet N, Sarter H, Gower-Rousseau C, Adrianjafy C, Olympie A, Buisson A, et al. Patient-reported Outcomes in a French Nationwide Survey of Inflammatory Bowel Disease Patients. J Crohns Colitis. 2017;11(2):165–74.

176. Projections de population 2013-2050 pour les départements et les régions | Insee [Population projections 2013-2050] [Internet]. [cited 2024 Jan 31]. Available from: https://www.insee.fr/fr/statistiques/2859843

177. Tables de mortalité par sexe, âge et niveau de vie [Mortality tables]| Insee [Internet]. [cited 2024 Jan 31]. Available from: https://www.insee.fr/fr/statistiques/3311422

178. van den Heuvel TRA, Jeuring SFG, Zeegers MP, van Dongen DHE, Wolters A, Masclee AAM, et al. A 20-Year Temporal Change Analysis in Incidence, Presenting Phenotype and Mortality, in the Dutch IBDSL Cohort-Can Diagnostic Factors Explain the Increase in IBD Incidence? J Crohns Colitis. 2017;11(10):1169–79.

179. Shaffer SR, Kuenzig ME, Windsor JW, Bitton A, Jones JL, Lee K, et al. The 2023 Impact of Inflammatory Bowel Disease in Canada: Special Populations-IBD in Seniors. J Can Assoc Gastroenterol. 2023;6(Suppl 2):S45–54.

180. Vieujean S, Caron B, Jairath V, Benetos A, Danese S, Louis E, et al. Is it time to include older adults in inflammatory bowel disease trials? A call for action. Lancet Healthy Longev. 2022;3(5):e356–66.

181. Shah SC, Khalili H, Gower-Rousseau C, Olen O, Benchimol EI, Lynge E, et al. Sex-based differences in incidence of inflammatory bowel diseases-pooled analysis ofpopulation-based studies from western countries. Gastroenterology. 2018;155(4):1079–1089.e3.

182. Cosnes J. Smoking, physical activity, nutrition and lifestyle: environmental factors and their impact on IBD. Dig Dis. 2010;28(3):411–7.

183. OFDT. Tabac : évolution de l'usage occasionnel ou régulier parmi les 18-75 ans [Evolution of tobacco use in 18-75 years] [Internet]. OFDT; 2019 [cited 2024 Jan 31]. Available from: https://www.ofdt.fr/pdf/561

184. Drees, Etudes et Résultats No 868. La longue diminution des appendicectomies en France | Direction de la recherche, des études, de l'évaluation et des statistiques [Diminution of appendicectomies on France] [Internet]. 2014 [cited 2024 Jan 31]. Available from: https://drees.solidarites-sante.gouv.fr/publications/etudes-et-resultats/la-longue-diminution-des-appendicectomies-en-france-0

185. Smyth M, Chan J, Evans K, Penner C, Lakhani A, Newlove T, et al. Cross-sectional analysis of quality of life in pediatric patients with inflammatory bowel disease in British Columbia, Canada. J Pediatr. 2021;238:57–65.e2.

186. Chouliaras G, Margoni D, Dimakou K, Fessatou S, Panayiotou I, Roma-Giannikou E. Disease impact on the quality of life of children with inflammatory bowel disease. World J Gastroenterol. 2017;23(6):1067–75.

187. Buie MJ, Coward S, Shaheen AA, Holroyd-Leduc J, Hracs L, Ma C, et al. Hospitalization rates for inflammatory bowel disease are decreasing over time: a population-based cohort study. Inflamm Bowel Dis. 2023;29(10):1536–45.

188. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. Lancet. 1980;1(8167):514.

189. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. Gut. 1998;43(1):29–32.

190. Irvine E, Zhou Q, Thompson A. The Short Inflammatory Bowel Disease Questionnaire: a quality of life instrument for community physicians managing inflammatory bowel disease. CCRPT Investigators. Canadian Crohn's Relapse Prevention Trial. Am J Gastroenterol. 1996;91(8):1571–8.

191. Rey G, Rican S, Jougla E. Measurement of mortality inequalities by cause of death - Ecological approach using a social disadvantage index. [Mesure des inégalités de mortalité par cause de décès - Approche écologique à l'aide d'un indice de désavantage social.]. Bull Epidémiologique Hebd. 2011;(8–9):87–90.

192. Ferguson A, Sedgwick DM, Drummond J. Morbidity of juvenile onset inflammatory bowel disease: effects on education and employment in early adult life. Gut. 1994;35(5):665–8.

193. Mackner LM, Bickmeier RM, Crandall WV. Academic achievement, attendance, and school-related quality of life in pediatric inflammatory bowel disease. J Dev Behav Pediatr. 2012;33(2):106–11.

194. Mayberry MK, Probert C, Srivastava E, Rhodes J, Mayberry JF. Perceived discrimination in education and employment by people with Crohn's disease: a case control study of educational achievement and employment. Gut. 1992;33(3):312–4.

195. Hummel TZ, Tak E, Maurice-Stam H, Benninga MA, Kindermann A, Grootenhuis MA. Psychosocial developmental trajectory of adolescents with inflammatory bowel disease. J Pediatr Gastroenterol Nutr. 2013;57(2):219–24.

196. El-Matary W, Dufault B, Moroz SP, Schellenberg J, Bernstein CN. Education, employment, income, and marital status among adults diagnosed with inflammatory bowel diseases during childhood or adolescence. Clin Gastroenterol Hepatol. 2017;15(4):518–24.

197. Singh H, Nugent Z, Brownell M, Targownik LE, Roos LL, Bernstein CN. Academic performance among children with inflammatory bowel disease: a population-based study. J Pediatr. 2015;166(5):1128–33.

198. Malmborg P, Everhov ÅH, Söderling J, Ludvigsson JF, Bruze G, Olén O. Earnings during adulthood in patients with childhood-onset inflammatory bowel disease: a nationwide population-based cohort study. Aliment Pharmacol Ther. 2022;56(6):1007–17.

199. Rasmussen J, Nørgård BM, Nielsen RG, Bøggild H, Qvist N, Brund RBK, et al. Implication of inflammatory bowel disease diagnosed before the age of 18 for achieving an upper secondary education: a nationwide population-based cohort study. Inflamm Bowel Dis. 2024;30(2):247–56.

200. Calsbeek H, Rijken M, Dekker J, van Berge Henegouwen GP. Disease characteristics as determinants of the labour market position of adolescents and young adults with chronic digestive disorders. Eur J Gastroenterol Hepatol. 2006;18(2):203–9.

201. Hodson R. Precision medicine. Nature. 2016;537(7619):S49.

202. van Schaik FDM, Oldenburg B, Hart AR, Siersema PD, Lindgren S, Grip O, et al. Serological markers predict inflammatory bowel disease years before the diagnosis. Gut. 2013;62(5):683–8.

203. Lee SH, Turpin W, Espin-Garcia O, Raygoza Garay JA, Smith MI, Leibovitzh H, et al. Anti-Microbial Antibody Response is Associated With Future Onset of Crohn's Disease Independent of Biomarkers of Altered Gut Barrier Function, Subclinical Inflammation, and Genetic Risk. Gastroenterology. 2021;161(5):1540–51.

204. Torres J, Petralia F, Sato T, Wang P, Telesco SE, Choung RS, et al. Serum Biomarkers Identify Patients Who Will Develop Inflammatory Bowel Diseases Up to 5 Years Before Diagnosis. Gastroenterology. 2020;159(1):96–104.

205. Loly C, Belaiche J, Louis E. Predictors of severe Crohn's disease. Scand J Gastroenterol. 2008;43(8):948–54.

206. Beaugerie L, Seksik P, Nion-Larmurier I, Gendre JP, Cosnes J. Predictors of Crohn's Disease. Gastroenterology. 2006;130(3):650–6.

207. Lichtenstein GR, Olson A, Travers S, Diamond RH, Chen DM, Pritchard ML, et al. Factors associated with the development of intestinal strictures or obstructions in patients with Crohn's disease. Am J Gastroenterol. 2006;101(5):1030–8.

208. Solberg IC, Høivik ML, Cvancarova M, Moum B, IBSEN Study Group. Risk matrix model for prediction of colectomy in a population-based study of ulcerative colitis patients (the IBSEN study). Scand J Gastroenterol. 2015;50(12):1456–62.

209. Jung C, Colombel JF, Lemann M, Beaugerie L, Allez M, Cosnes J, et al. Genotype/phenotype analyses for 53 Crohn's disease associated genetic polymorphisms. PloS One. 2012;7(12):e52223.

210. Louis E, Michel V, Hugot JP, Reenaers C, Fontaine F, Delforge M, et al. Early development of stricturing or penetrating pattern in Crohn's disease is influenced by disease location, number of flares, and smoking but not by NOD2/CARD15 genotype. Gut. 2003;52(4):552–7.

211. Cleynen I, González JR, Figueroa C, Franke A, McGovern D, Bortlík M, et al. Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. Gut. 2013;62(11):1556–65.

212. Ryan JD, Silverberg MS, Xu W, Graff LA, Targownik LE, Walker JR, et al. Predicting complicated Crohn's disease and surgery: phenotypes, genetics, serology and psychological characteristics of a population-based cohort. Aliment Pharmacol Ther. 2013;38(3):274–83.

213. Weersma RK, Stokkers PCF, van Bodegraven AA, van Hogezand RA, Verspaget HW, de Jong DJ, et al. Molecular prediction of disease risk and severity in a large Dutch Crohn's disease cohort. Gut. 2009;58(3):388–95.

214. Siegel CA, Horton H, Siegel LS, Thompson KD, Mackenzie T, Stewart SK, et al. A validated web-based tool to display individualised Crohn's disease predicted outcomes based on clinical, serologic and genetic variables. Aliment Pharmacol Ther. 2016;43(2):262–71.

215. Heliö T, Halme L, Lappalainen M, Fodstad H, Paavola-Sakki P, Turunen U, et al. CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. Gut. 2003;52(4):558–62.

216. Abreu MT, Taylor KD, Lin YC, Hang T, Gaiennie J, Landers CJ, et al. Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. Gastroenterology. 2002;123(3):679–88.

217. Adler J, Rangwalla SC, Dwamena BA, Higgins PDR. The prognostic power of the NOD2 genotype for complicated Crohn's disease: a meta-analysis. Am J Gastroenterol. 2011;106(4):699–712.

218. Kugathasan S, Denson LA, Walters TD, Kim MO, Marigorta UM, Schirmer M, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. Lancet. 2017;389(10080):1710–8.

219. Lee JC, Biasci D, Roberts R, Gearry RB, Mansfield JC, Ahmad T, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. Nat Genet. 2017;49(2):262–8.

220. Dubinsky MC, Kugathasan S, Kwon S, Haritunians T, Wrobel I, Wahbeh G, et al. Multidimensional prognostic risk assessment identifies association between IL12B variation and surgery in Crohn's disease. Inflamm Bowel Dis. 2013;19(8):1662–70.

221. Satsangi J, Welsh KI, Bunce M, Julier C, Farrant JM, Bell JI, et al. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet. 1996;347(9010):1212–7.

222. Ungaro RC, Hu L, Ji J, Nayar S, Kugathasan S, Denson LA, et al. Machine learning identifies novel blood protein predictors of penetrating and stricturing complications in newly diagnosed paediatric Crohn's disease. Aliment Pharmacol Ther. 2020;53(2):281–90.

223. Lee JC, Lyons PA, McKinney EF, Sowerby JM, Carr EJ, Bredin F, et al. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. J Clin Invest. 2011;121(10):4170–9.

224. Biasci D, Lee JC, Noor NM, Pombal DR, Hou M, Lewis N, et al. A blood-based prognostic biomarker in IBD. Gut. 2019;68(8):1386–95.

225. Noor NM, Lee JC, Bond S, Dowling F, Brezina B, Patel KV, et al. A biomarker-stratified comparison of top-down versus accelerated step-up treatment strategies for patients with newly diagnosed Crohn's disease (PROFILE): a multicentre, open-label randomised controlled trial. Lancet Gastroenterol Hepatol. 2024;9(5):415–27.

226. Plaza J, Mínguez A, Bastida G, Marqués R, Nos P, Poveda JL, et al. Genetic Variants Associated with Biological Treatment Response in Inflammatory Bowel Disease: A Systematic Review. Int J Mol Sci. 2024;25(7):3717.

227. Verstockt B, Verstockt S, Dehairs J, Ballet V, Blevi H, Wollants WJ, et al. Low TREM1 expression in whole blood predicts anti-TNF response in inflammatory bowel disease. EBioMedicine. 2019;40:733–42.

228. Savelkoul EHJ, Thomas PWA, Derikx LAAP, den Broeder N, Römkens TEH, Hoentjen F. Systematic Review and Meta-analysis: Loss of Response and Need for Dose Escalation of Infliximab and Adalimumab in Ulcerative Colitis. Inflamm Bowel Dis. 2022;29(10):1633–47.

229. Hastie T, Tibshirani R, Friedman J. High-Dimensional Problems: p N. In: The Elements of Statistical Learning. Springer, New York, NY; 2009. p. 649–98.

230. Wold, H. Estimation of principal components and related models by iterative least squares. In: Multivariate Analysis. New-York: P.R. Krishnaiaah; 1966. p. 391–420. (Academic Press).

231. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B. 1996;58(1):267–88.

232. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 2000;42(1):80–6.

233. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. J R Stat Soc Ser B. 2005;67(2):301–20.

234. Boulesteix AL, Sauerbrei W. Added predictive value of high-throughput molecular data to clinical data and its validation. Brief Bioinform. 2011;12(3):215–29.

235. De Bin R, Sauerbrei W, Boulesteix AL. Investigating the prediction ability of survival models based on both clinical and omics data: two case studies. Stat Med. 2014;33(30):5310–29.

236. De Bin R, Boulesteix AL, Benner A, Becker N, Sauerbrei W. Combining clinical and molecular data in regression prediction models: insights from a simulation study. Brief Bioinform. 2020;21(6):1904–19.

237. Bühlmann P, Hothorn T. Boosting Algorithms: Regularization, Prediction and Model Fitting. Stat Sci. 2007;22(4):477–505.

238. Zhao Q, Shi X, Xie Y, Huang J, Shia B, Ma S. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. Brief Bioinform. 2015;16(2):291–303.

239. Boulesteix AL, De Bin R, Jiang X, Fuchs M. IPF-LASSO: Integrative L1-Penalized Regression with Penalty Factors for Prediction Based on Multi-Omics Data. Comput Math Methods Med. 2017;2017:7691937.

240. Tenenhaus A, Philippe C, Guillemot V, Le Cao KA, Grill J, Frouin V. Variable selection for generalized canonical correlation analysis. Biostatistics. 2014;15(3):569–83.

241. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat. 2004;32(2):407–99.

242. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33:1–22.

243. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385–95.

244. Hastie T, Tibshirani R, Wainwright, Martin. Statistical Learning with Sparsity : The Lasso and Generalizations. Chapman and Hall/CRC; 2015.

245. Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc Ser B. 2010;72(4):417–73.

246. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B. 2006;68(1):49–67.

247. Simon N, Friedman J, Hastie T, Tibshirani R. A Sparse-Group Lasso. J Comput Graph Stat. 2013;22(2):231–45.

248. Grimonprez Q, Blanck S, Celisse A, Marot G. MLGL: An R Package Implementing Correlated Variable Selection by Hierarchical Clustering and Group-Lasso. J Stat Softw. 2023;106:1–33.

249. Zou H. The Adaptive Lasso and Its Oracle Properties. J Am Stat Assoc. 2006;101(476):1418–29.

250. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965;30(2):179–85.

251. Vernier-Massouille G, Balde M, Salleron J, Turck D, Dupas JL, Mouterde O, et al. Natural history of pediatric Crohn's disease: a population-based cohort study. Gastroenterology. 2008;135(4):1106–13.

252. Herzog D, Fournier N, Buehr P, Rueger V, Koller R, Heyland K, et al. Prevalence of intestinal complications in inflammatory bowel disease: a comparison between paediatric-onset and adult-onset patients. Eur J Gastroenterol Hepatol. 2017;29(8):926–31.

253. Colombel JF, Narula N, Peyrin-Biroulet L. Management Strategies to Improve Outcomes of Patients With Inflammatory Bowel Diseases. Gastroenterology. 2017;152(2):351–361.e5.

254. Khanna R, Bressler B, Levesque BG, Zou G, Stitt LW, Greenberg GR, et al. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. Lancet. 2015;386(10006):1825–34.

255. Bae JM, Lee HH, Lee BI, Lee KM, Eun SH, Cho ML, et al. Incidence of psoriasiform diseases secondary to tumour necrosis factor antagonists in patients with inflammatory bowel disease: a nationwide population-based cohort study. Aliment Pharmacol Ther. 2018;48(2):196–205.

256. Kirchgesner J, Lemaitre M, Carrat F, Zureik M, Carbonnel F, Dray-Spira R. Risk of Serious and Opportunistic Infections Associated With Treatment of Inflammatory Bowel Diseases. Gastroenterology. 2018;155(2):337–346.e10.

257. Lichtenstein GR, Rutgeerts P, Sandborn WJ, Sands BE, Diamond RH, Blank M, et al. A pooled analysis of infections, malignancy, and mortality in infliximab- and immunomodulator-treated adult patients with inflammatory bowel disease. Am J Gastroenterol. 2012;107(7):1051–63.

258. Savoye G, Salleron J, Gower-Rousseau C, Dupas JL, Vernier-Massouille G, Fumery M, et al. Clinical predictors at diagnosis of disabling pediatric Crohn's disease. Inflamm Bowel Dis. 2012;18(11):2072–8.

259. Dubinsky MC, Lin YC, Dutridge D, Picornell Y, Landers CJ, Farrior S, et al. Serum immune responses predict rapid disease progression among children with Crohn's disease: immune responses predict disease progression. Am J Gastroenterol. 2006;101(2):360–7.

260. Gupta N, Cohen SA, Bostrom AG, Kirschner BS, Baldassano RN, Winter HS, et al. Risk factors for initial surgery in pediatric patients with Crohn's disease. Gastroenterology. 2006;130(4):1069–77.

261. Amre DK, Lu SE, Costea F, Seidman EG. Utility of serological markers in predicting the early occurrence of complications and surgery in pediatric Crohn's disease patients. Am J Gastroenterol. 2006;101(3):645–52.

262. Jakobsen C, Cleynen I, Andersen PS, Vermeire S, Munkholm P, Paerregaard A, et al. Genetic susceptibility and genotype-phenotype association in 588 Danish children with inflammatory bowel disease. J Crohns Colitis. 2014;8(7):678–85.

263. Atia O, Kang B, Orlansky-Meyer E, Ledder O, Lev Tzion R, Choi S, et al. Existing prediction models of disease course in pediatric Crohn's disease are poorly replicated in a prospective inception cohort. J Crohns Colitis. 2022;16(7):1039–48.

264. Mittag F, Römer M, Zell A. Influence of Feature Encoding and Choice of Classifier on Disease Risk Prediction in Genome-Wide Association Studies. PloS One. 2015;10(8):e0135832.

265. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338(7708):1432–5.

266. Dunkler D, Sauerbrei W, Heinze G. Global, Parameterwise and Joint Shrinkage Factor Estimation. J Stat Softw. 2016;69:1–19.

267. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352:i6.

268. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565–74.

269. Dubinsky MC, Kugathasan S, Mei L, Picornell Y, Nebel J, Wrobel I, et al. Increased immune reactivity predicts aggressive complicating Crohn's disease in children. Clin Gastroenterol Hepatol. 2008;6(10):1105–11.

270. Ricciuto A, Aardoom M, Meyer EO, Navon D, Carman N, Aloi M, et al. Predicting Outcomes in Pediatric Crohn's Disease for Management Optimization: Systematic Review and Consensus Statements From the Pediatric Inflammatory Bowel Disease-Ahead Program. Gastroenterology. 2021;160(1):403–436.e26.

271. Gasparetto M, Payne F, Nayak K, Kraiczy J, Glemas C, Philip-McKenzie Y, et al. Transcription and DNA Methylation Patterns of Blood-Derived CD8+ T Cells Are Associated With Age and Inflammatory Bowel Disease But Do Not Predict Prognosis. Gastroenterology. 2021;160(1):232–244.e7.

272. Poulain D, Sendid B, Standaert-Vitse A, Fradin C, Jouault T, Jawhara S, et al. Yeasts: neglected pathogens. Dig Dis. 2009;27 Suppl 1:104–10.

273. Müller S, Styner M, Seibold-Schmid B, Flogerzi B, Mähler M, Konrad A, et al. Anti-Saccharomyces cerevisiae antibody titers are stable over time in Crohn's patients and are not inducible in murine models of colitis. World J Gastroenterol. 2005;11(44):6988–94.

274. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. J Natl Cancer Inst. 2010;102(3):152–60.

275. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br J Cancer. 2014;110(3):551–5.

276. Gustav L. Jakobsson JH Emil Sternegård, Ola Olén, Pär Myrelid, Rickard Ljung, Hans Strid, Ludvigsson JF. Validating inflammatory bowel disease (IBD) in the Swedish National Patient Register and the Swedish Quality Register for IBD (SWIBREG). Scand J Gastroenterol. 2017;52(2):216–221.

277. Benchimol EI, Guttmann A, Mack DR, Nguyen GC, Marshall JK, Gregor JC, et al. Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. J Clin Epidemiol. 2014;67(8):887–96.

278. Kirchgesner J, Desai RJ, Beaugerie L, Kim SC, Schneeweiss S. Calibrating Real-World Evidence Studies Against Randomized Trials: Treatment Effectiveness of Infliximab in Crohn's Disease. Clin Pharmacol Ther. 2022;111(1):179–86.

279. Kirchgesner J, Desai RJ, Schneeweiss MC, Beaugerie L, Kim SC, Schneeweiss S. Emulation of a randomized controlled trial in ulcerative colitis with US and French claims data: Infliximab with thiopurines compared to infliximab monotherapy. Pharmacoepidemiol Drug Saf. 2022;31(2):167–75.

280. Hasanzad M, Sarhangi N, Ehsani Chimeh S, Ayati N, Afzali M, Khatami F, et al. Precision medicine journey through omics approach. J Diabetes Metab Disord. 2022;21(1):881–8.

281. Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, et al. Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics. 2016;17(2):S15.

282. Pierre-Jean M, Deleuze JF, Le Floch E, Mauger F. Clustering and variable selection evaluation of 13 unsupervised methods for multi-omics data integration. Brief Bioinform. 2020;21(6):2011–30.

283. Cantini L, Zakeri P, Hernandez C, Naldi A, Thieffry D, Remy E, et al. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. Nat Commun. 2021;12(1):124.

284. Vandereyken K, Sifrim A, Thienpont B, Voet T. Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet. 2023;24(8):494–515.

285. Edén P, Ritz C, Rose C, Fernö M, Peterson C. "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. Eur J Cancer. 2004;40(12):1837–41.

286. Bøvelstad HM, Nygård S, Borgan Ø. Survival prediction from clinico-genomic models - a comparative study. BMC Bioinformatics. 2009;10(1):413.

287. Volkmann A, De Bin R, Sauerbrei W, Boulesteix AL. A plea for taking all available clinical information into account when assessing the predictive value of omics data. BMC Med Res Methodol. 2019;19(1):162.

288. Wang ZW, Xu QN, Li CT, Liu XL. Age Estimation Based on DNA Methylation and Its Application Prospects in Forensic Medicine. Fa Yi Xue Za Zhi. 2023;39(1):72–82.

289. Jardillier R, Koca D, Chatelain F, Guyon L. Prognosis of lasso-like penalized Cox models with tumor profiling improves prediction over clinical data alone and benefits from bi-dimensional pre-screening. BMC Cancer. 2022;22(1):1045.

290. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373–9.

291. Lee JWJ, Plichta D, Hogstrom L, Borren NZ, Lau H, Gregory SM, et al. Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. Cell Host Microbe. 2021;29(8):1294–1304.e4.

292. Charpentier C, Salleron J, Savoye G, Fumery M, Merle V, Laberenne JE, et al. Natural history of elderly-onset inflammatory bowel disease: a population-based cohort study. Gut. 2014;63(3):423–32.

293. Fumery M, Pariente B, Sarter H, Savoye G, Spyckerelle C, Djeddi D, et al. Long-term outcome of pediatric-onset Crohn's disease: A population-based cohort study. Dig Liver Dis. 2019;51(4):496–502.

294. Ananthakrishnan AN, Donaldson T, Lasch K, Yajnik V. Management of Inflammatory Bowel Disease in the Elderly Patient: Challenges and Opportunities. Inflamm Bowel Dis. 2017;23(6):882–93.

295. Pierre-Jean M, Mauger F, Deleuze JF, Le Floch E. PIntMF: Penalized Integrative Matrix Factorization method for multi-omics data. Bioinformatics. 2021;38(4):900–7.

296. Joosten E a. G, DeFuentes-Merillas L, de Weert GH, Sensky T, van der Staak CPF, de Jong C a. J. Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. Psychother Psychosom. 2008;77(4):219–26.

297. Baars JE, Markus T, Kuipers EJ, van der Woude CJ. Patients' preferences regarding shared decision-making in the treatment of inflammatory bowel disease: results from a patient-empowerment study. Digestion. 2010;81(2):113–9.

298. Schoefs E, Vermeire S, Ferrante M, Sabino J, Lambrechts T, Avedano L, et al. What are the Unmet Needs and Most Relevant Treatment Outcomes According to Patients with Inflammatory Bowel Disease? A Qualitative Patient Preference Study. J Crohns Colitis. 2023;17(3):379–88.

299. Nachury M, Bouhnik Y, Serrero M, Filippi J, Roblin X, Kirchgesner J, et al. Patients' real-world experience with inflammatory bowel disease: A cross-sectional survey in tertiary care centres from the GETAID group. Dig Liver Dis. 2021;53(4):434–41.

300. Eindor-Abarbanel A, Pinchevski N, Shalem T, Agajany N, Ophir N, Weiss B, et al. Parental perspectives on pediatric inflammatory bowel disease: Unraveling concerns, and study participation willingness. J Pediatr Gastroenterol Nutr. 2024;78(4):862–70.

301. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36.

302. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide. Cambridge University Press; 2011.

303. Andra SS, Austin C, Arora M. The tooth exposome in children's health research. Curr Opin Pediatr. 2016;28(2):221–7.

304. Arora M, Austin C. Teeth as a biomarker of past chemical exposure. Curr Opin Pediatr. 2013;25(2):261–7.

305. Arora M, Reichenberg A, Willfors C, Austin C, Gennings C, Berggren S, et al. Fetal and postnatal metal dysregulation in autism. Nat Commun. 2017;8:15493.

Annexe : Matériel supplémentaire de l'article sur le score PREDICT-EPIMAD

Version acceptée de l'article publié dans la revue Inflammatory Bowel Diseases en 2023.

Supplementary material

Supplementary Methods:

Supplementary table S1: List of the 369 SNPs included in statistical analyses. MAF : minor allele frequency.

chromosome	gene	SNP	Minor Allele (CEU dbSNP)	Major allele (CEU dbSNP)	MAF in Epimad cohort (n=156)
chr1	AGT	rs699	G	A	0.43
chr1	ATF3	rs10735510	G	C	0.47
chr1	C1Orf141	rs2144658	A	G	0.33
chr1	CD2	rs3136706	T	C	0.28
chr1	CD247	rs864537	G	A	0.39
chr1	CD48	rs1553456	A	G	0.23
chr1	CRCT1	rs4845783	G	A	0.30
chr1	DENND1B	rs1998598	G	A	0.34
chr1	ECM1	rs11205387	A	G	0.29
chr1	FAM5C	rs1954603	C	T	0.08
chr1	FASLG	rs9286879	G	A	0.27
chr1	FCGR2A	rs1801274	G	A	0.49
chr1	IL10	rs17259637	G	A	0.08
chr1	IL10	rs1800896	C	T	0.47
chr1	IL10	rs3024493	A	C	0.17
chr1	IL10	rs3024505	A	G	0.16
chr1	IL12RB2	rs881087	C	T	0.34
chr1	IL12RB2	rs881089	T	C	0.46
chr1	IL23R	rs11209026	A	G	0.04
chr1	IL23R	rs11805303	T	C	0.38
chr1	IL6R	rs4129267	T	C	0.45
chr1	ITLN1	rs2274910	T	C	0.29
chr1	KIF1B	rs10492972	C	T	0.26
chr1	KIF21B	rs7554511	A	C	0.25
chr1	MMEL1	rs6684865	A	G	0.31
chr1	NLRP3	rs10732302	A	G	0.14

chr1	NLRP3	rs35829419	A	C	0.04
chr1	NLRP3	rs4353135	G	T	0.29
chr1	NLRP3	rs6672995	A	G	0.16
chr1	OTUD3	rs4654925	C	G	0.48
chr1	OTUD3	rs6426833	G	A	0.47
chr1	PER3	rs2797685	T	C	0.17
chr1	PHC2	rs4653044	C	A	0.44
chr1	PTPN22	rs2476601	A	G	0.07
chr1	PTPRC	rs1326279	T	A	0.31
chr1	RGS1	rs2816316	C	A	0.15
chr1	RPL5	rs6604026	C	T	0.27
chr1	RPS6KA1	rs12025634	T	C	0.14
chr1	SCAMP3	rs1076556	G	A	0.28
chr1	SRP9	rs4653433	A	G	0.36
chr1	TLR5	rs851192	G	C	0.39
chr1	TNFRSF14	rs6667605	T	C	0.49
chr1	TNFRSF9	rs2453021	T	C	0.39
chr1	TNFSF18	rs2236876	A	G	0.28
chr1	WNT4	rs7524102	G	A	0.17
chr2	AFF3	rs1437377	C	T	0.10
chr2	ARPC2	rs12612347	G	A	0.46
chr2	ATG16L1	rs2241880	A	G	0.39
chr2	CFLAR	rs2041765	G	A	0.46
chr2	DNMT3A	rs13428812	G	A	0.30
chr2	E2F6	rs6751915	T	G	0.31
chr2	GCKR	rs780093	T	C	0.44
chr2	IL18R1	rs3771166	A	G	0.41
chr2	IL18RAP	rs2058660	G	A	0.27
chr2	IL1R2	rs2310173	T	G	0.44
chr2	IL8RA	rs11676348	T	C	0.51
chr2	PLCL1	rs6738825	A	G	0.48
chr2	PUS10	rs13003464	G	A	0.45
chr2	REL	rs842647	G	A	0.22
chr2	SATB2	rs1992950	G	A	0.29
chr2	SP140	rs28445040	T	C	0.25
chr2	IL1RL1	rs10197862	G	A	0.17
chr2	STAT1	rs2030171	A	G	0.34
chr2	STAT4	rs7574865	T	G	0.18
chr2	THADA	rs10495903	T	C	0.13
chr2	TNFRSF1A	rs17758146	T	C	0.23
chr3	APH	rs4855881	G	A	0.45
chr3	ATG7	rs2305295	C	T	0.34
chr3	CADM2	rs7611991	A	G	0.28
chr3	GPX1	rs17650792	G	A	0.44
chr3	GPX1	rs1800668	A	G	0.31
chr3	IHPK1	rs9872864	G	A	0.49

chr3	IL12A	rs17810546	G	A	0.10
chr3	LPP	rs1464510	A	C	0.43
chr3	MST1	rs11718165	G	A	0.28
chr3	SATB1	rs13073817	A	G	0.31
chr3	TAK1	rs13098497	G	A	0.24
chr3	TLR9	rs4082828	C	A	0.12
chr3	USP4	rs11720964	T	C	0.48
chr4	EPHA5	rs11735820	G	T	0.34
chr4	GC	rs2282679	G	T	0.30
chr4	IL21	rs1398553	A	G	0.33
chr4	KIAA1109	rs6822844	T	G	0.12
chr4	NFKB	rs230530	G	A	0.47
chr4	RBPJ	rs6817712	G	A	0.26
chr4	PHOX2B	rs16853571	C	A	0.10
chr4	TLR1	rs5743595	G	A	0.16
chr5	CAMK2A	rs3822607	G	A	0.36
chr5	CEP72	rs4957048	A	G	0.22
chr5	CPEB4	rs359457	C	T	0.40
chr5	DAB2	rs1373692	A	C	0.35
chr5	DAB2	rs9292777	G	A	0.35
chr5	DAP	rs267939	C	T	0.44
chr5	IL12B	rs10045431	A	C	0.22
chr5	IL12B	rs1363670	C	G	0.12
chr5	IL12B	rs3213094	T	C	0.20
chr5	IL7R	rs6897932	T	C	0.26
chr5	IRF1	rs2070727	A	C	0.33
chr5	IRGM	rs10065172	T	C	0.17
chr5	IRGM	rs13361189	C	T	0.17
chr5	IRGM	rs4958847	A	G	0.23
chr5	LRAP	rs2549794	C	T	0.44
chr5	NDFIP1	rs11167764	A	C	0.20
chr5	PITX1	rs254560	A	G	0.39
chr5	PTGER4	rs4613763	C	T	0.15
chr5	SCL22A4	rs1050152	T	C	0.48
chr5	SCL22A4	rs2631367	G	C	0.44
chr5	SLC22A4	rs12521868	T	G	0.47
chr5	SLC22A4	rs3792876	T	C	0.09
chr5	SLC22A5	rs2631362	G	A	0.26
chr5	TMEM174	rs7702331	G	A	0.40
chr5	TSLP	rs1837253	T	C	0.24
chr5	ENSG00000249738	rs1422878	T	C	0.33
chr6	BACH2	rs1847472	A	C	0.32
chr6	CCR6	rs2301436	T	C	0.49
chr6	CNR1	rs4707436	A	G	0.28
chr6	HLA	rs12191877	T	C	0.17
chr6	HLA	rs1799964	C	T	0.25

chr6	HLA	rs9263739	T	C	0.16
chr6	HLADQ	rs2858330	T	C	0.46
chr6	HLADQ	rs6457617	C	T	0.49
chr6	HLADQ	rs9272346	G	A	0.45
chr6	HLADQA1	rs477515	A	G	0.24
chr6	HLADRA	rs3135388	A	G	0.10
chr6	HLADRA	rs9268877	A	G	0.43
chr6	IL17A	rs7747909	A	G	0.25
chr6	LOC391040	rs6933404	C	T	0.20
chr6	LYRM4	rs12529198	G	A	0.06
chr6	MAP3K7IP2	rs7758080	G	A	0.30
chr6	MAPK14	rs9470219	A	C	0.44
chr6	PRDM1	rs548234	C	T	0.31
chr6	PRDM1	rs6903235	G	A	0.29
chr6	RAGE	rs1800624	T	A	0.30
chr6	RPS6KA2	rs9459678	C	T	0.39
chr6	SLC22A23	rs17309827	G	T	0.35
chr6	SPDEF	rs3798544	A	G	0.11
chr6	TAGAP	rs212388	C	T	0.40
chr6	TRAF3IP2	rs33980500	T	C	0.09
chr6	TRIM10	rs259940	G	A	0.27
chr6	Unknown	rs7746082	C	G	0.33
chr6	VEGFA	rs943072	G	T	0.11
chr7	ABCB1	rs2235048	A	G	0.49
chr7	ABCB5	rs1011559	C	A	0.17
chr7	ARPC1A	rs3801288	G	A	0.14
chr7	C7Orf33	rs12704040	C	G	0.42
chr7	CNTNAP2	rs7807268	G	C	0.48
chr7	GNA12	rs798502	C	A	0.28
chr7	IKZF1	rs1456896	C	T	0.27
chr7	IL6	rs1800795	C	G	0.35
chr7	IRF5	rs3807306	T	G	0.42
chr7	IRF5	rs4728142	A	G	0.39
chr7	MAGI2	rs2160322	G	C	0.34
chr7	MMD2	rs4724190	A	G	0.42
chr7	NOD1	rs2075818	C	G	0.25
chr7	NOD1	rs2235099	A	G	0.24
chr7	NOD1	rs2906766	C	T	0.30
chr7	NOD1	rs2907748	T	C	0.27
chr7	NOD1	rs6958571	C	A	0.27
chr7	PIK3CG	rs1526083	G	A	0.38
chr7	PIK3CG	rs342293	G	C	0.41
chr7	SLC26A3	rs2108225	A	G	0.46
chr7	SLC26A3	rs4598195	C	A	0.42
chr7	SLC26A3	rs4730273	A	C	0.30
chr7	SLC26A3	rs4730276	A	G	0.38

chr7	TNPO	rs10488631	C	T	0.12
chr8	BLK	rs2736340	T	C	0.22
chr8	DEFB1	rs11362	T	C	0.43
chr8	DEFB4	rs9720835	A	G	0.11
chr8	FBXO43	rs2453626	T	C	0.39
chr8	RIPK2	rs43134	A	G	0.45
chr8	RIPK2	rs447618	C	T	0.45
chr8	TMEM75	rs6651252	C	T	0.13
chr8	TRIB1	rs1551398	G	A	0.37
chr9	CARD9	rs10781499	A	G	0.46
chr9	IL33	rs1342326	C	A	0.19
chr9	IL33	rs1929992	C	T	0.28
chr9	JAK2	rs10758669	C	A	0.37
chr9	JAK2	rs10974944	G	C	0.29
chr9	JAK2	rs10975003	C	T	0.31
chr9	JAK2	rs1536800	T	C	0.25
chr9	RPS2P34	rs668853	C	T	0.38
chr9	TLR4	rs7864330	G	T	0.06
chr9	TNFSF15	rs10759734	G	A	0.24
chr9	TNFSF15	rs3810936	T	C	0.26
chr9	TNFSF15	rs4979459	G	T	0.50
chr9	TNFSF15	rs6478109	A	G	0.26
chr9	TNFSF15	rs7869487	C	T	0.26
chr9	TRAF1	rs3761847	G	A	0.47
chr10	C10orf67	rs1398024	T	G	0.26
chr10	CCNY	rs3936503	A	G	0.35
chr10	CREM	rs12242110	G	A	0.37
chr10	DLG5	rs1248694	T	C	0.31
chr10	DLG5	rs2165047	T	C	0.29
chr10	GPR120	rs17484310	A	T	0.10
chr10	IL2RA	rs12722489	T	C	0.09
chr10	KLF6	rs6601764	C	T	0.44
chr10	MBL2	rs11003123	A	G	0.20
chr10	MBL2	rs1800450	T	C	0.15
chr10	MBL2	rs4935047	A	G	0.41
chr10	MBL2	rs7084554	C	T	0.20
chr10	NKX2_3	rs10883365	A	G	0.40
chr10	IATPR	rs2755996	T	C	0.10
chr10	PRF1	rs3758562	G	A	0.33
chr10	PRKCQ	rs4750316	C	G	0.15
chr10	RPS12P16	rs17582416	G	T	0.42
chr10	SGMS1	rs3011770	C	T	0.22
chr10	TCF7L2	rs3814570	T	C	0.26
chr10	UBE2D1	rs1819658	T	C	0.18
chr10	ZMIZ1	rs1250550	A	C	0.30
chr10	ZNF365	rs10761659	A	G	0.45

chr10	ZNF365	rs10995271	C	G	0.37
chr11	C11orf30	rs2155219	G	T	0.43
chr11	C11orf30	rs7130588	G	A	0.34
chr11	C11orf30	rs7927894	T	C	0.40
chr11	CADM1	rs11601041	C	T	0.46
chr11	CADM1	rs11607801	A	C	0.47
chr11	CADM1	rs4938200	T	C	0.38
chr11	CADM1	rs220874	C	T	0.41
chr11	ETS1	rs7117768	G	C	0.27
chr11	FADS1	rs102275	C	T	0.36
chr11	FAM55A	rs661946	T	C	0.34
chr11	IL18	rs2043055	G	A	0.39
chr11	IRF7	rs11246213	G	A	0.22
chr11	PHRF1	rs4963128	T	C	0.29
chr11	LSP1	rs11041476	A	G	0.36
chr11	LSP1	rs907611	A	G	0.36
chr11	MAML2	rs483905	A	G	0.27
chr11	MAML2	rs543104	A	G	0.27
chr11	NELL1	rs1793004	C	G	0.24
chr11	LINC02714	rs11223996	C	T	0.12
chr11	NT_167190	rs1892953	G	A	0.38
chr11	NUCB2	rs10766384	G	A	0.28
chr11	PRDX5	rs694739	G	A	0.35
chr11	RELA	rs1049728	C	G	0.05
chr11	RPS6KB2	rs1476792	C	T	0.44
chr11	TRAF6	rs5030411	G	A	0.41
chr12	APAF1	rs17028658	C	T	0.07
chr12	APAF1	rs2288729	A	G	0.35
chr12	ATXN2	rs616668	G	T	0.19
chr12	ATXN2	rs653178	T	C	0.45
chr12	GLI1	rs10783827	G	T	0.29
chr12	IFNG	rs1558744	A	G	0.41
chr12	IFNG	rs2069727	C	T	0.43
chr12	IL23A	rs2066808	G	A	0.07
chr12	IL26	rs2870946	C	T	0.09
chr12	LRRK2	rs11175593	T	C	0.03
chr12	RASSF8	rs16929496	C	T	0.21
chr12	SH2B3	rs3184504	C	T	0.45
chr12	SH2B3	rs739496	G	A	0.19
chr12	TCF1	rs2393791	C	T	0.42
chr12	TRHDE	rs12831974	C	T	0.13
chr13	LACC1	rs3764147	G	A	0.27
chr13	FOXO1	rs941823	T	C	0.21
chr13	FOXO1	rs9548988	C	T	0.51
chr13	HMGB1	rs1045411	T	C	0.23
chr13	SMAD9	rs12855930	T	C	0.30

chr13	TBC1D4	rs11617463	A	C	0.08
chr13	TNFSF11	rs2062305	G	A	0.49
chr13	USP12	rs17085007	C	T	0.15
chr14	AKT1	rs2494731	C	G	0.31
chr14	GALC	rs8005161	T	C	0.08
chr14	HIF1A	rs2301113	C	A	0.24
chr14	LGALS3	rs4652	C	A	0.45
chr15	AGBL1	rs1431242	T	C	0.25
chr15	HERC2	rs916977	T	C	0.22
chr15	PIAS1	rs8038236	T	C	0.41
chr15	SMAD3	rs17293632	T	C	0.25
chr16	BRD7	rs11644386	C	T	0.16
chr16	BRD7	rs4785412	C	T	0.16
chr16	RMI2	rs243327	G	A	0.49
chr16	CDH1	rs1728785	A	C	0.20
chr16	CIITA	rs4781011	T	G	0.28
chr16	CIITA	rs6498131	T	C	0.41
chr16	CYLD	rs3135503	G	T	0.31
chr16	CYLD	rs8060598	C	T	0.31
chr16	FAM92B	rs8050910	G	T	0.38
chr16	GNPTG	rs3826051	T	C	0.37
chr16	IL27	rs151181	C	T	0.40
chr16	IL27	rs8049439	C	T	0.39
chr16	IRF8	rs16940186	C	T	0.27
chr16	IRF8	rs305080	T	C	0.32
chr16	IRF8	rs903202	C	T	0.43
chr16	ITGAM	rs13338069	G	A	0.12
chr16	NOD2	rs17221417	C	G	0.46
chr16	NOD2	rs2066843	C	T	0.47
chr16	NOD2	rs2066844	T	C	0.17
chr16	NOD2	rs2066845	C	G	0.07
chr16	NOD2	rs2066847	I	D	0.13
chr16	NOD2	rs2076756	G	A	0.50
chr16	NOD2	rs5743289	T	C	0.39
chr16	PRKCB1	rs2106375	C	A	0.17
chr16	PRKCB1	rs7404095	T	C	0.45
chr16	PRKCB1	rs8056879	A	G	0.38
chr16	SNX20SLIC1	rs7202124	G	A	0.17
chr16	SOCS1	rs193779	A	G	0.24
chr16	TNP2	rs415595	G	A	0.46
chr16	TRAF7	rs2078282	A	G	0.27
chr16	UBE2I	rs12446893	G	A	0.13
chr17	AURKB	rs2289590	C	A	0.43
chr17	BIRC4BP	rs1533031	G	A	0.36
chr17	CCL2	rs3091315	G	A	0.21
chr17	desert	rs991804	T	C	0.21

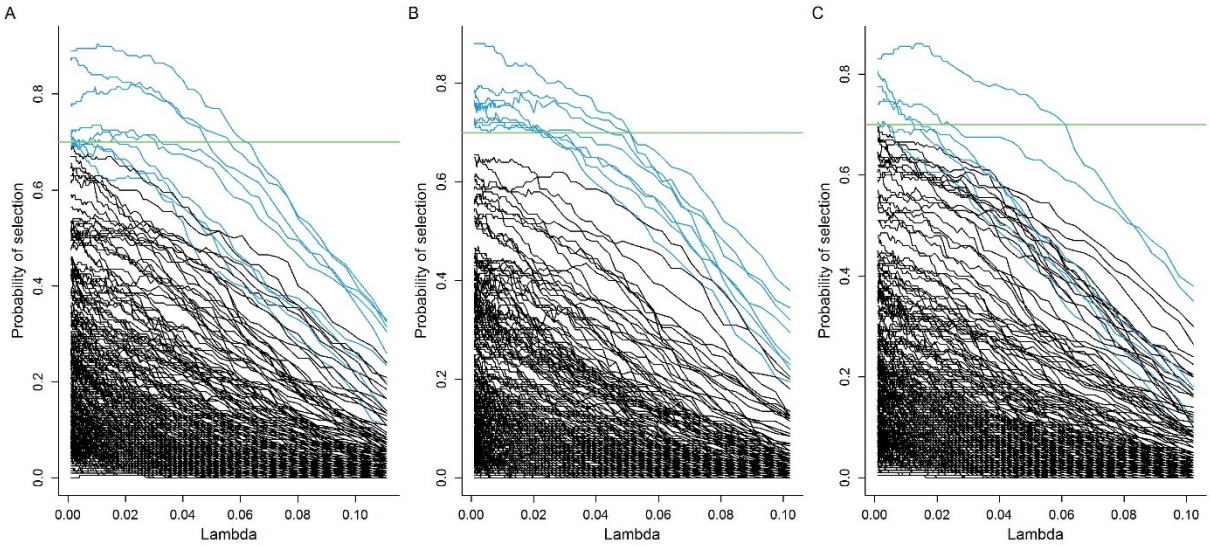
chr17	EIF5A	rs11658072	C	T	0.42
chr17	IBD22	rs12948909	C	A	0.21
chr17	IKZF3	rs907091	T	C	0.50
chr17	NOS2	rs2274894	T	G	0.42
chr17	NOS2	rs2872753	G	A	0.44
chr17	NT_010799	rs1676578	G	A	0.38
chr17	ORMDL3	rs2305480	A	G	0.49
chr17	ORMDL3	rs8076131	G	A	0.49
chr17	RPS6KB1	rs180530	A	G	0.50
chr17	STAT3	rs744166	G	A	0.35
chr18	PTPN2	rs2542151	G	T	0.14
chr18	PTPN2	rs2847281	G	A	0.40
chr18	PTPN2	rs2847286	G	A	0.37
chr19	BCL3	rs2927488	A	G	0.29
chr19	CARD8	rs2043211	T	A	0.35
chr19	CARD8	rs6509366	A	G	0.36
chr19	CD209	rs11465421	G	T	0.48
chr19	CEACAM6	rs10416839	T	G	0.35
chr19	CEACAM6	rs11669653	A	T	0.29
chr19	CEACAM6	rs8106277	C	T	0.43
chr19	FUT2	rs281379	G	A	0.50
chr19	ICAM3	rs7257871	C	T	0.24
chr19	IL12RB1	rs17852635	A	G	0.32
chr19	KEAP1	rs11085735	A	C	0.09
chr19	MYO9B	rs2305767	C	T	0.48
chr19	NFKBIB	rs3136644	G	A	0.22
chr19	NLRP12	rs10418264	C	T	0.43
chr19	NLRP12	rs34436714	A	C	0.27
chr19	nuclear_factorIC	rs4807451	G	A	0.06
chr19	RSHL1	rs12463359	T	G	0.39
chr19	SBNO2	rs740495	G	A	0.28
chr19	SLC7A10	rs10500264	A	G	0.19
chr19	SLC7A10	rs736289	C	T	0.29
chr19	TYK2	rs12720356	C	A	0.11
chr19	TYK2	rs280519	G	A	0.48
chr20	CD40	rs1883832	T	C	0.28
chr20	CEBPB	rs6095811	T	C	0.37
chr20	FOXA2	rs1203905	G	A	0.06
chr20	HNF4A	rs6017342	A	C	0.40
chr20	TNFRSF6B	rs2315008	T	G	0.22
chr21	AIRE	rs878081	T	C	0.26
chr21	ETS2	rs7282723	T	C	0.36
chr21	FLJ45139	rs2836754	T	C	0.35
chr21	ICOSLG	rs378299	C	T	0.46
chr21	ICOSLG	rs7278940	T	C	0.30
chr21	ICOSLG	rs762421	G	A	0.39

chr21	NRIP1	rs1736135	C	T	0.40
chr22	C1QTNF6	rs229522	A	G	0.29
chr22	C1QTNF6	rs229527	A	C	0.46
chr22	HMOX1	rs2071749	A	G	0.45
chr22	IL2RB	rs2284033	A	G	0.38
chr22	MAP3K7IP1	rs2413583	T	C	0.15
chr22	MTMR3	rs713875	C	G	0.48
chr22	NCF4	rs4821544	C	T	0.38
chr22	PNPLA3	rs738409	G	C	0.21
chr22	RAC2	rs739041	G	A	0.43
chr22	UBE2L3	rs140489	A	G	0.23
chr22	XBP1	rs35873774	C	T	0.06
chrX	TRMT2B	rs4827884	A	G	0.38
chrX	FOXP3	rs5915330	G	A	0.41
chrX	IKBKG	rs4526543	G	A	0.39
chrX	MAGEB6	rs4898186	A	T	0.45
chrX	NLGN4X	rs4352986	A	G	0.38
chrX	PABPC5	rs1029318	A	C	0.42
chrX	PABPC5	rs11093286	A	G	0.42
chrX	TLR7	rs179009	G	A	0.24
chrX	TMSL3	rs1483191	A	G	0.45

Statistical analysis:

Lasso logistic regression: Lasso regression allows to model the relationship between genetic variants and the phenotype of interests and performs selection of variables at the same time than estimation of the multivariable model.[1] Since estimation of regularization parameter λ by cross-validation tends to select a too high number of variables we used "stability selection" to select relevant variables and to avoid overfitting : we randomly selected 200 bootstrap samples of size n (number of patients) by sampling with replacement in the original dataset.[2] The idea is that relevant variables will be more often selected in replicated samples than non-relevant variables that are likely to be more sensitive to resampling. Selected variables were those achieving a probability of selection of more than 70% over the 200 bootstrap replications. Stability paths are presented on Supplementary Figure 1. Lasso regression was performed using the R glmnet package.

Supplementary Figure S1: Stability paths for the selection of genetic variants by lasso with stability selection at 0.7 threshold. This figure represents selection probabilities over 200 bootstrap replications for each variable in function of the penalization parameter λ of the lasso. Each curve is associated to one variable. Variables reaching the 0.7 threshold (horizontal green line) were selected. These selected variables are represented in blue. A) Composite outcome at 5 years B) Outcome surgery at 5 years C) Outcome complicated of behavior at 5 years.



Adaptive logistic lasso: After separate selections of clinical, serological and genetic variants, the final models were performed using adaptive lasso.[3] Adaptive lasso is a modified version of lasso regression that performs better selection of the relevant variables and optimal estimation of the variable coefficients. Adaptive weight vector was given by $\hat{w} = 1/|\hat{\beta}|$ with $\hat{\beta}$ obtained from ridge regression. Optimal lasso penalisation parameter λ was estimated by 5-fold cross-validation and maximization of the area under the curve (AUC).

Discriminative performances and internal validation: Internal validation were done by using bootstrap resampling with 1000 repetitions to estimate the AUC corrected for over-optimism and the shrinkage factors.[4,5]

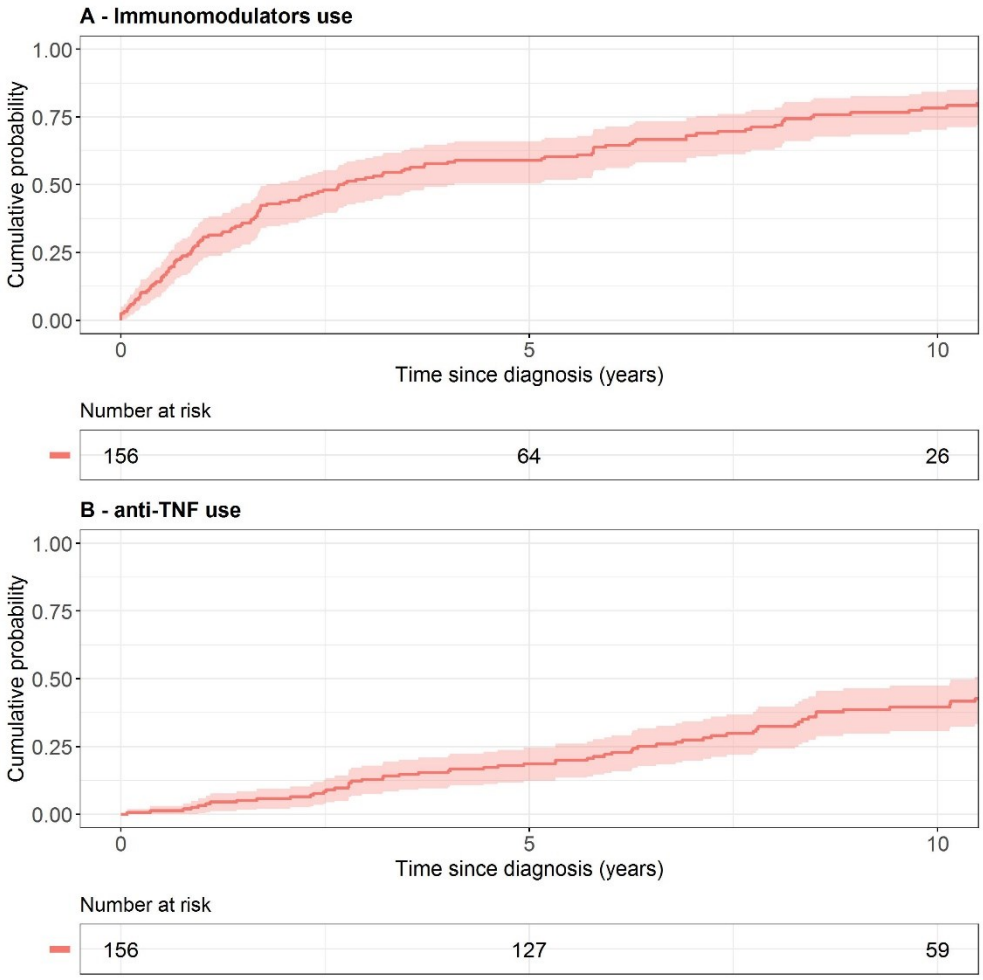
The shrinkage factor is the estimated shrinking in the regression coefficients to improve the prediction in future patients. Apparent predictive performances are performances calculated directly on the original data set (i.e. the original data set is used both as training and test set). Predictive performances were corrected for optimism bias using 1000 bootstrap samples. Bootstrap samples were drawn with replacement from the original data set. The adaptive lasso regression model was estimated on each bootstrap sample and the deriving scoring system was then applied both to the original data and to the bootstrap sample (i.e. the bootstrap sample here referred to the training sample and the original data referred to the test sample). For each bootstrap replication, an estimation of optimism bias was calculated as the difference between performances in the original and bootstrap samples. Optimism bias is then averaged over the 1000 bootstrap samples and subtracted from apparent performance in order to give the performance corrected for optimism bias.

Decision curve analysis:

The net benefit is calculated as (proportion of true positives) – (proportion of false positives)*pt/(1−pt), where pt is the threshold probability.[6,7] Net benefit was standardized with observed event rate in our sample so that maximal standardized benefit is equal to one. As recommended, the decision curve was only plotted for a reasonable range of threshold probabilities. This range was chosen by discussion with expert gastroenterologists.

Supplementary Results:

Supplementary Figure S2 : Cumulative probability of medical therapy exposure in 156 patients with pediatric-onset Crohn's disease. (A) Immunomodulators, (B) anti–tumor necrosis factor (TNF).



Supplementary table S2: External cohort's clinical characteristics (n=60)

Variable	n (%)
Follow-up in years (median, IQR)	7.8 [5.0 -12.3]
Male gender	34 (56.7 %)
Clinical data at diagnosis	
Age (median, IQR)	12.3 [10.2 – 15.4]
Familial antecedents of IBD	10 (16.7 %)
Location at diagnosis	
Ileal (L1)	8 (13.8 %)
Colonic (L2)	18 (31.0 %)
Ileo-colonic (L3)	32 (55.2 %)
Upper digestive disease (L4)	14 (23.3 %)
Ano-perineal lesions	22 (36.7 %)
Extra-intestinal symptoms	18 (30.0 %)
Clinical outcomes at 5 years	
Surgery	33 (58.9 %)
Complicated behavior (B2/B3) †	14 (35.9 %)
Composite outcome : surgery and/or complicated behavior	37 (64.9 %)

† date of change in behavior was unknown so denominator included only patients for whom we were able to assess a change of behavior during the first 5 years of the disease. IQR : interquartile range.

Supplementary Table S3: Odds ratios of selected SNPs in the pooled sample of discovery and external cohort.

Outcome	Selected variants†	OR and 95% CI	p-value
Composite outcome at 5 years	chr1_PHC2_rs4653044	1.26 [0.86 ; 1.85]	0.23
	chr1_TLR5_rs851192	0.70 [0.46 ; 1.06]	0.09
	chr10_IATPR_rs2755996	2.29 [1.12 ; 4.67]	0.02
	chr10_UBE2D1_rs1819658	0.55 [0.32 ; 0.96]	0.03
	chr13_FOXO1_rs9548988	0.80 [0.53 ; 1.20]	0.28
	chr13_TNFSF11_rs2062305	1.95 [1.29 ; 2.96]	<0.01
	chr17_IKZF3_rs907091	1.72 [1.14 ; 2.68]	0.02
	chr4_NFKB_rs230530	0.56 [0.38 ; 0.84]	<0.01
	chr7_IKZF1_rs1456896	2.27 [1.40 ; 3.67]	<0.001
Intestinal resection at 5 years	chr1_PHC2_rs4653044	1.31 [0.86 ; 1.99]	0.21
	chr12_LRRK2_rs111755	3.35 [0.95 ; 11.75]	0.06
	chr13_FOXO1_rs954898	0.76 [0.49 ; 1.18]	0.22
	chr13_HMGB1_rs1045411	0.34 [0.18 ; 0.66]	<0.01
	chr16_NOD2_rs2066845	2.24 [1.02 ; 4.88]	0.04
	chr17_AURKB_rs228959	1.49 [0.09 ; 0.36]	0.08
	chr17_ORMDL3_rs80761	0.55 [0.35 ; 0.88]	0.01
	chr3_IHPK1_rs9872864	0.50 [0.31 ; 0.81]	<0.01
Complicated behavior at 5 years	chr1_ECM1_rs11205387	1.23 [0.74 ; 2.02]	0.42
	chr1_TLR5_rs851192	0.62 [0.39 ; 0.99]	0.04
	chr10_KLF6_rs6601764	0.50 [0.32 ; 0.79]	<0.01
	chr14_AKT1_rs2494731	1.72 [1.05 ; 2.84]	0.03
	chr6_PRDM1_rs548234	0.60 [0.37 ; 0.98]	0.04
	chr7_IKZF1_rs1456896	1.85 [1.23 ; 3.05]	0.01

† variants selected using lasso and stability selection at 0.7 threshold in discovery (Epimad) cohort.
Variants highlighted in grey are those achieving p≤0.05 in pooled sample and included in final multivariable models. OR : Odds ratio calculated using univariable logistic regression ; 95% CI : 95% confidence interval.

Supplementary Table S4 : Final models coefficients. Best predictive model is presented in bold (PREDICT-EPIMAD).

Chromosome	candidate gene	SNP	Response Modalities	Coding	β^{+} coefficients - model 1	β coefficients - model 2 (without pAnca)
Main outcome : composite outcome at 5 years						
Intercept					**-1.948**	-2.001
chr4	NFKB	rs230530	GG, GA, AA	2,1,0	**-0.665**	-0.722
chr7	IKZF1	rs1456896	CC, CT, TT	2,1,0	**1.047**	1.000
chr10	UBE2D1	rs1819658	TT, TC, CC	2,1,0	**-1.248**	-1.145
chr10	IATPR	rs2755996	TT, TC, CC	2,1,0	**1.115**	1.045
chr13	TNFSF11	rs2062305	GG, GA, AA	2,1,0	**0.808**	0.767
chr17	IKZF3	rs907091	TT, TC, CC	2,1,0	**0.958**	0.923
Disease location			ileal (L1)/ colonic(L2)/ ileo-colonic (L3)	0/1/0	**-1.078**	-1.139
pAnca			positive/negative	1/0	**-1.828**	
				Threshold*	**0.35**	0.28
				Shrinkage†	**0.88**	0.89
Outcome : Surgery at 5 years						
Intercept					-0.993	-1.163
chr3	IHPK1	rs9872864	GG, GA, AA	0,1,2	0.670	0.8404
chr13	HMGB1	rs1045411	TT, TC, CC	2,1,0	-0.971	-1.169
chr16	NOD2	rs2066845	CC, CG, GG	2,1,0	0.968	1.092
chr17	ORMDL3	rs8076131	GG, GA, AA	2,1,0	-0.692	-0.876
pAnca			positive/negative	1/0	-1.181	
				Threshold*	0.15	0.21
				Shrinkage†	1.02	1.12
Outcome : Complication of behavior at 5 years						
Intercept					0.105	-0.258
chr1	TLR5	rs851192	GG, GC, CC	2,1,0	-0.783	-0.758
chr6	PRDM1	rs548234	CC, CT, TT	2,1,0	-1.126	-0.979
chr7	IKZF1	rs1456896	CC, CT, TT	2,1,0	0.853	0.829
chr10	KLF6	rs6601764	CC, CT, TT	2,1,0	-0.603	-0.593
chr14	AKT1	rs2494731	CC, CG, GG	2,1,0	0.693	0.730
pAnca			positive/negative	1/0	-1.773	
				Threshold*	0.36	0.23
				Shrinkage†	0.98	0.99

$^{+}$ β coefficients calculated using adaptive lasso logistic regression model; ***** Thresholds maximizing both sensitivity and specificity by means of Youden's J statistic are calculated using AUC in discovery sample; † Shrinkage factor were calculated using 1000 bootstrap samples.

Supplementary references

1 Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Ser B Methodol* 1996; 58: 267–88.

2 Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Ser B Stat Methodol* 2010; 72: 417–73.

3 Zou H. The Adaptive Lasso and Its Oracle Properties. *J Am Stat Assoc* 2006; 101: 1418–29.

4 Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; 338. DOI:10.1136/bmj.b605.

5 Dunkler D, Sauerbrei W, Heinze G. Global, Parameterwise and Joint Shrinkage Factor Estimation. *J Stat Softw* 2016; 69: 1–19.

6 Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; 352: i6.

7 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak Int J Soc Med Decis Mak* 2006; 26: 565–74.