

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Caractérisation épidémiologique de la maladie de Crohn au Québec

Par

Anne-Marie Lowe

Programme de Sciences Biomédicales

Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures
En vue de l'obtention du grade de « Maître es sciences » (M.Sc.)
En Sciences Biomédicales
Option générale

Janvier 2008

© Anne-Marie Lowe 2008



Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Caractérisation épidémiologique de la maladie de Crohn au Québec

Présenté par :

Anne-Marie Lowe

A été évalué par un jury composé des personnes suivantes :

.....

Maria Victoria Zunzunegui (PhD), présidente-rapporteuse

.....

Paul Brassard (MD), directeur de recherche

.....

Paul Rivest (MD), membre du jury

Résumé

Une étude épidémiologique de la maladie de Crohn (MC) au Québec a été faite afin de caractériser au niveau populationnel l'incidence de cette maladie. En employant une définition opérationnelle de cas adaptée pour le Québec, 10 065 cas ont été identifiés à partir de banques de données administratives provinciales pour la période d'étude (1995-2000). La validation de la définition opérationnelle de cas a révélé que sa spécificité est de 83% et sa sensibilité est de 97%. L'incidence moyenne de la MC au Québec, standardisée pour l'âge et le sexe, est de 22,4 nouveaux cas / 10⁵ personnes / année. Des variables potentiellement associées à la MC ont été modélisées selon le modèle de régression de Poisson afin d'expliquer la variation des taux d'incidence de la MC au Québec. Huit variables ont été associées de manière significative à la MC. Les variables significatives ont été ajustées les unes par rapport aux autres dans un modèle multivarié à partir duquel trois variables ont été associées de manière significative à la MC. Ainsi, une forte proportion d'immigrants dans la population est associée négativement à la MC ; l'incidence de cinq maladies entériques à déclaration obligatoire est associée positivement à la MC, tout comme une forte proportion d'individus de descendance juive. Ces variables ont permis d'expliquer 21% de la variation de l'incidence de la MC au Québec. D'autres facteurs tels la susceptibilité génétique à la MC ou l'effet d'un microorganisme pathogène présent dans l'environnement devraient éventuellement être pris en considération afin d'expliquer la variation résiduelle.

Mots-clés : épidémiologie, maladie de Crohn, incidence, validation, géographie, régression de Poisson, *Mycobacterium avium* sp *paratuberculosis*

Abstract

An epidemiological study of Crohn's disease (CD) in Quebec has been done to characterize the incidence of this disease at an ecological level. By using a case definition adapted for Quebec, 10 065 cases have been identified from provincial administrative databases for the study period (1995-2000). The operational case definition has been validated and has revealed a specificity of 83% and a sensitivity of 97%. The mean standardized incidence of CD for age and gender in Quebec is 22.4 new cases / 10^5 persons / year. Potentially associated variables to CD have been analyzed by using the Poisson regression model to explain CD incidence rates variation in Quebec. Eight variables have been significantly associated to CD. The significant variables have been adjusted for each other in a multivariate model from which three variables have emerged. Therefore, a high proportion of immigrants in the population is negatively associated to CD, the incidence of five reportable enteric disease is positively associated to CD as well as a high proportion of people of Jewish descent in the population. These variables taken together explain 21% of CD incidence variance in the Quebec population. Other factors as genetic susceptibility to CD or the effect of an environmental pathogen should eventually be taken into consideration in the models to explain the residual variance.

Keywords : epidemiology, Crohn's disease, incidence, validation, geography, Poisson regression, *Mycobacterium avium* sp *paratuberculosis*

Table des matières

<i>Résumé</i>	<i>iii</i>
<i>Abstract</i>	<i>iv</i>
<i>Table des matières</i>	<i>v</i>
<i>Liste des tableaux</i>	<i>viii</i>
<i>Liste des figures</i>	<i>ix</i>
<i>Liste des équations</i>	<i>x</i>
<i>Liste des sigles et abréviations</i>	<i>xi</i>
<i>Remerciements</i>	<i>xiii</i>
<i>Introduction</i>	<i>1</i>
<i>Partie 1 : Recension de la littérature</i>	<i>4</i>
1.1 Description de la maladie de Crohn	4
1.1.1 Signes cliniques et histopathologiques.....	4
1.1.2 Modèle épidémiologique de causalité.....	5
1.2 Épidémiologie de la maladie de Crohn	6
1.2.1 Incidence et prévalence	6
i) Épidémiologie dans le monde	6
ii) Épidémiologie en Amérique du Nord.....	7
1.2.2 Variations géographiques de l'incidence de la maladie de Crohn.....	9
1.3 Prédiposition de l'hôte à la maladie de Crohn	11
1.3.1 Gènes associés à la maladie de Crohn.....	11
1.3.2 Le gène <i>NOD2/CARD15</i> et son rôle dans la reconnaissance bactérienne.....	11
1.3.2 Le sexe.....	13
1.3.2 L'âge au moment du diagnostic.....	14
1.4 Facteurs potentiellement associés à la maladie de Crohn : variables individuelles, géographiques, socioéconomiques et reliées à la condition de santé	15
1.4.1 Lieu de résidence – urbanité, ruralité.....	15
1.4.2 Tabagisme.....	16
1.4.3 Utilisation de contraceptifs oraux	17
1.4.4 Statut socio-économique	19
1.4.5 Implication des maladies entériques dans le développement de la MC et hypothèse hygiénique.....	19
i) L'hypothèse hygiénique.....	20
ii) Flore intestinale et développement du système immunitaire.....	20
iii) Infections gastro-intestinales à l'enfance.....	21
iv) L'infection par un microorganisme et la MC.....	22
v) Maladies entériques à déclaration obligatoire et MC	22
1.4.6 Origine ethnique – immigration, autochtones, descendance juive	23
1.4.7 Agent pathogène : <i>Mycobacterium avium</i> sp. <i>paratuberculosis</i>	25
<i>Partie 2 : Hypothèses et objectifs spécifiques</i>	<i>30</i>
2.1 Hypothèses	30
2.1.1 Épidémiologie de la maladie de Crohn et variables associées	30

2.1.2 Définition opérationnelle de cas	30
2.2 Objectifs	30
<i>PARTIE 3 : Méthodologie.....</i>	32
3.1 Données	32
3.1.1 Identification des cas de maladie de Crohn	32
3.1.2 Définition opérationnelle de cas	32
3.1.3 Validation des cas identifiés par la définition de cas	33
3.1.4 Données de population à risque	36
3.1.5 Données géographiques – régions de CLSC	37
3.1.6 Données démographiques et associées à la condition de santé	37
3.1.7 Source des données et définition des banques de données.....	38
3.2 Analyse descriptive	41
3.2.1 Devis descriptif – l'étude écologique.....	41
3.2.2 Taux d'incidence de la maladie de Crohn au Québec.....	42
3.2.3 Standardisation directe des taux d'incidence pour l'âge et le sexe.....	44
3.2.4 Visualisation des données d'incidence.....	45
3.3 Analyses associatives.....	46
3.3.1 Variables indépendantes principales	46
3.3.2 Définition des variables principales	48
i) Immigration.....	48
ii) Ascendance juive.....	48
iii) Population autochtone	48
iv) Contraceptifs oraux.....	49
v) Urbanité.....	50
vi) Incidence de cinq maladies entériques à déclaration obligatoire	50
vii) Tabagisme	51
viii) Statut socio-économique - indice de défavorisation.....	51
3.4 Modélisation des données	52
3.4.1 Modèle général de régression	52
3.4.2 Stratégie générale de modélisation	54
3.4.3 Modèles de régression : analyses univariées	55
3.4.4 Modèles de régression : analyses multivariées	56
3.4.5 Colinéarité entre les variables.....	56
3.4.6 Corrélation entre les variables	57
3.4.7 Modélisation des mesures répétées et des données corrélées dans l'espace .	57
3.4.8 Calcul du pseudo- R^2 du modèle de régression de Poisson.....	58
3.5 Tests statistiques des analyses	59
3.6 Seuil alpha	59
3.7 Logiciels utilisés	59
<i>PARTIE 4 : Résultats - présentation et analyse.....</i>	60
4.1 Épidémiologie de la maladie de Crohn au Québec – analyses descriptives ..	60
4.1.1 Nombre de cas de maladie de Crohn identifiés	60
4.1.2 Validation et rendement de la définition opérationnelle de cas.....	60
4.1.3 Incidence moyenne de la maladie de Crohn au Québec.....	61
4.1.4 Visualisation des données d'incidence.....	62
4.2.1 Analyses univariées.....	69
4.2.2 Analyses multivariées	71

	vii
i) modèle de régression multivarié complet	71
ii) modèle de régression multivarié réduit.....	72
4.2.3 Pseudo-R ² du modèle final de régression de Poisson	73
<i>PARTIE 5 : Discussion</i>	74
<i>Conclusion</i>	87
<i>Sources documentaires</i>	88
<i>Annexes</i>	<i>i</i>
Annexe 1 : Liste des variables principales et sous-variables dans la base de données	i
Annexe 2 : Calculs de la standardisation directe	xi
Annexe 3 : Calculs taux de tabagisme	xii
Annexe 4 : Légende d’attribution de l’indice de défavorisation – statut socioéconomique	xiv
Annexe 5 : Recodification des CLSC selon les modifications faites par le MSSS	xv
Annexe 6 : Procédure GENMOD de SAS	xxi
Annexe 7 : Correspondance entre RSS et CLSC	xxiii

Liste des tableaux

Tableau I :	Incidence et prévalence de la maladie de Crohn dans le monde ...	p. 7
Tableau II :	Incidence et prévalence de la maladie de Crohn en Amérique du Nord.....	p. 9
Tableau III :	Tableau 2X2 utilisé afin de mesurer la validité de la définition de cas.....	p. 35
Tableau IV :	Description des variables indépendantes principales.....	p. 47
Tableau V :	Tableau 2X2 - résultats de la validation de la définition de cas.....	p. 60
Tableau VI :	Incidence de la MC dans six provinces canadiennes.....	p. 62
Tableau VII :	CLSC ayant les taux d'incidence standardisés les plus élevés au Québec.....	p. 63
Tableau VIII :	CLSC ayant les taux d'incidence standardisés les plus faibles au Québec.....	p. 63
Tableau IX :	IRR bruts des modèles univariés incluant les variables « âge » et « sexe ».....	p. 69
Tableau X :	IRR bruts et ajustés de modèles univariés et multivariés – association de variables à la maladie de Crohn	p. 70
Tableau XI :	IRR ajustés du modèle multivarié réduit qui inclut les variables significatives du modèle complet	p. 73

Liste des figures

Figure 1 :	Modèle de causalité de la triade épidémiologique	p. 5
Figure 2 :	Triade épidémiologique de causalité - l'hôte	p. 11
Figure 3 :	Triade épidémiologique de causalité - les facteurs environnementaux	p. 15
Figure 4 :	Triade épidémiologique de causalité - implication d'un agent pathogène	p. 25
Figure 5 :	Stratégie d'introduction des variables dans les modèles de régression	p. 54
Figure 6 :	Répartition géographique des taux d'incidence de la maladie de Crohn par régions de CLSC, province de Québec (1995- 2000)	p. 64
Figure 7 :	Répartition géographique des taux d'incidence de la maladie de Crohn par régions de CLSC, excluant les CLSC du nord, zoom sur l'île de Montréal, province de Québec (1995-2000)	p. 65
Figure 8 :	Répartition géographique des taux d'incidence de la maladie de Crohn par régions de CLSC <i>urbains</i> , province de Québec (1995-2000)	p. 67
Figure 9 :	Répartition géographique des taux d'incidence de la maladie de Crohn par régions de CLSC <i>ruraux</i> , province de Québec (1995-2000)	p. 68

Liste des équations

Équation 1 :	Calcul de sensibilité selon le tableau 2X2	p. 35
Équation 2 :	Calcul de spécificité selon le tableau 2X2	p. 36
Équation 3 :	Calcul de la valeur prédictive positive selon le tableau 2X2	p. 36
Équation 4 :	Calcul de la valeur prédictive négative selon le tableau 2X2	p. 36
Équation 5 :	Calcul du taux d'incidence brut	p. 42
Équation 6 :	Calcul du taux d'incidence brut par groupe d'âge	p. 43
Équation 7 :	Calcul du taux d'incidence brut selon le sexe	p. 43
Équation 8 :	Loi de Poisson	p. 53
Équation 9 :	Calcul du ratio de taux d'incidence	p. 55
Équation 10 :	Relation entre la variable indépendante et dépendante – modèle univarié	p. 55
Équation 11 :	Relation entre les variables dépendantes et indépendantes – modèle multivarié	p. 56
Équation 12 :	Calcul du pseudo- R^2 de la régression de Poisson	p. 58
Équation 13 :	Modèle de régression multivarié complet	p. 71
Équation 14 :	Modèle de régression multivarié réduit	p. 72

Liste des sigles et abréviations

AB :	Alberta
ACP :	Amplification en chaîne par polymérase
ADN :	Acide désoxyribonucléique
BAAR :	(anglais) Acid-fast bacilli, bacilles acido-alcoolo résistant
BCG :	Bacille de Calmette-Guérin
CAI :	Commission d'accès à l'information
CB:	Colombie-Britannique
CIM:	Classification internationale des maladies
CLSC :	Centre local de service communautaire
CO :	Contraceptifs oraux
CU:	Colite ulcéreuse
CUSM :	Centre universitaire de santé McGill
DEV :	Déviance
DSP :	Direction de la Santé publique
ECCO :	(anglais) European Crohn's and Colitis Organisation
EXP :	Exponentielle
GALT:	Tissus lymphoïdes associés au tube digestif
GEE :	(anglais) Generalized estimating equations
GREZOSP :	Groupe de recherche en épidémiologie des zoonoses et santé publique
IC :	Intervalle de confiance
IRR:	(anglais) Incidence rate ratio, rapport de taux d'incidence
kg:	Kilogramme
km:	Kilomètre
LZA-SSH:	Laboratoire de lutte contre les zoonoses d'origine alimentaire – section Saint-Hyacinthe
LZA-ST:	Laboratoire de lutte contre les zoonoses d'origine alimentaire – section Toronto
log :	Logarithme
LSPQ :	Laboratoire de Santé publique du Québec
MAA:	<i>Mycobacterium avium</i> subsp. <i>avium</i>
MAC:	Complexe <i>Mycobacterium avium</i>
MADO :	Maladie à déclaration obligatoire
MAP :	<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i>

MB :	Manitoba
MC :	Maladie de Crohn
Md :	Médiane
mg :	Milligramme
MII:	maladie inflammatoire des intestins
MMR :	(anglais) Measles, mumps, rubella; rougeole, oreillons, rubéole
MN :	Minnesota
MSSS :	Ministère de la Santé et des Services sociaux
NAD :	(anglais) North American Datum
NAM :	Numéro d'assurance maladie
NÉ:	Nouvelle-Écosse
OR :	(anglais) Odds ratio, rapport de cotes
Ord. :	Ordonnance
p :	(anglais) p value, valeur de p
PCR:	(anglais) Polymerase chain reaction, amplification en chaîne par polymérase
QC :	Québec
R :	Coefficient de corrélation
R ² :	Coefficient de détermination
RAMQ :	Régie de l'assurance maladie du Québec
Réf. :	Référence
RSS :	Région socio-sanitaire
Ru :	Rural
SES :	Statut socio-économique
SK :	Saskatchewan
SNP :	(anglais) Single nucleotide polymorphism, polymorphisme d'un nucleotide simple
sp. :	Sous-espèce
TOL :	tolérance
U :	Urbain
VIF :	(anglais) Variance inflation factor, facteur d'inflation de la variance
VP+ :	Valeur prédictive positive
VP- :	Valeur prédictive négative
' :	Prime
% :	pourcentage
Σ :	Sommation
μ :	Moyenne

Remerciements

Cette étude est le fruit d'un travail d'équipe qui fut enrichissant tant au point de vue scientifique qu'au point de vue personnel. J'aimerais remercier ceux et celles qui ont donné de leur temps et qui ont partagé leur expertise afin que ce projet soit mené à bien. À nos collaborateurs de la Faculté de Médecine vétérinaire : Dr Michel Bigras-Poulin, épidémiologiste et directeur du Groupe de recherche en épidémiologie des zoonoses et santé publique (GREZOSP) et Dr Pascal Michel, épidémiologiste et coordinateur de la section Saint-Hyacinthe du Laboratoire de lutte contre les zoonoses d'origine alimentaire (LZA-SSH), merci pour leur support, leur expertise en épidémiologie ainsi que pour leurs suggestions et leurs idées; Mlle Laurie St-Onge, géographe au LZA-SSH, merci pour son souci du détail et pour sa précieuse aide dans la création de documents géographiques de qualité supérieure. À nos collaborateurs du Centre Universitaire de santé McGill (CUSM) : Dr Marcel A. Behr, professeur associé à l'université McGill et investigateur principal du laboratoire d'épidémiologie clinique, merci pour son support technique, son expertise en microbiologie et pour ses idées inspirantes; Dr Alain Bitton, gastroentérologue, merci pour sa participation dans la section de validation; Mlle Sarah Vahey et M. Abbas Kezouh, statisticiens à la Division d'épidémiologie clinique de l'Hôpital-Royal-Victoria, merci pour leurs conseils et leur aide sans borne dans la manipulation des données. Merci au Secteur Vigie et Protection de la Direction de santé publique de l'Agence de la santé et des services sociaux de Montréal pour la distribution des données d'incidence des MADO entériques. Du département de géographie du Ministère de la Santé et des Services Sociaux (MSSS), merci à Simon Mélançon pour son efficacité et pour les données populationnelles qu'il nous a fournies.

Un merci tout spécial à ma famille, Francine Bachand, Normand Lowe et Catherine Lowe, à Nicolai Østergaard et à mes amis, tout particulièrement Maninder Setia, Shamiza Hussein et Wendy Somerville, qui m'ont encouragée et supportée tout au long de ce projet. Finalement, je tiens sincèrement à remercier mon directeur de recherche, Dr Paul Brassard, d'avoir cru en mon potentiel dès le tout début de cette aventure. Au cours de ces deux années, Dr Brassard m'a fait découvrir l'épidémiologie, a su me guider et m'accompagner de manière assidue en faisant preuve d'une générosité sans borne, tant au niveau du temps que de sa personne.

Introduction

Les maladies inflammatoires des intestins (MII) regroupent la maladie de Crohn (MC) et la colite ulcéreuse (CU) (1). Jusqu'à ce jour, leur étiologie reste méconnue. Dans le cas de la MC, une cause multifactorielle demeure toutefois l'hypothèse la plus acceptée dans la littérature, c'est-à-dire qu'elle serait due à une combinaison de prédispositions génétiques, de facteurs démographiques et de facteurs environnementaux, incluant un agent pathogène (2). De plus, cette maladie affecte principalement les individus des pays développés et l'incidence rapportée sur plusieurs continents suggère une variation dans la répartition des cas (3).

Des données épidémiologiques ont récemment été obtenues à propos de la MC pour cinq provinces canadiennes, soit la Colombie-Britannique, l'Alberta, la Saskatchewan, le Manitoba et la Nouvelle-Écosse. L'incidence de la MC est donc estimée être de 13,4 nouveaux cas / 10^5 personnes / année au Canada (4). Au Québec, peu a été fait concernant l'épidémiologie de la MC. La prévalence des principales susceptibilités génétiques prédisposant à la MC chez des patients a toutefois été récemment étudiée (5).

Bien que l'étiologie de la MC soit indéterminée, des facteurs de risque prédisposant à cette maladie ont été identifiés et plusieurs études à ce sujet sont disponibles dans la littérature. Des variables individuelles, géographiques, démographiques et reliées à la condition de santé ont ainsi été associées à la MC dans certaines provinces du Canada (6;7). En effet, la MC se développerait surtout chez des individus dans la vingtaine et de façon générale, l'incidence est plus élevée chez les femmes. De plus, une association a été faite entre la MC et le lieu de résidence urbain, le tabagisme et la consommation de contraceptifs oraux chez les femmes. Il a aussi été observé que certaines populations dans le monde sont plus susceptibles au développement de cette maladie, telles les individus d'origine juive, et que d'autres le sont moins, comme les autochtones et les populations asiatiques. Ces tendances soutiennent le rôle d'une susceptibilité génétique dans le développement de la MC.

En effet, plusieurs mutations génétiques ont été associées à la MC, telles *IRGM*, *ATG16L1*, *NOD2/CARD15* et *IL23R*. Ces découvertes ont entraîné une meilleure connaissance des processus biologiques associés à ces gènes et prises ensembles, les évidences génétiques impliquent fortement des déficits dans la prise en charge de

bactéries intracellulaires dans la pathogenèse de la MC (8). Finalement, une association avec le taux d'incidence de certaines maladies entériques a été observée dans certaines études.

Plusieurs organismes pathogènes ont ainsi été associés à la MC, sans qu'aucune conclusion ne soit toutefois tirée. Cependant, il existe certaines évidences épidémiologiques qui suggèrent que *Mycobacterium avium* sp. *paratuberculosis* (MAP) serait le microorganisme le plus fortement associé à la MC (9). En effet, MAP est la cause d'une MII chez les animaux. Des symptômes histopathologiques et cliniques semblables ont été observés entre la maladie des animaux et celle des humains (10) (11). Un lien de causalité reste tout de même à être établi.

Étant donné que la MC n'est pas une maladie à déclaration obligatoire (MADO), l'étude de son incidence implique donc l'identification des nouveaux cas à partir de banques de données et de registres. Une définition opérationnelle de cas a été développée et validée au Manitoba. Elle permet donc d'identifier les cas de MC avec une grande sensibilité et spécificité à partir des banques de données administratives des provinces canadiennes. Cette définition se base sur le nombre de diagnostics de MC posés pour un même individu pendant une période de temps prédéfinie (12).

Lors d'une étude précédente, des cas de MC avaient été identifiés au Québec en se basant sur cette définition opérationnelle de cas qui avait alors été adaptée afin d'être appliquée à la banque de données de la Régie de l'assurance maladie du Québec (RAMQ) et au registre hospitalier Med Echo pour la période 1993-2002. L'aire de diffusion de cette étude était alors les 166 territoires de centre local de service communautaire (CLSC) et l'existence d'une hétérogénéité dans la répartition des cas avait alors été suggérée (13).

Dans cet ouvrage, nous décrivons l'étude de type écologique que nous avons menée afin de caractériser la MC au Québec. Un de nos principaux objectifs était de mieux comprendre la variation de l'incidence de la MC dans cette province. Aussi, le taux d'incidence moyen de la MC au Québec a été calculé pour la période 1995-2000 et une standardisation adéquate des taux a permis de comparer les résultats obtenus pour le Québec à ceux des autres provinces canadiennes. Par ailleurs, afin d'estimer la vraisemblance des cas identifiés à partir de la banque de données de la RAMQ et du

registre Med Echo, nous avons validé la définition opérationnelle de cas du Québec employée dans cette étude.

L'objectif général du présent ouvrage peuvent ainsi être résumés par: caractériser de manière épidémiologique la MC au Québec afin de mieux comprendre les caractéristiques individuelles, géographiques, démographiques et reliées à la condition de santé qui lui sont associées. De manière plus générale, nous croyons que l'identification de facteurs de risque associés à la MC pourra éventuellement servir à guider les intervenants de santé du Québec qui interagissent avec les individus qui en sont atteints.

Partie 1 : Recension de la littérature

1.1 Description de la maladie de Crohn

1.1.1 Signes cliniques et histopathologiques

La MC est une maladie inflammatoire des intestins, tout comme l'est la CU (1). Ces deux maladies se manifestent de diverses façons et à différents niveaux, ce qui permet de les diviser individuellement en sous-types associés aux différents phénotypes (14). La MC a été décrite pour la première fois en 1932 par trois médecins, soit Burril B. Crohn, Leon Ginzburg, and Gordon D. Oppenheimer qui décrivent leurs patients comme étant affectés par une entité pathologique et clinique qu'ils appelèrent « iléite régionale ». L'iléite régionale, une maladie unique et distincte de la tuberculose intestinale, fut par la suite rebaptisée « maladie de Crohn » en l'honneur du chercheur (15).

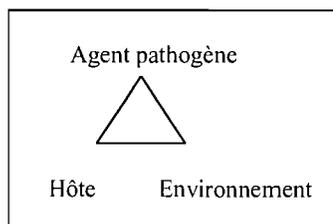
La MC peut être décrite comme un problème de santé complexe qui se manifeste de façon hétérogène (16). De manière générale, des événements pathologiques distincts sont observés dans les parties des intestins affectées, mais pas dans les segments sains, ni dans les intestins des individus en santé. Cette séquence d'événements pathologiques peut comprendre des dommages vasculaires, une artérite focale, une déposition de fibrine, une occlusion artérielle (principalement au niveau de la *muscularis propria*) et un infarctus des tissus peut se produire, c'est-à-dire une nécrose des tissus due à un apport insuffisant en sang oxygéné, ou une néovascularisation (17). Par ailleurs, trois variables de la classification internationale des maladies de Vienne ont été choisies afin d'établir le pronostic de la MC: l'âge au diagnostic, le comportement de la maladie et sa localisation (18).

Généralement, la maladie se développe chez les adultes et est le plus souvent diagnostiquée entre 20 et 40 ans (16). Des études populationnelles ont démontré que l'âge au diagnostic est en réalité distribué de façon bimodale, avec un premier pic d'incidence dans la vingtaine et un second pic, plus faible, dans la cinquantaine (19). Malgré le fait que cette maladie inflammatoire puisse affecter toutes les parties du tractus gastro-intestinal, de la bouche à l'anus, elle affecte plus fréquemment l'iléus terminal et le colon. Une particularité de la MC est l'intermittence de tissus en santé et de segments d'intestins en inflammation (16).

1.1.2 Modèle épidémiologique de causalité

Bien que l'étiologie de la MC reste jusqu'à ce jour méconnue, le modèle de la « triade épidémiologique » a été suggéré pour décrire cette maladie chronique. En effet, ce modèle implique une interaction entre les caractéristiques intrinsèques de l'hôte, un agent pathogène et l'environnement (voir figure 1, p. 5) (2).

Figure 1 : Modèle de causalité de la triade épidémiologique



De façon générale, il est accepté que le développement de la MC soit dû à la combinaison d'une susceptibilité génétique et de divers facteurs environnementaux incluant les conditions démographiques, géographiques, socio-économiques et sanitaires. Des antigènes d'agents microbiens pourraient aussi être impliqués (20). Lors d'un débat dirigé présenté lors du congrès du *European Crohn's and Colitis Organisation (ECCO) 2007*, des arguments en faveur de l'importance des facteurs environnementaux dans le développement de la MC ont été soulevés. Il a de plus été suggéré que l'effet de certains d'entre eux pourrait se traduire dans la saisonnalité des cas, la variation géographique de l'incidence (variations nord-sud/ouest-est et variabilité locale) et dans l'agrégation des cas par mois de naissance. Par ailleurs, le statut socio-économique, la présence d'appendicectomie, la possibilité que des bactéries pathogènes de l'environnement puisse interagir avec l'hôte, le rôle des bactéries commensales de l'intestin et la prise de contraceptifs oraux chez les femmes seraient d'autres facteurs potentiellement associés à la MC (21).

La prévalence, la modulation et l'expression phénotypique seraient donc possiblement dépendantes d'un ou plusieurs facteurs environnementaux et une composante génétique serait aussi impliquée. Ce phénomène complexe peut être aisément observé lors de l'étude des différences géographiques, des changements démographiques d'un bassin d'individus et de la migration des populations (22). À ce sujet, il a été postulé que la population actuelle de l'Inde serait sélectionnée de manière évolutive pour une constitution allélique optimale lui permettant de combattre des infections de tous sexes.

Ainsi, dans leur environnement d'origine, les Indiens s'exposent à divers agents infectieux au cours de leur vie, ce qui favorise le bon fonctionnement de leur système immunitaire. Cependant, lorsque ces mêmes individus se retrouvent dans les conditions plus hygiéniques de pays occidentaux tels le Royaume-Uni ou le Canada suite à l'immigration, leur système immunitaire se dérègle (23). Cela suggère donc que dépendamment du bagage génétique qu'un individu possède, le fait de changer d'environnement pourrait déclencher le développement de la MC.

1.2 Épidémiologie de la maladie de Crohn

1.2.1 Incidence et prévalence

i) Épidémiologie dans le monde

La comparaison des taux d'incidence des MII entre différentes études engendre quelques difficultés. En effet, les taux de détection et les critères diagnostiques peuvent différer entre les sites de recherche, tout comme l'accès aux procédures diagnostiques (ex. endoscopie) peut varier selon le temps et les centres (24). Malgré ces obstacles, il est généralement accepté que le taux d'incidence de la MC est plus élevé dans les pays développés que dans les pays en voie de développement (25). En effet, il a augmenté rapidement entre les années 1960 et 1970 en Amérique du Nord et en Europe du Nord et depuis, il continuerait à augmenter à un rythme plus lent.

Par ailleurs, la prévalence rapportée de la MC en Europe démontre une grande variabilité des taux, de 10 à 150 cas / 10^5 personnes / année (24) et les taux d'incidence de la MC y varient de 0,7 à 9,8 cas / 10^5 personnes / année. De plus, la population européenne de 2002 ayant été estimée à environ 580 millions d'individus, entre 23 000 et 41 000 nouveaux cas de MC seraient diagnostiqués chaque année en Europe (16).

Il a été observé que l'on retrouve plus de cas de MC en Europe et en Amérique du Nord que sur les autres continents (voir tableau I, p. 7). Cependant, l'incidence des MII a augmenté à certains endroits du monde où elle avait toujours été faible, comme au Japon, en Corée du sud, à Singapour, au nord de l'Inde et en Amérique latine (3). En Asie, l'incidence de la MC varierait de 0,5 à 1,0 cas / 10^5 personnes / année. Toutefois, ces résultats sont à prendre avec un grain de sel puisqu'il n'existe qu'un faible nombre

d'études valides faites en Asie et mis à part au Japon, il n'existe pas de registre des MII dans les pays asiatiques (26).

Par ailleurs, le premier cas de MC en Côte-d'Ivoire a été identifié en 2004 (3).

Tableau I : Incidence et prévalence de la maladie de Crohn dans le monde¹

AUTEUR (S) (RÉFÉRENCE)	PAYS	PÉRIODE	TAUX D'INCIDENCE PAR 100 000	PRÉVALENCE PAR 100 000
Bjornsson et al.	Islande	1990-1994	5,5	-
Moum et al.	Norvège	1990-1993	5,8	-
Langholz et al. Munkholm et a.	Danemark - Copenhague	1980-1987	4,1	54
Kyle	Écosse (nord-est)	1985-1987	9,8	147
Russel et al.	Pays-Bas - Limburg, sud	1991-1994	6,9	
Trallori et al.	Italie - Florence	1990-1992	3,4	40
Odes et al.	Israël (sud)	1987-1992	4,2	50,6
Morita et al.	Japon	1991	0,5	5,8
Wright et al.	Afrique du Sud - Cape Town (blancs)	1980-1984	2,6	-
	Afrique du Sud - Cape Town (couleur)	1980-1985	1,8	-
	Afrique du Sud - Cape Town (noirs)	1980-1986	0,3	-
Bernstein et al.	Canada (CB;AB;SK;MA;N-É)	1998-2000	13,4*	233,7

*Ajusté directement pour la population totale des cinq provinces

ii) Épidémiologie en Amérique du Nord

En Amérique du Nord, les taux d'incidence de la MC varient entre 3,6 et 20,2 cas / 10⁵ personnes / année. La prévalence varie entre 26 et 318,5 cas / 10⁵ personnes / année (voir tableau II, p. 9). Lorsque ces taux sont extrapolés à l'estimé des populations du Canada et des États-Unis de 2003, cela implique qu'environ 10 000 à 47 000 résidents de ces deux pays recevraient un diagnostic de la MC annuellement et que

¹ Adapté de : Loftus, EV. *Clinical epidemiology of Inflammatory bowel disease: incidence, prevalence and environmental influences*. Gastroenterology, 2004(126)1504-1517 (16)

630 000 personnes souffriraient de la maladie dans ces deux pays combinés (16). Par ailleurs, il a été estimé qu'environ 170 000 Canadiens souffriraient présentement de MII (4).

Certaines études se sont penchées sur l'épidémiologie de la MC à un niveau plus local. Par exemple, dans le comté d'Olmsted de l'état américain du Minnesota (MN), le taux d'incidence de la MC a augmenté entre 1940 et 1970 et est resté stable entre 1970 et 2000 (27). Par ailleurs, en 1999, une étude épidémiologique de la MC a été effectuée dans la province canadienne du Manitoba et l'incidence rapportée pour la période 1989-1994 était de 14,6 nouveaux cas / 10^5 personnes / année. La prévalence en 1994 était estimée à 198,5 cas / 10^5 personnes / année (12).

L'incidence et la prévalence de la MC a par la suite été estimée dans quatre autres provinces canadiennes pour la période 1998-2000 (4). Au Canada, la province de la Nouvelle-Écosse aurait le taux d'incidence le plus élevé, c'est-à-dire 20,2 cas / 10^5 personnes / année. À l'opposé, c'est dans la province de la Colombie-Britannique que l'on retrouverait le taux d'incidence le plus faible du pays : 8,8 nouveaux cas / 10^5 personnes / année (voir tableau II, p. 9) (4). Toutefois, des résultats ont révélé que dans cette province de l'ouest, le nombre de cas de MC a augmenté de six fois entre 1995 et 2005 chez les enfants de moins de 16 ans dont les parents étaient originaires d'Asie du Sud (22).

Tableau II : Incidence et prévalence de la maladie de Crohn en Amérique du Nord²

AUTEUR (S) (RÉFÉRENCE)	PAYS	PÉRIODE	TAUX D'INCIDENCE PAR 100 000	PRÉVALENCE PAR 100 000
Hiatt et al.	É.U. - nord de Californie	1980-1981	7,0	-
Kurata et al.	É.U. - sud de Californie	1987-1988	3,6	26,0
Stowe et al.	É.U. - Monroe County, NY	1980-1989	3,9	-
Loftus et al.	É.-U. - Olmsted County, MN	1984-1993	6,9	162,0
Loftus et al.	É.-U. - Olmsted County, MN	1990-2000	7,9	174
Bernstein et al.	Canada (AB;SK;MA;N-É)	1998-2000	13,4 *	233,7
Bernstein et al.	Colombie-Britannique	1998-2000	8,8	160,7
Bernstein et al.	Alberta	1998-2000	16,5	283,0
Bernstein et al.	Saskatchewan	1998-2000	13,5	263,8
Bernstein et al.	Manitoba	1998-2000	15,4	271,4
Bernstein et al.	Nouvelle-Écosse	1998-2000	20,2	318,5

*Ajusté directement pour la population totale des cinq provinces

1.2.2 Variations géographiques de l'incidence de la maladie de Crohn

Une variation nord-sud dans l'incidence des MII a été observée en Europe et aux États-Unis (28). Les plus hauts taux d'incidence et de prévalence de la MC ont par ailleurs été rapportés dans le nord de l'Europe, au Royaume-Uni et en Amérique du Nord (voir tableau I, p. 7). Dans certains pays d'Asie et du sud de l'Europe, les taux d'incidence et de prévalence sont en croissance. D'un autre côté, les taux d'incidence les plus bas rapportés dans la littérature sont en Amérique du Sud, dans le sud-est de l'Asie, en Afrique (excluant l'Afrique du Sud) et en Australie (16). On observe ainsi une variation croissante de l'incidence de la MC du sud vers le nord. Plusieurs hypothèses tentent d'expliquer cette variation et l'une d'entre elles se base sur le fait qu'elle pourrait n'être que le reflet d'une inégalité dans l'accès et la qualité des soins de santé, ainsi que de certaines différences dans les niveaux d'industrialisation, d'hygiène et d'assainissement des différentes régions du monde (29).

² Adapté de : Loftus, EV. *Clinical epidemiology of Inflammatory bowel disease: incidence, prevalence and environmental influences*. *Gastroenterology*, 2004(126)1504-1517 (16)

En France, un gradient nord-sud de l'incidence de la MC a clairement été mis en évidence entre les différents départements de santé du pays. Les endroits ayant les risques relatifs atténués les plus élevés de la MC sont localisés dans le tiers nord de la France (28). Il est à noter que ces différences dans la présentation de la maladie pourraient aussi refléter les différents patrons génétiques propres aux populations de ces pays. À ce sujet, il a été démontré qu'il y a une hétérogénéité génétique chez les individus d'Europe où l'on observe une répartition inégale des variations génétiques prédisposant à la MC entre les peuplades du nord par rapport à celles du sud (30).

La possibilité qu'un « gradient » nord-sud de l'incidence de la MC soit réel a été soulevée lors du congrès de l'ECCO 2007 et une hypothèse a été suggérée pour expliquer ce phénomène. En effet, cette variation pourrait être directement reliée au taux de gras des individus qui varierait dépendamment d'où les populations se trouvent. Cette hypothèse se base partiellement sur le fait qu'en 1932, Dr Burrill B. Crohn avait relevé des anomalies dans les niveaux de gras des tissus mésentériques des patients atteints d'« iléite régionale » (voir section 1.1.1). Ainsi, il a été suggéré que les modes de vie des individus originaires de régions plus au nord entraîneraient une plus grande adiposité chez ces personnes comparativement aux individus originaires du sud. Cette hypothèse va d'ailleurs dans le même sens que les données d'incidence et de prévalence du nord par rapport au sud (31).

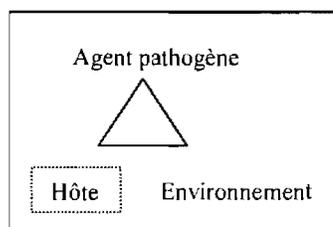
À ce sujet, il est intéressant de noter que plusieurs études ont tenté d'éclaircir le lien possible entre le taux de gras d'un individu et le développement de la MC. Par exemple, il a été postulé que les adipocytes sont des producteurs d'hormones, de facteurs de croissance et de protéines impliquées dans la reconnaissance bactérienne et que 95% des bactéries viables retrouvées dans les tissus mésentériques sont localisées dans les adipocytes qui agissent tels des macrophages (31). Ainsi, le gras pourrait favoriser le développement de certains facteurs de risque associés à la MC, comme la survie bactérienne chez un hôte.

1.3 Prédilection de l'hôte à la maladie de Crohn

1.3.1 Gènes associés à la maladie de Crohn

La cartographie du génome humain qui s'est terminée en 2003 a permis d'identifier des *loci* de susceptibilité aux MII tels *IBD1* sur le chromosome 16 (32), *IBD3* sur le chromosome 6 et *IBD5* sur le chromosome 5 (33). Depuis, plusieurs prédispositions génétiques à la MC ont été identifiées, introduisant ainsi l'apport de la constitution allélique des gènes dans le développement de la MC (voir figure 2, p. 11).

Figure 2 : Triade épidémiologique de causalité - l'hôte



Récemment, trois gènes de susceptibilité ont été identifiés par un consortium de chercheurs canadiens et américains : les gènes *PHOX2B*, *NCF4* et *ATG16L1* (34). Il est intéressant de constater que deux de ces gènes jouent un rôle dans la prise en charge bactérienne. En effet, *ATG16L1* fait partie d'une famille de gènes impliqués dans le processus d'autophagie, un mécanisme biologique impliqué dans la dégradation des protéines, la prise en charge des antigènes, la régulation du signal cellulaire ainsi que dans plusieurs autres voies essentielles à l'initiation et à la régulation de la réponse inflammatoire. De plus, le gène *NCF4* code pour une protéine qui joue un rôle important dans l'activité de la NADPH-oxydase ainsi que dans la génération d'espèces oxygénées activées, essentielles au processus de phagocytose. Ces nouveaux gènes associés à la MC s'ajoutent aux variantes des gènes *IL23R* et *NOD2/CARD15* aussi associées au développement de la MC.

1.3.2 Le gène *NOD2/CARD15* et son rôle dans la reconnaissance bactérienne

Une région de susceptibilité à la MC a été identifiée sur le chromosome 16 par le Dr Jean-Pierre Hugot en 1996, soit *IBD1* (1). Par la suite, des polymorphismes d'un nucléotide simple (SNP) ont été découverts dans le gène *NOD2/CARD15* qui est situé dans ce locus de susceptibilité. Il a par la suite été démontré que la protéine *nod2* codée

par ce gène est impliquée dans la reconnaissance bactérienne et dans l'activation d'une voie de défense immunologique. En effet, la protéine nod2 possède une région riche en leucines répétées qui possède une activité de liaison aux lipopolysaccharides imbriqués dans le peptidoglycan de la membrane plasmique des bactéries. Cette liaison permet ainsi à la protéine nod2 de reconnaître un envahisseur bactérien potentiel et d'entraîner par la suite la stimulation de la voie immunologique NF-kB (35).

Trois mutations qui confèrent une susceptibilité accrue à la MC ont ainsi été recensées à l'intérieur du gène *NOD2/CARD15* (36) (35). Des études sur la transmission de ces mutations génétiques ont démontré qu'une transmission préférentielle de la mutation 3020insC (ou SNP13 ou Leu1007insC) se fait à partir de parents hétérozygotes à des enfants chez qui la MC se développe à un moment de leur vie (36). De plus, la fréquence allélique de 3020insC dans les groupes de patients affectés par la MC équivaut à 8,2%. Par ailleurs, il a été observé que cette mutation est retrouvée chez 8,4% des patients juifs, chez 8,1% des patients caucasiens non-juifs, chez 4% des témoins sains et chez 3% des patients ayant une colite ulcéreuse (35). Deux autres mutations du gène *NOD2/CARD15*, SNP8 (ou Arg702Trp) et SNP12 (ou Gly908Arg), ont aussi été recensées plus souvent chez les patients MC par rapport aux patients ayant une colite ulcéreuse et aux témoins sains.

Au Québec, il a été démontré que la prévalence des trois mutations de *NOD2/CARD15* chez les patients atteints de MC (45%) est significativement plus élevée que dans un groupe contrôle (9%). De plus, une relation entre le génotype et le phénotype a été observée, c'est-à-dire qu'il y aurait une association significative entre la présence de variations dans le gène *NOD2/CARD15* et la maladie présente au niveau iléal du tube digestif opposé à un phénotype de la maladie affectant strictement le colon (5). Le génotypage de 45 nouveaux SNPs a de plus été accompli dans la région du locus comprenant *NOD2/CARD15*.

Par ailleurs, une question d'intérêt a été soulevée par Vermeire et ses collaborateurs concernant le rôle du gène *NOD2/CARD15* dans la susceptibilité à la MC :

« *Quelle est la contribution des variantes de NOD2/CARD15 (Arg702Trp, Gly908Arg, Leu1007insC) dans le développement de la MC chez des*

*populations de patients divergentes géographiquement?*³ » (en référence aux populations européenne, québécoise et nord-américaine) (5).

1.3.2 Le sexe

Le sexe est un aspect intrinsèque de l'individu qui est associé à la MC. En effet, la MC n'est pas distribuée également entre les sexes féminin et masculin. Tout d'abord, il a été observé que l'incidence de la MC pédiatrique est plus élevée chez les garçons que chez les filles. Par exemple, à Stockholm, les garçons âgés de moins de 16 ans ont une incidence de la MC significativement plus élevée que celle des filles du même âge; pour la période 1990-2001, l'incidence de la MC chez les garçons atteignait presque le double de celle retrouvée chez les filles ($p = 0,0144$), c'est-à-dire $6,2 \text{ cas} / 10^5 \text{ personnes} / \text{année}$ (IC 95% : 5,0-8,2) par rapport à $3,6 \text{ cas} / 10^5 \text{ personnes} / \text{année}$ (IC 95% : 2,5-4,8) (37). Un phénomène semblable est rapporté en Colombie-Britannique où une prédominance masculine a été observée parmi les patients atteints de MC pédiatrique chez la population originaire d'Asie du Sud. En effet, le ratio masculin/féminin chez les individus âgés de moins de 16 ans atteints de MC pour la période 1996-2005 était de 3,5 (22)

Par ailleurs, le ratio des sexes s'inverse lorsqu'on observe la MC chez l'adulte. En effet, l'incidence de la MC chez les adultes est plus élevée chez les femmes que chez les hommes. Ce phénomène est rapporté par plusieurs études épidémiologiques, dont une menée à Copenhague où il a été démontré que chez les individus âgés entre 15 à 24 ans, l'incidence moyenne retrouvée chez les femmes dépassait près de deux fois celle des hommes pour la période 1979-1987. En effet, une incidence moyenne de $12,8 \text{ cas} / 10^5 \text{ personnes} / \text{année}$ a été observée chez les femmes de 15 à 24 ans, tandis que l'incidence chez les hommes était de $6,0 \text{ cas} / 10^5 \text{ personnes} / \text{année}$ (38). À plus grande échelle, le ratio femmes/hommes de l'incidence de la MC au Danemark a été estimé à 1,34 (24). Par ailleurs, au Manitoba pendant la période 1989-1994, le ratio des taux d'incidence (IRR) de la MC chez les femmes par rapport aux hommes était équivalent à 1,38 (IC 95% : 1,22 – 1,57), les femmes ayant un taux d'incidence supérieur de 38% par rapport aux hommes (12) tandis qu'en Nouvelle-Zélande, il a été établi que 61,4% des patients atteints de MC sont du sexe féminin (39). Toutefois, certaines études ont démontré qu'au Japon, ce sont les hommes qui seraient le plus à risque de développer la MC (26)

³ Traduit de l'anglais : "*What is the contribution of these variants in geographically diverse patient populations?*" (5)

où cette maladie serait de 2 à 2,5 fois plus commune chez les hommes que chez les femmes (40;41).

1.3.2 L'âge au moment du diagnostic

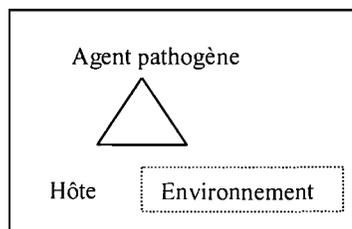
La MC est une maladie qui se développe généralement chez les individus jeunes. En effet, il a été démontré qu'au Manitoba, le taux d'incidence le plus élevé pour la période 1989-1994 est celui du groupe d'âge des 20-29 ans avec un taux de 28,4 cas / 10^5 personnes / année, par rapport à un taux moyen pour tous les groupes d'âge de 14,6 cas / 10^5 personnes / année (12). De tels résultats ont d'ailleurs été corroborés dans une cohorte italienne où le plus haut taux d'incidence spécifique à l'âge était aussi retrouvé entre 20 et 29 ans (42). Par ailleurs, l'incidence de la MC a augmenté de six fois entre 1962 et 1987 dans le comté de Copenhague et l'incidence maximale durant cette période est celle du groupe d'âge de 15-24 ans (38). Une autre étude danoise a démontré des résultats semblables pour le comté de Jutland nord où les taux d'incidence les plus élevés sont retrouvés dans le groupe d'âge des 15-29 ans. Un taux d'incidence moyen de 15,4 cas / 10^5 personnes / année (IC 95% : 13,3-17,6) a été observé chez les femmes de cette tranche d'âge par rapport à 8,8 cas / 10^5 personnes / année (IC 95% : 7,2-10,4) chez les hommes du même âge (24). Par ailleurs, il a été démontré que jusqu'à 25% des patients affectés par une MII reçoivent leur diagnostic pendant l'enfance ou l'adolescence (1).

L'association de variations de certains gènes avec la MC a aussi permis de démontrer qu'il existe des relations génotype-phénotype (30) et en ce qui a trait à l'âge au diagnostic de la MC, il a été suggéré que les patients qui possèdent deux allèles mutés du gène *NOD2/CARD15* auraient un plus grand risque de développer la maladie à un jeune âge (30;43). Par ailleurs, il a été observé dans une population d'Europe du Nord, c'est-à-dire un regroupement d'individus d'origine écossaise ou irlandaise, qu'une mutation dans le gène *CD14* serait aussi associée au diagnostic à un jeune âge (30).

1.4 Facteurs potentiellement associés à la maladie de Crohn : variables individuelles, géographiques, socioéconomiques et reliées à la condition de santé

Une des théories proposées afin d'expliquer l'étiologie de la MC implique une réaction à la diète ou à d'autres facteurs environnementaux chez un individu prédisposé génétiquement (44) (voir figure 3, p. 15).

Figure 3 : Triade épidémiologique de causalité - les facteurs environnementaux



1.4.1 Lieu de résidence – urbanité, ruralité

Il a été observé qu'il y aurait plus de cas de MC dans les régions urbaines que dans les régions rurales. Cette observation n'implique cependant pas automatiquement la géographie du territoire comme variable associée à la MC, mais peut en fait refléter la présence de facteurs de risque environnementaux dans certaines régions par rapport à d'autres. De plus, l'industrialisation d'un territoire, un accès inégal à des soins de santé spécialisés ou encore l'hygiène et l'assainissement des eaux sont des facteurs additionnels à considérer (29). En effet, il a été observé en Inde que la disparition des helminthes (vers parasites de l'intestin) coïncide dans le temps avec l'émergence des MII. Ainsi, l'amélioration de l'hygiène et des moyens sanitaires aurait potentiellement un rôle à jouer dans l'apparition précoce de la MC dans les population urbaines par rapport aux rurales (45). Ces observations vont dans le sens de l'hypothèse hygiénique qui prédit que les individus ayant été moins en contact avec des organismes pathogènes durant l'enfance seraient plus à risque de développer des MII (voir section 1.4.5).

1.4.2 Tabagisme

Le tabagisme est l'un des facteurs de risque le plus fortement associé à la MC dans la littérature. En effet, fumer augmenterait le risque de développer la MC (46). Une méta-analyse à ce sujet par Calkins a permis d'établir que le risque de développer la MC est deux fois plus élevé chez les fumeurs comparé aux individus n'ayant jamais fumé de leur vie (OR combiné = 2,0, IC 95% :1,65-2,47) (46). De plus, fumer aggrave le cours de la MC chez ceux qui en sont atteints, pouvant entraîner une utilisation plus fréquente de stéroïdes et d'immunosuppresseurs. Des chirurgies à répétition peuvent aussi être nécessaires chez les cas plus graves (46). Par ailleurs, il a été observé que cesser de fumer améliore le cours de la MC et diminue le risque des individus sains de la développer (47). Il a été démontré qu'il existe une relation dose-effet entre le niveau d'exposition à la fumée de cigarette et le risque de MC, plus particulièrement chez les fumeurs actifs. En effet, lorsqu'un individu cesse de fumer, son risque de développer les signes cliniques reliés à la MC diminue avec les années passées sans fumer. Cela suggère que les dommages intestinaux causés par la fumée ainsi que l'expression clinique de la maladie pourraient avoir un caractère réversible (48).

Une exposition active au tabac est associée avec un diagnostic retardé de la MC. En effet, une étude récente a révélé que parmi un groupe de patients ayant eu un diagnostic tardif de la maladie (plus de 40 ans), environ la moitié d'entre eux étaient des fumeurs, par rapport à seulement 23% des patients ayant reçu un diagnostic précoce (moins de 40 ans) (49). Cela supporterait par ailleurs la théorie que l'influence génétique est plus déterminante lors du développement de la maladie à un plus jeune âge, tandis que les facteurs environnementaux seraient reliés de façon plus importante à un diagnostic tardif (49).

L'exposition à la fumée secondaire aurait aussi un rôle à jouer dans le développement de la MC puisque les patients atteints de la MC ont 72% plus de chance d'avoir été soumis à une exposition prénatale à la fumée de cigarette que des individus en santé. De plus, deux fois plus d'individus atteints de MC que de témoins sains ont subi une exposition passive à la fumée de cigarette pendant l'enfance, c'est-à-dire un ou deux parents fumeurs (50).

D'un autre côté, fumer la cigarette n'est pas considéré comme un facteur de risque pour toutes les MII; fumer aurait un effet bénéfique chez les patients atteints de CU (51). En effet, l'incidence de la CU est plus faible chez les fumeurs. De plus, les individus atteints de CU qui fument ont un meilleur cheminement clinique ainsi qu'un plus faible nombre de chirurgies encourues que les individus atteints non-fumeurs (52). Par ailleurs, une étude d'association rétrospective a été conduite chez des gens issus de même famille. Ce modèle de recherche permet ainsi de générer des observations à partir d'individus qui partagent un bagage génétique similaire en ce qui a trait à leurs prédispositions aux maladies inflammatoires de l'intestin. Dans le cas où dans une même famille, un individu développe la MC et un autre, la CU, il a été démontré que le fait de fumer au moment du diagnostic est associé de façon significative avec le développement de la MC et non pas avec celui de la CU. Ainsi, la MC a 10,5 fois plus de chances de se développer chez l'individu fumeur par rapport au développement de la CU chez sa fratrie (53).

Les mécanismes sous-jacents à l'action de la cigarette sur le développement de l'inflammation ne sont pas complètement définis. Ce phénomène est alimenté par l'inexistence d'un modèle animal adéquat et par la complexité de la fumée de tabac qui contient des centaines de substances différentes, incluant la nicotine, des radicaux libres et le monoxyde de carbone (46). Parmi plusieurs études impliquant des modèles animaux préliminaires, l'une d'entre elles a permis de démontrer que chez le rat, une pré-exposition à la fumée de cigarette aggrave une condition de colite causée de manière expérimentale (54). La fumée de cigarette engendrerait en effet un stress oxydatif considérable à l'organisme qui affecterait les muqueuses intestinales (55). Il a été suggéré que l'effet néfaste causé par la fumée de cigarette dans le cas d'une colite pourrait être dû à la combinaison d'effets métaboliques divers dont une diminution des niveaux d'antioxydant tel le glutathion, une surproduction de l'agent pro-inflammatoire leukotriène-B4 et une accumulation de neutrophiles dans les tissus du colon. Ces derniers augmentent la concentration des métabolites réactifs de l'oxygène à cet endroit du système digestif (54).

1.4.3 Utilisation de contraceptifs oraux

Une méta-analyse menée en 1995 portant sur le rôle des contraceptifs oraux dans le développement des MII a conclu qu'il existe une association modeste entre

l'utilisation des contraceptifs oraux et le développement de la MC (56). C'est en combinant les résultats de neuf études que les auteurs sont parvenus à augmenter le pouvoir statistique de leurs analyses et ainsi, questionner l'association possible entre l'utilisation des contraceptifs oraux et le développement de la MC et de la CU. Un risque relatif de 1,44 (IC 95% : 1,12-1,86) pour la MC a été obtenu en association avec l'utilisation des contraceptifs oraux (valeur ajustée pour le tabagisme). Les auteurs de cette méta-analyse ont inclus les résultats d'études publiées entre 1975 et 1993. En 1994, Boyko et ses collaborateurs ont observé lors d'une étude cas-témoins que les femmes ayant fait usage de contraceptifs oraux pendant plus de six ans étaient 5 fois plus à risque de développer la MC que les témoins, et ce, même après ajustement pour la race, le tabagisme, les revenus et les histoires de grossesses. Des résultats différents, mais allant dans le même sens que les précédents ont été obtenus en 2001 en Espagne, où il a été démontré que l'usage de contraceptifs oraux est un facteur de risque à la MC qui triplerait presque les chances de développer la MC (OR = 2,8, I.C. 95% : 1,009-7,774) (57).

Par ailleurs, Bernstein et ses collaborateurs ont démontré que dans la province canadienne du Manitoba, les individus atteints de MC initient la prise de contraceptifs oraux à un âge significativement plus bas par rapport aux individus sains. En effet, il a été observé que les sujets dont le diagnostic de MC avait été fait avant l'âge de 24 ans avaient débuté la prise de contraceptifs oraux plus jeunes ($17,7 \pm 2,6$ ans ($p = 0,023$)) que les témoins (6).

Des études histopathologiques ont proposé un mécanisme pathogénique selon lequel un infarctus gastro-intestinal multifocal entraînerait un développement de la MC, au cours duquel le tabagisme et les contraceptifs oraux agiraient comme de potentiels agents thrombogéniques (17;58). En effet, la pilule contraceptive aurait des effets modulateurs sur certains facteurs de coagulation tels le fibrinogène et la prothrombine qui pourraient augmenter la tendance à former des thromboses (59). Des observations ont été faites à propos de la concentration d'hormones contenue dans les pilules contraceptives en lien avec certains facteurs impliqués dans la coagulation. En effet, les pilules qui contiennent 50 mg et plus d'éthinylestradiol augmenteraient les niveaux de substrats pro-coagulants et diminueraient les niveaux d'anticoagulants ainsi que l'activité fibrinolytique (58). Les pilules contraceptives seraient ainsi la cause d'un stimulus prothrombique pouvant entraîner des changements microvasculaires. Chez des

individus atteints de la MC, cela résulterait en une aggravation de l'activité de la maladie (58).

1.4.4 Statut socio-économique

Le statut socioéconomique est généralement un prédicteur important de santé, c'est pourquoi les études qui impliquent une exposition à des facteurs environnementaux devraient en tenir compte. Afin de contrôler de façon adéquate les facteurs de confusion lors de l'élaboration de modèles épidémiologiques, il est ainsi essentiel d'inclure ce phénomène aux multiples facettes dans toutes les étapes de la recherche (60)

À ce sujet, la littérature suggère que les individus atteints de MII poursuivent une éducation scolaire à un niveau plus haut que la population générale (61) (62). De plus, ils proviendraient de secteurs socioéconomiques plus élevés que la population générale (63-65). Par ailleurs, l'apparition précoce de la MC dans les groupes ayant un statut socioéconomique élevé soulève l'importance du rôle joué par l'amélioration de l'hygiène et des moyens sanitaires dans l'étiologie des MII (45). Il a été démontré dans la province canadienne du Manitoba que le taux d'emploi des patients atteints d'MII au moment de recevoir leur diagnostic n'est pas différent de celui attendu dans la population générale.

1.4.5 Implication des maladies entériques dans le développement de la MC et hypothèse hygiénique

Jusqu'à présent, deux hypothèses principales soutiennent l'association de facteurs infectieux à la pathophysiologie de la MC, c'est-à-dire une perte de tolérance du système immunitaire intestinal envers les bactéries environnementales ou envers les bactéries commensales (66). En ce qui a trait au développement du système immunitaire intestinal, la littérature regorge de conclusions paradoxales à ce sujet. Par exemple, des études épidémiologiques ont suggéré que l'« hypothèse hygiénique » s'appliquerait à la MC, c'est-à-dire que les infections contractées pendant l'enfance engendrent le développement d'une protection contre les maladies ultérieures. D'un autre côté, d'autres études ont suggéré que les individus ayant eu des infections à un bas âge auraient un risque plus élevé de développer la MC (67).

i) L'hypothèse hygiénique

Tout d'abord, il a été suggéré que l'absence d'infections pendant l'enfance pourrait être un facteur de risque au développement des MII (68). Cette association soutient l'hypothèse hygiénique qui fut initialement suggérée par l'épidémiologiste britannique David Strachan en 1989. Ce chercheur avait alors noté que les rhino-conjonctivites allergiques saisonnières étaient moins communes chez les plus jeunes enfants de larges familles (69). Cette hypothèse fut reprise concernant l'étiologie de la MC tel qu'exposé par Gent et ses collaborateurs :

« Une explication possible serait celle où la MC est initiée par une infection entérique et l'amélioration des conditions hygiéniques des habitations limite l'exposition des enfants aux organismes entériques qui programment le système immunitaire digestif, rendant ainsi le système digestif plus susceptible aux attaques infectieuses plus tard dans la vie.⁴ »
(70)

ii) Flore intestinale et développement du système immunitaire

En se basant sur l'hypothèse que la MC se développe lorsqu'un agent pathogène spécifique infecte un hôte chez qui le système de défense intestinal est déficient, il devient intéressant d'évaluer le rôle joué par la composition de la flore intestinale dans le développement des barrières immunitaires de l'être humain. En effet, les bactéries du système digestif jouent un rôle essentiel dans la maturation du système immunitaire. On retrouverait un nombre plus grand de bactéries dans les intestins que de cellules dans le corps humain (71) c'est-à-dire environ 1 à 2 kg de bactéries dans le système digestif d'un humain (72). Par ailleurs, il a été démontré chez les animaux que la présence de bactéries dans l'environnement est nécessaire au bon développement du système immunitaire. En effet, des animaux de laboratoire ont été maintenus dans un environnement exempt de contaminants et on a observé une déficience au niveau des composantes immunologiques des tissus lymphoïdes associés à leur tube digestif (GALT) : faible nombre de cellules lymphocytaires intestinales de types T et B, petits follicules lymphoïdes sans centre germinal, et finalement, faible niveau d'immunoglobulines sécrétées (73). De plus, il a été démontré que les organismes de

⁴ Traduit de l'anglais: "One possible explanation is that Crohn's disease is initiated by an enteric infection and improved hygiene from provision of better household amenities has limited the exposure of infants to enteric organisms that programme the immune system of the gut, thereby rendering the bowel more susceptible to triggering infection later in life." (70)

la famille des helminthes protégeraient des désordres immunologiques comme l'inflammation intestinale (45). Ainsi, la composition de la flore bactérienne intestinale aurait la capacité d'augmenter ou de diminuer la perméabilité du système digestif à d'autres types de bactéries potentiellement pathogènes (72). Il existe aussi une hypothèse concernant le rôle de la vaccination dans la maturation normale du système immunitaire et dans le développement et le maintien de l'équilibre des mécanismes de régulation. En effet, il est suggéré que des irrégularités dans l'évolution du système immunitaire à l'enfance pourraient avoir un impact sur la santé plus tard dans la vie (74). Des résultats controversés ont été obtenus lors de l'étude de l'impact de la vaccination en bas âge avec le vaccin combiné contre la rougeole, les oreillons et la rubéole (vaccin MMR). En effet, une association entre cette vaccination et un syndrome gastro-intestinal a été démontrée (75), mais une étude écologique menée en Angleterre n'a pas révélé d'association entre l'introduction du vaccin MMR et l'augmentation de la MC entre les années 1991-2002 (76). Il n'existe ainsi aucun lien causal jusqu'à ce jour.

iii) Infections gastro-intestinales à l'enfance

D'un autre côté, il a aussi été proposé qu'une fréquence élevée de gastroentérites ou de diarrhées pendant l'enfance serait associée aux patients souffrant de MII (16) tout comme une histoire d'infections fréquentes à un jeune âge ou une exposition à des antibiotiques (77). Dans le cas de MC pédiatrique, des marqueurs d'infection tels qu'aller à la garderie pendant les six premiers mois suivant la naissance, partager le lieu de résidence avec plusieurs autres individus et avoir un animal de compagnie à la maison ont été associés au développement de la maladie. De plus, le moment où l'infection survient pourrait être important dans l'étiologie de la MC chez les enfants (67). Par ailleurs, des changements dans la flore intestinale peuvent affecter les fonctions immunitaires d'un individu et une exposition subséquente à des organismes pathogènes inhabituels peut mener à une maladie gastro-intestinale (78). Dans le cas de la MC, il est reconnu que des individus ayant des prédispositions génétiques à cette maladie sont plus sujets au développement d'une inflammation chronique et incontrôlée qui pourrait être issue d'un contact avec un facteur extrinsèque tel un agent pathogène environnemental (79).

iv) L'infection par un microorganisme et la MC

Quant à l'organisme pathogène qui déclencherait la MC, plusieurs hypothèses ont été suggérées. Il a été observé que les gens atteints d'une MII ont généralement une concentration de bactéries muco-sales plus élevée que les individus en santé. De plus, il y aurait une augmentation progressive de la concentration de ces bactéries muco-sales avec la sévérité de la maladie, autant dans le colon en inflammation que dans le colon sain (80). Par ailleurs, il a été démontré que la microflore fécale des patients atteints de MII diffère de celle des témoins sains (81). En fait, une proportion considérable de la microflore dominante des patients ayant une MC active (environ 30% des espèces présentes lors d'un épisode actif de MC) appartient à des groupes phylogénétiques non-communs, suggérant qu'il existe des différences majeures entre la MC et un état de santé normal. De plus, la communauté microbienne fécale des individus atteints de MC a été identifiée comme étant instable en structure selon le temps (82). Cette instabilité pourrait refléter une susceptibilité accrue aux microorganismes environnementaux chez les individus atteints de la MC (83). Il devient donc complexe de faire la différence entre les organismes pathogènes possiblement associés au développement de la MC et ceux de passage dans les intestins des patients malades.

A ce sujet, plusieurs microorganismes environnementaux font l'objet de recherches quant à leur possible association à la MC. Il est possible de regrouper ces études en deux larges spectres : 1) les recherches qui postulent que la MC est initiée par des bactéries commensales de l'être humain, ou 2) par des microorganismes pathogènes pouvant être présents dans l'environnement tels *Mycobacterium*, *Yersinia*, *Escherichia coli*, *Campylobacter*, *Clostridium*, *Chlamydia*, etc. ou encore certains virus tels herpès et rotavirus (84).

v) Maladies entériques à déclaration obligatoire et MC

Ces observations préliminaires ont mené un groupe de recherche du Manitoba à évaluer si le taux d'incidence des maladies entériques à déclaration obligatoire les plus fréquentes ne serait pas associé au taux d'incidence de la MC. Ils ont obtenu une association négative entre la MC et l'incidence combinée de quatre maladies entériques (campylobactériose, salmonellose, shigellose et giardiase). En effet, il ont observé qu'il existerait une probabilité 50% plus élevée de développer la MC dans les endroits où les

taux d'incidence de ces infections entériques sont faibles (85). Les auteurs mentionnent que leurs résultats vont dans le sens de l'hypothèse hygiénique.

1.4.6 Origine ethnique – immigration, autochtones, descendance juive

L'importance de la race et de l'origine ethnique dans le risque de développer une MII est supportée par plusieurs études. En effet, il a été démontré qu'il existe des disparités entre différentes races quant à l'incidence, la localisation de la MC et aux complications extra-intestinales qui en résultent (86). Les MII ont été historiquement considérées comme affectant principalement les individus blancs. Cette impression pouvait toutefois résulter du fait que la majorité des études à ce sujet étaient faites dans des groupes au sein desquels les minorités non-blanches étaient sous-représentées (87). Désormais, les MII sont reconnues comme étant en croissance chez diverses populations ethniques (88).

Il est possible de croire que la variation dans l'incidence de la MC entre différents groupes ethniques serait en partie due au fait que les marqueurs de risque génétiques ne sont pas distribués de façon uniforme entre les populations. En effet, il a été démontré que la fréquence allélique des polymorphismes du gène *NOD2/CARD15* varie de façon significative parmi les sous-groupes ethniques (89). Par exemple, les SNP dans le gène *NOD2/CARD15* associés à la MC n'ont pas été identifiés dans des échantillons de la population de Gambie, du Japon et de la Chine, toutes en apparence résistantes à la MC. En effet, ces SNP seraient plutôt des allèles de risque pour les Caucasiens (90) (26).

Par ailleurs, chez les Maoris de Nouvelle-Zélande, une différence marquée dans la fréquence allélique du gène *NOD2/CARD15* a été observée par rapport aux groupes caucasiens de ce pays. Une corrélation linéaire a de plus été observée entre la diminution de la fréquence allélique d'une variante génétique retrouvée en association avec les polymorphismes de *NOD2/CARD15* et l'augmentation du degré d'ancêtres maoris (91). De ce fait, la MC est une maladie non-commune chez les Maoris. On observe un phénomène semblable chez les Indiens d'Amérique chez qui l'incidence de la MC est extrêmement rare (92).

Aussi, la prévalence et l'incidence de la MC est de deux à quatre fois plus élevée chez les individus de descendance juive ashkénaze que chez n'importe quel autre groupe ethnique (93). De plus, le phénomène d'anticipation génétique est observé chez les Juifs

ashkénazes d'Amérique, c'est-à-dire que la maladie se développe plus tôt chez les enfants d'individus atteints de la MC chez qui la maladie s'était développée plus tardivement (29). Il a été démontré que le bagage génétique des Juifs ashkénazes les prédisposerait au développement de la MC. En effet, les individus juifs ashkénazes atteints de MC sont porteurs des trois mutations de *NOD2/CARD15* associées à la MC à un taux de 47,4% par rapport à 27,45% chez les Juifs sépharades (94).

L'immigration offre quant à elle une opportunité d'étudier l'association entre l'environnement, les changements dans le mode de vie des individus et leur impact sur l'expression des MII (22). Les prédispositions génétiques à la MC jouent un rôle important dans l'étiologie de cette maladie, mais il a été observé que certaines populations immigrantes ont un taux d'incidence de la MC à la hausse dans leur pays d'accueil, tandis que dans leur pays d'origine, cette maladie est rare, voire même inexistante. Le fait que les polymorphismes génétiques connus jusqu'à ce jour expliquent jusqu'à 20% du risque associé à la MC indique que d'autres facteurs, génétiques et/ou environnementaux, contribuent probablement aussi à la susceptibilité à la maladie (95). Aux États-Unis par exemple, il existe des différences significatives dans les sous-types de MII ainsi qu'entre les différents marqueurs sérologiques selon les groupes ethniques et raciaux qui sont affectés. En effet, les individus afro-américains et les caucasiens sont affectés de façon prédominante par la MC, tandis que les Mexicains américains affectés par une MII sont plutôt atteints de colite ulcéreuse (88). Au Royaume-Uni, il a été démontré que les jeunes Asiatiques (Indiens, Pakistanais et Bengalis) de seconde génération sont significativement 6,1 fois plus à risque de développer les MII avant l'âge de 26 ans par rapport à la population européenne indigène. (96). De plus, lors d'une étude sur l'incidence des MII chez les Bengalis (originaires du Bangladesh) résidant à Londres-Est, il a été démontré qu'il y a une augmentation significative du nombre de ces individus qui développent la MC par rapport à la population en général. Il est à noter que les MII sont reconnues pour leur faible prévalence chez les individus vivant au Bangladesh. (97).

On observe un phénomène semblable en Colombie-Britannique. En effet, le Canada, a connu un virage démographique durant les vingt dernières années avec l'arrivée d'immigrants provenant de régions géographiques variées. Il a été démontré que l'incidence de la MC chez les patients d'origine indienne est passée de 5,95 cas / 10⁵ personnes / année en 1996 à 18,01 cas / 10⁵ personnes / année en 2001 (22).

Par ailleurs, la première instance de MC familiale rapportée chez une population immigrante a été observée dans une famille indo-canadienne. En effet, tous les membres de cette famille ont développé la MC après une résidence prolongée au pays. (98). Un tel phénomène a par la suite été observé en Belgique dans une famille d'immigrants d'origine marocaine où le père a développé la MC après leur arrivée et par la suite, quatre des huit enfants de la famille l'ont développée. Toutefois, aucune variation génétique associée à la MC n'a été retrouvée dans cette famille. Cette étude suggère la présence d'un facteur environnemental majeur qui ne serait pas relié à l'hygiène à l'enfance (99).

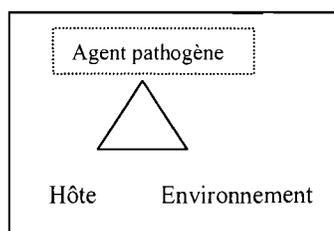
D'un autre côté, il a été postulé que dans le cas des individus de race noire, les disparités apparentes de la MC identifiées selon la race seraient plutôt dues à des facteurs sociaux-économiques et non pas à la maladie elle-même. En effet, il a été observé que les individus de race noire auraient des présentations rapportées de la MC similaires à celles des individus caucasiens et le cours de la maladie se déroule aussi de façon similaire. (100).

L'apparition précoce de la MC chez les caucasiens par rapport aux populations immigrantes soutient le rôle joué par l'amélioration de l'hygiène et des moyens sanitaires des pays développés dans l'étiologie des MII (45).

1.4.7 Agent pathogène : *Mycobacterium avium* sp. *paratuberculosis*

Le travail initié par Ogura et ses collaborateurs en 2001 sur la variante Leu1007fsinsC du gène *NOD2/CARD15* suggère que cette susceptibilité génétique pourrait potentiellement être reliée à un déficit du système immunitaire dans la reconnaissance des bactéries présentes dans un environnement gastro-intestinal (5) (voir figure 4, p. 25).

Figure 4 : Triade épidémiologique de causalité - implication d'un agent pathogène



Plus récemment, d'autres variantes génétiques ont été associées à la MC et ces découvertes semblent supporter l'évidence qu'un (ou des) agent(s) pathogène(s) serait impliqué dans le développement de la MC. À ce sujet, Parkes et ses collaborateurs ont émis cette observation :

« Prises ensemble, les évidences génétiques concernant IRGM, ATG16L1 (convergeant vers le processus d'autophagie), CARD15 et IL23R impliquent fortement des déficits dans la réponse immunitaire précoce, particulièrement dans les processus d'immunité innée et la prise en charge de bactéries intracellulaires dans la pathogénèse de la maladie de Crohn.⁵ » (8)

Plusieurs organismes pathogènes ont été suggérés comme étant impliqués dans la causalité de la MC tel qu'introduit à la section 1.4.5 et celui qui serait le plus plausible est *Mycobacterium avium* sp. *paratuberculosis* (MAP), une bactérie intracellulaire membre de la famille des mycobactéries. Il est désormais reconnu que certaines maladies dont la cause établie est une infection d'origine mycobactérienne, telles la tuberculose ou la lèpre, dépendent aussi d'une susceptibilité génétique de l'hôte (101). Ainsi, la MC et la tuberculose auraient en commun le fait qu'elles sont toutes deux des maladies qui dépendent d'une transmission génétique complexe et qui se développent seulement chez une proportion d'individus exposés à l'agent causal (95).

MAP est la cause d'une maladie gastro-intestinale des animaux qui affecte principalement les ruminants appelée paratuberculose, ou maladie de Johne. C'est en 1895 que deux vétérinaires, Johne et Frothingam, décrivent les caractéristiques d'un microorganisme présent dans les intestins enflammés d'une vache et cette bactérie fut plus tard identifiée comme étant MAP (10). Par la suite, c'est en 1913 que le chirurgien écossais Dr T.K. Dalziel émet l'observation que les symptômes histologiques retrouvés chez des patients humains atteints d'une entérite étaient similaires à ceux qui avaient été antérieurement décrits chez la vache. C'est à partir de cette observation qu'il suggéra l'existence d'une cause commune entre les deux maladies animale et humaine. Il émit donc l'hypothèse que MAP serait associée à la MC (11).

MAP fait partie de la famille des Mycobacteriaceae et la composition de sa paroi cellulaire riche en lipides lui confère les caractéristiques des bacilles acido-alcool-résistants (BAAR). En microscopie, les BAAR se colorent en rouge lorsqu'ils sont

⁵ Traduit de l'anglais : « Taken together, the genetic evidence regarding IRGM, ATG16L1 (converging on autophagy pathways), CARD15 and IL23R strongly implicates defects in the early immune response, particularly innate immune pathways and the handling of intracellular bacteria, in the pathogenesis of Crohn's disease. » (8)

colorés par la méthode de Ziehl-Neelsen. MAP est difficile à mettre en culture pour deux principales raisons : c'est un microorganisme à croissance lente qui a une capacité réduite de capter le fer du milieu ambiant. En effet, MAP ne peut produire le sidérophore mycobactin commun aux mycobactéries. C'est pourquoi il est essentiel d'ajouter aux milieux de culture du fer chélaté par un sidérophore synthétique, le Mycobactin J.

Par ailleurs, MAP fait partie du « Complexe *Mycobacterium avium* » (MAC) et possède un génome qui est à 99% homologue à celui de la sous-espèce *M. avium avium* (MAA), une cause de tuberculose chez les oiseaux. MAA et MAP sont deux des quatre sous-espèces du MAC (102).

Depuis l'hypothèse suggérée par Dalziel en 1913, de nombreuses études ont été menées afin d'isoler MAP à partir de patients atteints de la MC. En 1984, une espèce de mycobactéries non-classifiée a été isolée chez des cas de MC. Ce microorganisme avait des caractéristiques communes à MAP, telles l'apparence, la dépendance au Mycobactin J pour croître et 18 mois d'incubation avaient été requis pour obtenir une isolation primaire (103). Plus récemment, une bactérie faisant partie de l'espèce *Mycobacterium avium* a été détectée par la méthode de microscopie à fluorescence dans 59% des échantillons de tissus humains provenant de chirurgies effectuées chez des patients atteints de MC, et dans 14% des tissus témoins. Le rapport de cotes associant ce microorganisme à la MC dans cette étude est ainsi de 8,6 ($p = 0,0002$). Les limitations des techniques de microscopie n'ont toutefois pas permis aux auteurs de confirmer que le microorganisme était MAP (104).

L'avènement des méthodes moléculaires telle la technique d'amplification en chaîne par polymérase (ACP, mieux connue sous l'acronyme anglais PCR) a alimenté l'espoir de démontrer une association entre MAP et la MC. Il a été démontré que MAP possède des séquences génétiques qui sont absentes des autres mycobactéries telle la séquence d'insertion IS900 (105). De nombreuses études ont cherché à établir l'existence (ou l'inexistence) d'une association entre MAP et la MC en se basant sur l'amplification de cette région du génome de MAP. Cependant, bien que IS900 ait largement été utilisé comme cible moléculaire, des doutes ont été soulevés quant à sa spécificité et sa validité en tant que proxy direct de MAP. En effet, des éléments « semblables à IS900⁶ » ont été retrouvés chez d'autres organismes et d'autres séquences d'insertion du MAC partagent

⁶ Traduction des termes techniques anglais : *IS900-like elements*

des similarités considérables avec IS900 (106). Jusqu'à ce jour, la manière la plus simple et sûre d'établir une distinction sans ambiguïté entre les espèces et sous-espèces du complexe MAC au niveau moléculaire est le séquençage de la région 3' du gène *hsp65* (107).

Certaines études ont observé une association entre MAP et la MC, mais les résultats varient. Une revue systématique jointe à une méta-analyse d'études qui emploient des techniques basées sur les acides nucléiques pour détecter MAP chez des patients atteints de la MC a évalué que l'estimé combiné de différence de risque était de 0,23 (IC 95% : 0,14-0,32), démontrant qu'il y a une association entre MAP et la MC (108). Par exemple, la cible génétique IS900 de MAP a été détectée par PCR chez un plus grand nombre de cas de MC que de témoins (83,3% par rapport à 10,3% (OR : 43,3, $p < 0,0001$)) et MAP a été mise en culture à partir d'échantillons de tissus de cas de MC (chez 63,3% des cas et 10,3% des témoins (OR : 14,9, $p < 0,0001$)) (109). D'un autre côté, MAP n'a pas été associée à 11 cas de MC dans une étude qui avait utilisé l'amplification de *hsp65* pour détecter la bactérie (110).

Par ailleurs, une des limitations de la technique PCR part du fait que la détection d'une séquence génomique par cette méthode d'amplification ne peut que démontrer la présence d'acides désoxyribonucléiques (ADN) d'une bactérie dans un milieu, sans toutefois rendre compte de sa viabilité, donc de son potentiel de propagation et d'infection (111). Il n'existe de plus aucune validation des tests moléculaires de détection dû au fait qu'il est difficile d'obtenir un contrôle sans équivoque avec lequel comparer les résultats de PCR puisque MAP est un organisme fastidieux à cultiver (112).

Plus récemment, soit en mars 2007, il a été démontré que l'utilisation de traitements antibiotiques appropriés pourrait être bénéfique pour certains patients atteints de MC. En effet, des antibiotiques ciblant les mycobactéries atypiques (clarithromycine, rifabutine et levofloxacine) ont permis d'éliminer les traces d'ADN de MAP qui étaient présentes dans le sang d'un patient atteint de MC et cet événement est associé à une rémission de la maladie (113).

Ces obstacles causent une dichotomisation des opinions quant au lien de causalité entre MAP et la MC. Bien que certaines études démontrent qu'il existe une association entre

MAP et la MC, le critère d'« association » n'est qu'un paramètre parmi d'autres employés en épidémiologie pour identifier l'existence d'une causalité entre deux éléments (104). Malgré tout, il existe bon nombre d'évidences épidémiologiques suggérant un lien de causalité entre cette bactérie et la maladie (9). À ce sujet, une méta-analyse a démontré qu'il existe une forte association entre MAP et MC (OR = 7.0) tout en mentionnant l'hétérogénéité des découvertes des différentes études (114). Par ailleurs, l'incapacité des laboratoires diagnostics à détecter MAP chez les humains peut simplement refléter les limitations techniques actuelles, reliées au fait que MAP est un organisme à croissance lente qui requiert des conditions de culture spéciales (101).

Whittington et ses collaborateurs ont émis une suggestion intéressante qui pourrait s'appliquer à la recherche de causalité entre MAP et la MC :

« Les priorités de recherche en épidémiologie seraient de faire des cultures de sources d'eau dans des zones où il existe une agrégation géographique locale de maladie de Johne et de modéliser la propagation de la maladie de Johne par l'eau ou par aérosol, en utilisant les systèmes de données d'information géographique ainsi que les cartes de distribution des fermes affectées par la maladie de Johne⁷. » (115)

MAP a été mise en évidence sur tous les continents, mais son incidence reste sous-estimée dans la plupart des pays du fait que la maladie de Johne n'est pas une maladie à déclaration obligatoire (116). Au Canada, un programme volontaire de surveillance et de contrôle de la paratuberculose a été mis en place. Au Québec, les enquêtes de la Fédération des Producteurs Laitiers du Québec et la Fédération des Producteurs Bovins du Québec ont permis d'évaluer que 41,7% des troupeaux d'élevage laitier sont positifs pour MAP et 2,4% d'animaux sont séropositifs (117).

⁷ Traduit de l'anglais : « *Epidemiological research priorities would be to culture water sources in areas where there is local geographic clustering of Johne's disease and to model the spread of Johne's disease by water or aerosol, using geographic information system data and distribution maps of Johne's disease-affected farms.* » (115)

Partie 2 : Hypothèses et objectifs spécifiques

2.1 Hypothèses

2.1.1 Épidémiologie de la maladie de Crohn et variables associées

Le taux d'incidence de la MC est distribué de façon hétérogène dans la province de Québec. La variance des taux d'incidence de la MC entre les CLSC peut être en partie expliquée par des facteurs associés à la maladie. La variance non-expliquée par ces facteurs pourrait-être due à la présence d'un microorganisme pathogène retrouvé dans l'environnement, c'est-à-dire MAP.

2.1.2 Définition opérationnelle de cas

La définition opérationnelle utilisée pour identifier les cas dans la banque de données de la RAMQ et dans le registre Med Echo possède une sensibilité ainsi qu'une spécificité d'au moins 80%.

2.2 Objectifs

2.2.1 Valider la définition opérationnelle de cas de MC en comparant les cas identifiés dans la banque de données administratives à des cas cliniques de MC.

2.2.2 Rassembler les données permettant la construction des variables et des sous-variables choisies (caractéristiques individuelles, démographiques, géographiques et reliées à la condition de santé) afin de créer une banque de données divisée selon les 166 régions de CLSC du Québec.

2.2.3 Déterminer le taux d'incidence moyen de la MC au Québec, standardisé pour l'âge et le sexe selon la population du Québec pour la période 1995-2000.

2.2.4 Déterminer le taux d'incidence moyen de la MC au Québec, standardisé pour l'âge et le sexe selon les populations de cinq provinces canadiennes pour la période 1995-2000.

2.2.5 Déterminer dans quelles régions on retrouve la plus haute et la plus faible incidence de la MC au Québec et visualiser ces informations grâce à des cartes.

2.2.6 En utilisant le modèle de régression de Poisson, faire des analyses afin de déterminer quelles variables populationnelles sont associées à la MC au Québec.

PARTIE 3 : Méthodologie

3.1 Données

3.1.1 Identification des cas de maladie de Crohn

La MC n'étant pas une maladie à déclaration obligatoire, l'identification des diagnostics à partir de banques de données est donc la façon la plus efficace d'apprécier la vitesse d'apparition de cette maladie. En effet, les nouveaux diagnostics apparus au cours d'une période de temps d'observation représentent les cas incidents (118). Lors d'une étude précédente, les cas incidents de MC avaient été identifiés pour la période 1993-2002 à partir de deux banques de données administratives : celle de la RAMQ et le fichier Med Echo (13) (voir section 3.1.7). Pour cet ouvrage, la période à l'étude est **1995-2000**.

Dans les registres RAMQ et Med Echo, les diagnostics de MC sont représentés par des codes diagnostiques qui proviennent de la neuvième édition de la classification internationale des maladies (CIM-9) (119). Cinq codes ont initialement été retenus pour identifier la MC (13), soit les codes 555 (entérite régionale), 5550 (entérite régionale de l'intestin grêle), 555.1 (entérite régionale du gros intestin), 555.2 (entérite régionale de l'intestin grêle et du gros intestin) et 555.9 (entérite régionale de localisation non précisée). Lors d'une étude précédente, les diagnostics de MC extraits des deux banques de données ont été regroupés en une seule cohorte à laquelle a été appliquée la définition opérationnelle de cas élaborée par Bernstein et ses collaborateurs (12) afin d'obtenir une banque de données spécifique aux individus identifiés comme étant des « cas de MC » pour le Québec.

3.1.2 Définition opérationnelle de cas

Bernstein et ses collaborateurs ont développé et validé une méthode d'identification des cas de MC à partir de la banque de données du régime d'assurance provincial de Santé Manitoba, un régime similaire à celui de la RAMQ. Dans ces banques de données administratives gouvernementales ainsi que dans le fichier Med Echo, les individus sont identifiés par un code unique qui est leur numéro d'assurance maladie (NAM). Le NAM d'une personne assurée est composé de 12 caractères alphanumériques : les trois premières lettres du nom de famille à la naissance, la

première lettre du prénom, les deux derniers chiffres de l'année de naissance, les deux chiffres du mois de naissance (auquel on a additionné 50 pour indiquer le sexe féminin si tel est le cas), les deux chiffres du jour de naissance et les deux chiffres correspondant à un code administratif de la Régie (120).

Selon la **définition opérationnelle de cas**, pour qu'un individu soit considéré un « cas » pour la période à l'étude à partir des banques de données de la RAMQ et de Med Echo, il a dû remplir les critères suivants :

- Correspondre à au moins un des codes diagnostics mentionnés à la section 3.1.1

ET

- si la durée de couverture par le régime de santé provincial est 2 ans et plus pendant la période à l'étude (initialement 1993-2002 (13)), avoir eu un minimum de 5 contacts médicaux distincts avec un code diagnostic de MC pendant la période à l'étude

OU

- dans le cas où la durée de couverture par le régime de santé provincial est inférieure à 2 ans, avoir accumulé un minimum de 3 contacts médicaux distincts avec le diagnostic de MC pendant ces deux ans

Dans cette étude, nous avons choisi de restreindre l'utilisation des cas identifiés entre les années 1995 et 2000. Ceci a pour but d'éviter de confondre des cas prévalents avec des cas incidents au début de la période 1993-2002. De plus, à la fin de la période 1993-2002, il se peut que les individus n'aient pas le temps de cumuler assez de contacts médicaux pour être identifiés comme des cas par la définition. Nous avons ainsi choisi d'éliminer les années 1993, 1994, 2001 et 2002.

3.1.3 Validation des cas identifiés par la définition de cas

Bien que la définition de cas élaborée par Bernstein et ses collaborateurs ait été antérieurement validée, nous avons jugé essentiel de valider les cas identifiés au Québec dans la banque de la RAMQ et du registre Med Echo à l'aide de cette définition pour la

période 1993 à 2002. Pour ce faire, nous avons utilisé comme groupe de référence un ensemble de cas confirmés de MII (MC ou CU) identifiés entre 1993 et 2001 dans les cliniques de gastroentérologie de deux hôpitaux du Centre universitaire de Santé McGill (CUSM) à Montréal, soit l'Hôpital Général de Montréal et l'Hôpital Royal-Victoria.

Afin d'être en mesure d'effectuer une comparaison entre les cas identifiés par la définition et les cas cliniques, un identifiant commun aux deux groupes a été choisi : le NAM. En effet, les individus faisant partie de la banque de données de la RAMQ sont identifiés par leur NAM afin qu'aucune information nominale ne soit accessible lors de la consultation de la banque de données. Ainsi, pour effectuer la comparaison entre les deux groupes de cas, le NAM de chaque cas clinique a dû être retracé par la RAMQ. Les informations requises par la RAMQ afin de retracer les NAM étaient les suivantes :

- Nom
- Prénom
- Date de naissance
- Code postal
- Diagnostic (MC ou CU)
- Sexe
- Date de diagnostic

Toutefois, puisque des informations nominales figuraient dans cette liste, nous avons dû obtenir une approbation des comités éthiques des deux centres de santé impliqués dans le processus de validation, l'accord de la Commission d'Accès à l'Information (CAI) concernant le transfert d'information nominale à la RAMQ ainsi qu'un consentement écrit et éclairé des patients.

Afin d'obtenir cette approbation, un groupe de 97 patients a tout d'abord été contacté par téléphone afin de valider les informations concernant leur lieu de résidence et la langue dans laquelle l'information devait leur être envoyée. Après avoir obtenu l'adresse correcte de 87 patients, une enveloppe leur a été envoyée le 17 novembre 2006, contenant un formulaire de consentement (version française ou anglaise selon la demande), une lettre décrivant le projet et préalablement signée par leur gastroentérologue, et une enveloppe de retour préaffranchie et pré-adressée. Dix jours ouvrables après l'envoi, les patients qui n'avaient pas retourné leur formulaire ont été

contactés par téléphone. Un deuxième rappel téléphonique a été fait vingt jours après l'envoi.

Par ailleurs, les NAM contenus dans la banque de données de la RAMQ avaient été encryptés lors de l'étude précédente afin que ces identifiants ne soient pas révélés lors de l'identification des cas dans la banque de données. Afin de rendre la comparaison possible entre les deux groupes de cas, la liste qui regroupait les cas cliniques a été envoyée à la RAMQ afin qu'ils encryptent les NAM de chacun des cas cliniques avec la même clef de cryptage que celle utilisée par le passé pour la banque de données de la RAMQ.

Finalement, la RAMQ nous a retourné une liste comprenant les cas identifiés par la définition et les cas cliniques, en indiquant si le diagnostic associé à chaque cas par la définition de cas correspondait au diagnostic posé en clinique.

Tableau III : tableau 2X2 utilisé afin de mesurer la validité de la définition de cas

		DIAGNOSTIC CLINIQUE	
		MC (+)	Autres (-)
DÉFINITION OPÉRATION- NELLE DE CAS	MC (+)	a	b
	Autres (-)	c	d

La sensibilité et la spécificité de la définition de cas ont pu être évaluées à partir de ces données. La sensibilité évalue combien de cas sont identifiés comme étant positifs (diagnostic de MC) par la définition de cas tout en étant un cas réel identifié par les critères cliniques. À partir des cellules du tableau III en page 35, la sensibilité correspond à :

$$a / (a+c)$$

Équation 1 : Calcul de sensibilité selon le tableau 2X2

La spécificité évalue le nombre de faux-positifs, c'est-à-dire le nombre de cas identifiés par la définition comme étant positifs (diagnostic de MC), mais qui ne le sont pas selon les critères cliniques. À partir des cellules du tableau III en page 35, la spécificité correspond à :

$$d / (b+d)$$

Équation 2 : Calcul de spécificité selon le tableau 2X2

Par ailleurs, le rendement, ou nombre de cas détectés par la définition opérationnelle de cas, a été mesuré par la valeur prédictive du test. La valeur prédictive positive (VP+) évalue la présence réelle ou non de la maladie chez un individu, par rapport aux résultats du test de dépistage – ici, la définition opérationnelle de cas (121). À partir des cellules du tableau III en page 35, la valeur prédictive positive correspond à :

$$a / (a+b)$$

Équation 3: Calcul de la valeur prédictive positive selon le tableau 2X2

La valeur prédictive négative (VP-) correspond à la probabilité qu'un individu réellement non malade ne soit pas détecté par la définition opérationnelle. À partir des cellules du tableau III en page 35, la valeur prédictive négative correspond à :

$$d / (c+d)$$

Équation 4: Calcul de la valeur prédictive négative selon le tableau 2X2

3.1.4 Données de population à risque

La population à risque a été établie comme étant la population du Québec en 1996 (milieu de la période initiale de 1993-2002). Les données de population pour chaque CLSC ont été fournies par le Ministère de la Santé et des Services sociaux

(MSSS) et les données initiales proviennent du recensement de 2001 de Statistique Canada.

3.1.5 Données géographiques – régions de CLSC

Depuis 2005, la province du Québec est divisée en 166 territoires de CLSC, couramment appelés « districts ». Il s'agit du plus bas niveau de découpage territorial dans le domaine socio-sanitaire (89). En 2005, le MSSS a effectué des fusions entre certains CLSC, faisant passer leur nombre de 169 à 166. En effet, les CLSC 1302 et 1305 ont été fusionnés pour donner le CLSC 1061 Témiscouata. De plus, les CLSC 2102 et 2106 ont été fusionnés pour donner le CLSC 2107 Saguenay. Finalement, les CLSC 5107 et 5108 ont été fusionnés pour donner le CLSC 5071 Sherbrooke. Des fusions entre d'autres CLSC ont aussi eu lieu dans les années antérieures (voir tableau de correspondance de la codification des CLSC en annexe 5). Ainsi, il a été nécessaire d'attribuer le code de CLSC 2005 aux cas de MC identifiés par la définition de cas dont la codification du CLSC d'origine était antérieure à 2004 afin d'assurer une correspondance valide entre les cas de mêmes CLSC.

3.1.6 Données démographiques et associées à la condition de santé

Des facteurs potentiellement associés à la MC ont été identifiés dans la littérature et des données populationnelles ont été collectées pour l'ensemble des CLSC du Québec afin de construire les variables à étudier. Les données démographiques et associées à la condition de santé de la population de chaque CLSC portaient sur :

- Le lieu de résidence de la population (urbain ou rural)
- La proportion de la population qui est immigrante
- La proportion d'individus de descendance juive dans la population
- L'utilisation de contraceptifs oraux chez les femmes en âge de procréer
- L'incidence de certaines maladies entériques à déclaration obligatoire
- Le pourcentage de tabagisme dans la population
- La proportion de la population qui se définit autochtone
- Le statut socio-économique moyen de la population

Les données obtenues proviennent de différentes sources (voir section 3.1.7) et sont décrites de manière détaillée à la section 3.3.2.

3.1.7 Source des données et définition des banques de données

Les données utilisées dans cet ouvrage proviennent de diverses sources définies dans cette section.

A. Banque de données de la Régie de l'Assurance Maladie du Québec (RAMQ)

La Régie de l'Assurance Maladie du Québec dispose d'une grande quantité de renseignements pouvant favoriser le travail des chercheurs et la prise de décision des professionnels de la santé. Tout en préservant la confidentialité des données, elle peut fournir de l'information sur les services couverts par les régimes d'assurance maladie et d'assurance médicaments (122). Cette banque de données inclut les personnes qui sont sous le régime public d'assurance maladie du Québec qui couvre l'ensemble de la population de la province. Deux types de codes sont entrés dans la banque de données de la RAMQ lorsqu'un patient rencontre un médecin à un bureau privé ou à l'urgence d'un hôpital : les codes de facturation et les codes diagnostiques (13). Ces codes sont entrés dans le système lorsque le patient est hospitalisé ou non. Le NAM, le sexe, la date de naissance, le code de CLSC associé au patient ainsi que la date des diagnostics sont des informations qui font partie de la banque de données. Les cas de MC utilisés dans cet ouvrage ont été extraits de cette banque de données.

B. Fichier central Med Echo

Le système Med Echo contient les données médico-hospitalières relatives aux hospitalisations survenues dans les centres hospitaliers dispensant des soins généraux et spécialisés. Les informations concernant les soins de courte durée et les chirurgies d'un jour sont donc compilées par les établissements avant d'être acheminées au MSSS pour alimenter la banque centrale de données Med Echo (123). Lors de l'hospitalisation d'un patient, des entrées sont faites dans ce fichier central. Il y a ainsi création d'un dossier sommaire d'hospitalisation qui contient un diagnostic principal, des diagnostics secondaires (de 0 à 15) et jusqu'à 9 codes de traitement. Les cas de MC utilisés dans cet ouvrage ont aussi été extraits de cette banque de données.

C. Ministère de la Santé et des Services sociaux (MSSS)

Le secteur de la santé et des services sociaux dispose d'un réseau de services et de ressources, tant humaines et matérielles que financières, dédiées au bien-être de la population. Le MSSS a créé un « Atlas de la santé et des services sociaux » qui offre un accès en ligne à des données validées concernant les caractéristiques sociales et économiques, les habitudes de vie et l'environnement de la population du Québec. L'objectif de l'Atlas de la santé et des services sociaux du Québec est donc d'esquisser un portrait géographique de cette organisation : dresser un inventaire des ressources, localiser la population par rapport à celles-ci et illustrer ses caractéristiques socio-économiques ainsi que sa distribution sur le territoire (124). Lors de cette étude, des données ont été obtenues directement via le site web et d'autres nous ont été distribuées sur demande par la section de géographie du MSSS. Les données concernant l'immigration, la descendance juive, la population autochtone, l'urbanité, le tabagisme et le statut socio-économique pour chaque territoire de CLSC sont toutes des données provenant de Statistique Canada, mais elles nous ont été distribuées par l'entremise du MSSS.

D. Statistique Canada

Statistique Canada est l'organisme statistique national du Canada. On y produit des statistiques qui aident les Canadiens à mieux comprendre leur pays, sa population, ses ressources, son économie, sa société et sa culture (125). Cette organisation effectue un recensement de la population tous les cinq ans. Des données concernant le profil des communautés sont ainsi cumulées. Les données populationnelles concernant l'immigration, la descendance juive, la population autochtone et le statut socio-économique ont été compilées lors du recensement de 2001 et nous ont été transmises via le MSSS.

E. Banque de données d'IMS Canada

IMS Canada est le principal fournisseur mondial de renseignements et de services de consultation stratégiques à l'intention de l'industrie pharmaceutique et du secteur de la santé. Cette entreprise recueille des données auprès de plus de 65 000 sources au Canada, notamment les hôpitaux, les pharmacies, les fabricants et grossistes de l'industrie du médicament et les médecins. Chaque année, plus de 200 millions de transactions sont traitées pour créer une base de données complète sur les courbes de

diagnostic et de traitement, les habitudes de prescription et la consommation de médicaments (126). Les données concernant les ordonnances de contraceptifs oraux proviennent de cette banque de données.

F. Banque de données des MADO du Québec

Le MSSS a créé un fichier central qui compile les cas de maladies à déclaration obligatoire déclarées aux directions de santé publique (DSP) des régions, puis enregistrées et validées par celles-ci aux registres nominaux et anonymes (127). Pour être à déclaration obligatoire, l'intoxication, l'infection ou la maladie doit être susceptible de causer une épidémie si rien n'est fait pour la contrer, reconnue comme une menace importante pour la santé, nécessitant une vigilance des autorités de santé publique ou la tenue d'une enquête épidémiologique, évitable par l'intervention des autorités de santé publique ou d'autres autorités (128). Les cas de MADO sont déclarées à deux niveaux :

- Par les médecins lorsque ceux-ci diagnostiquent une intoxication, une infection ou une maladie inscrite à la liste ou lorsqu'ils constatent la présence de signes cliniques caractéristiques de l'une de ces intoxications, infections ou maladies chez une personne vivante ou décédée.

- Par tout dirigeant d'un laboratoire ou d'un département de biologie médicale, privé ou public, lorsqu'une analyse de laboratoire faite dans le laboratoire ou le département qu'il dirige démontre la présence de l'une de ces intoxications, infections ou maladies (128).

Au Québec, la déclaration des MADO est une obligation légale pour tous les médecins et laboratoires (127) et c'est le Laboratoire de santé publique du Québec (LSPQ) qui gère le fichier central MADO. De plus, le lieu de résidence permanente des personnes affectées attribue une région d'appartenance aux cas (128). Dans cet ouvrage, les données concernant les taux d'incidence des cinq MADO entériques (campylobactériose, yersiniose, giardiase, salmonellose, shigellose) ont été rendues disponibles par le Secteur Vigie et Protection de la DSP de Montréal.

3.2 Analyse descriptive

Tout d'abord, la MC peut être identifiée comme étant une maladie rare, c'est-à-dire qu'il y a moins de 5% de densité d'incidence sur la période à l'étude. Dans une telle situation, le nombre de personnes à risque reste relativement le même tout au long de la période d'observation. Cette caractéristique implique que le rapport de taux devient un estimateur sans biais du ratio du risque et du rapport de cotes qui estiment tous la force d'opération des différents facteurs de risque sur l'apparition de la maladie (118). Dans cette étude, les IRR sont employés lors des analyses statistiques.

3.2.1 Devis descriptif – l'étude écologique

Le devis descriptif permet l'étude de corrélations entre des facteurs pouvant être associés à une maladie et la maladie elle-même, tout en s'intéressant aux aspects temps, espace et personnes. Ce genre de devis ne donne lieu à aucune comparaison avec un groupe de référence, ou groupe témoin, puisqu'il ne fait que décrire un groupe avec la maladie.

Par ailleurs, les études écologiques, ou populationnelles, sont un type de devis descriptif qui impliquent l'analyse de deux échelles d'informations, soit l'échelle individuelle et populationnelle. Elles fournissent une distribution de la fréquence de la maladie en fonction de marqueurs de risque estimés au niveau populationnel (118).

Dans cette étude, le devis descriptif permet l'analyse de la *distribution de la fréquence de la MC sur un territoire en fonction de facteurs qui lui sont potentiellement associés*. Dans ce cas, la maladie est représentée par une fréquence populationnelle sous forme de taux et les facteurs de risque sont reflétés par une fréquence populationnelle incluant les exposés de même que les non-exposés (118). En effet, les mesures écologiques obtenues dans cet ouvrage sont de type agrégé, c'est-à-dire qu'elles sont des résumés d'observations (ex. moyennes ou proportions) dérivées de différents groupes d'individus (129).

L'un des principaux avantages à mener des études écologiques s'applique aux recherches portant sur de larges groupes et tel est le cas dans cet ouvrage. En effet, il est possible d'obtenir des mesures concernant les variables d'intérêt en le faisant au niveau

écologique dans le cas où il serait impossible d'obtenir de façon réaliste ces mêmes mesures au niveau individuel. D'un autre côté, les études écologiques s'appliquent aussi bien aux investigations d'agrégats apparents d'une maladie dans des zones restreintes afin de comprendre les différences qui existent entre les taux d'incidence au sein d'une population (129).

Les études écologiques ont ainsi pour but de suggérer des hypothèses de recherche qui peuvent être récupérées par des devis de recherche capables d'ouvrir la voie à un traitement rigoureux de l'information. En effet, puisqu'elles n'isolent pas les exposés des non-exposés ni les malades des non-malades, les études écologiques peuvent aisément générer des biais et des conclusions erronées (118).

3.2.2 Taux d'incidence de la maladie de Crohn au Québec

Le taux d'incidence, ou la densité d'incidence, permet de mesurer la vitesse d'apparition d'une maladie ou encore la force d'opération d'un ou de facteur(s) de risque sur l'apparition de la maladie (118). Dans le présent ouvrage, le taux d'incidence de la MC représente la variable dépendante des modèles de régression sur lesquelles des analyses statistiques descriptives et associatives ont été faites.

Afin de calculer le **taux d'incidence brut** de la MC pour chaque CLSC, nous avons utilisé la somme des cas rapportés entre 1995 et 2000 comme numérateur. Par ailleurs, le dénominateur est constitué des personnes-temps, c'est-à-dire la somme du temps contribué par les personnes à risque suivies pendant le temps de l'étude. Les personnes à risque correspondent à la population du Québec en 1996 de chacun des 166 CLSC. Le taux d'incidence brut est calculé par le rapport suivant :

$$\frac{\Sigma \text{ cas incidents rapportés pour la période de temps d'observation}}{\text{Personnes – temps}}$$

Équation 5 : Calcul du taux d'incidence brut

Où

Σ : symbole de sommation (retrouvé dans les équations 5, 6 et 7)

Par ailleurs, les effets possibles de l'âge et du sexe des patients sur l'incidence de la MC ont été considérés et ces variables ont été modélisées. Tout d'abord, cinq groupes d'âge ont été créés afin de calculer le **taux d'incidence brut par groupe d'âge**. Les intervalles utilisés sont : 0-19 ans, 20-29 ans, 30-39 ans, 40-54 ans et 55 ans et plus. De plus, la population du Québec en 1996 de ces cinq groupes d'âge de chacun des 166 CLSC a été utilisée comme population à risque (voir section 3.1.4). Il est à noter que lors des analyses statistiques, le groupe d'âge des 0-19 ans a été choisi comme groupe de référence. Par la suite, le taux d'incidence a été rapporté par 100 000 de population par année. Le taux d'incidence brut par groupe d'âge est calculé par le rapport suivant :

$$\frac{\Sigma \text{ des cas incidents } \textit{par groupe d'âge} \text{ rapportés pour la période de temps d'observation}}{\text{Personnes – temps}}$$

Équation 6 : Calcul du taux d'incidence brut par groupe d'âge

Finalement, le **taux d'incidence brut selon le sexe** des patients a également été calculé pour la durée totale de l'étude. La population du Québec en 1996 correspondant aux deux sexes de chacun des 166 CLSC a été utilisée comme population à risque (voir section 3.1.4). Il est à noter que lors des analyses statistiques, le sexe « masculin » a été choisi comme groupe de référence. Le taux d'incidence brut selon le sexe est représenté par le rapport suivant :

$$\frac{\Sigma \text{ des cas incidents } \textit{par sexe} \text{ rapportés pour la période de temps d'observation}}{\text{Personnes – temps (durée totale d'observation)}}$$

Équation 7 : Calcul du taux d'incidence brut selon le sexe

3.2.3 Standardisation directe des taux d'incidence pour l'âge et le sexe

La standardisation directe, ou méthode de la population-type, permet de comparer des taux d'incidence en tenant compte des différences de distribution entre les populations (121). Pour ce faire, cette méthode utilise comme standard une population ayant une composition donnée pour un facteur pouvant introduire de la confusion lors des analyses. Ainsi, les taux d'incidence bruts correspondant aux différents niveaux de ce facteur de confusion sont appliqués aux effectifs de la population-type, eux aussi segmentés selon les différents niveaux du facteur, pour finalement donner les diagnostics ajustés. Le taux d'incidence standardisé est alors obtenu en faisant le quotient de la somme des diagnostics ajustés par l'effectif total de la population-type (130). Ainsi, lorsqu'une comparaison est faite entre des taux d'incidence standardisés pour un ou des facteur(s) particulier(s), l'observation de toute différence résiduelle entre les groupes ne peut pas être attribuée à ce facteur, mais à la variable d'intérêt introduite dans le modèle statistique à l'étude (121).

Dans le présent ouvrage, le taux d'incidence brut de la MC de chaque CLSC a été standardisé de façon directe pour l'**âge** et le **sexe** selon les effectifs de deux populations de référence : celle du Québec de l'année 1996 et celle de cinq provinces canadiennes de cette même année (les calculs de la standardisation directe sont présentés en annexe 2).

Tout d'abord, décrivons la standardisation pour la population du **Québec** de 1996. Lors du calcul des taux d'incidence bruts de chaque CLSC (voir section 3.2.2), les cas ont pu être divisés selon le sexe et en trois strates d'âge, c'est-à-dire : 1) 18 ans et moins 2) 19-64 ans et 3) plus de 65 ans. Ces groupes d'âge ont par la suite été utilisés pour diviser la population-type afin d'être en mesure d'appliquer les taux d'incidence bruts à la population-type.

Toutefois, il est à noter que les données initiales de la population référence ont dû être modifiées. En effet, celles-ci étaient déjà séparées en intervalles d'âge qui ont dû être élargis afin de mieux correspondre à la classification des cas. Trois groupes d'âge ont donc été générés pour la population de référence: 1) 0-19 ans 2) 20-64 ans et 3) plus de 65 ans. Ainsi, les strates d'âge selon lesquelles les cas ont été regroupés sont sensiblement différentes des groupes d'âge de la population-type. C'est pourquoi lors de

la standardisation directe, le taux d'incidence du groupe d'âge 18 ans et moins a été appliqué à la population du groupe d'âge 0-19 ans du Québec.

Par ailleurs, le taux d'incidence de la MC de chaque groupe d'âge a aussi été appliqué à la population combinée de **cinq provinces du Canada** (Colombie-Britannique, Alberta, Saskatchewan, Manitoba et Nouvelle-Écosse) pour l'année 1996 afin d'être en mesure d'établir des comparaisons entre le taux d'incidence standardisé moyen de la MC du Québec et les taux d'incidence standardisés moyens de la MC disponibles dans la littérature pour ces autres provinces canadiennes. La méthode de calcul est la même que pour la standardisation directe avec la population du Québec (voir annexe 2).

3.2.4 Visualisation des données d'incidence

Quatre cartes ont été élaborées afin de visualiser la répartition des taux d'incidence standardisés pour l'âge et le sexe pour la période 1995-2000 dans la province de Québec. Selon l'Atlas du Canada, « *les cartes sont des outils complexes permettant de visualiser et de diffuser des données géospatiales à échelle réduite* » (131). Les cartes du présent ouvrage ont été faites en collaboration avec la section de géographie médicale du LZA-SSH. L'aire de diffusion choisie est le territoire de CLSC et la codification des CLSC est celle de 2006 afin d'assurer une concordance avec les données d'incidence qui ont été faites selon la codification des territoires de CLSC de 2005 (à noter qu'il n'y a eu aucun changement dans la codification des CLSC entre 2005 et 2006). Des intervalles de classes ont été construits à l'aide de la méthode statistique de séparations naturelles (*Natural Breaks*), regroupant les taux d'incidence en six classes.

La méthode des **séparations naturelles** est basée sur la reconnaissance subjective de lacunes dans la distribution, où la diminution du nombre d'observations est significative. Tracer un histogramme des données permet de reconnaître ces lacunes. Cette méthode élaborée par George Jenks réduit la variation à l'intérieur des classes et augmente la variation entre les classes. Elle est très utile lorsque l'ensemble de données contient plusieurs valeurs modales (131).

Par ailleurs, la projection « Québec Lambert Conique Conforme » est le type de projection utilisée dans l'élaboration des cartes. Tout d'abord, une projection cartographique est un système de correspondance entre les coordonnées géographiques

et les points du plan de projection. Ces systèmes se basent sur la géométrie et plus couramment sur des formules mathématiques. La sélection de la meilleure projection cartographique dépend du but de l'utilisation de la carte. Par ailleurs, certaines des projections les plus simples se définissent d'après des propriétés géométriques générales, par exemple les cônes, les cylindres et les plans. Quant à la projection Québec Lambert Conique Conforme, elle est un type de plan, ou projection, utilisée pour dresser les cartes de régions situées au nord de l'équateur, tel le Canada, afin de diminuer la distorsion de l'ensemble des formes du territoire et des eaux. De manière plus spécifique, la projection Québec Lambert Conique Conforme est une projection spécifique pour le Québec (132).

Finalement, le *North American Datum* (NAD) 83 est une indication du réseau de repères géodésiques, c'est-à-dire des points précis sur la Terre dont les coordonnées sont connues et qui servent de référence. Le NAD 83 est généralement celui qui est utilisé pour l'Amérique du Nord.

3.3 Analyses associatives

3.3.1 Variables indépendantes principales

Les variables indépendantes étudiées dans les modèles de régression sont décrites au tableau IV en page 47. Ainsi, des données de type continu ont été catégorisées par la méthode des quantiles afin que les différents niveaux de chaque variable soient étudiés en relation avec l'incidence de la MC. La méthode de classification par quantiles implique que les données sont disposées par ordre ascendant et le nombre d'observations individuelles est compté. Les observations sont ensuite réparties dans le nombre sélectionné de classes, toutes les classes contenant un nombre semblable d'observations (131). Les classes ont pu être remodelées manuellement au besoin. Par ailleurs, la liste complète des sous-variables qui ont servi à construire les variables principales est fournie en annexe 1.

Tableau IV : Description des variables indépendantes principales

VARIABLES INDÉPENDANTES	TYPE DE VARIABLE	CATÉGORIES	UNITÉ	SOURCES*
1. Groupes d'âge	continue	0-19, 20-29, 30-39, 30-54, 55 et +	années	RAMQ- Med Echo, MSSS
2. Sexe	discrète catégorique	Féminin/masculin	-	RAMQ- Med Echo, MSSS
3. Urbanité	discrète catégorique	Urbain/rural	-	MSSS
4. Immigration	continue	0-10, 11-62	%	MSSS, Statistiques Canada
5. Descendance juive	continue	0-1, 2-19, 20-59	%	MSSS, Statistiques Canada
6. Contraceptifs oraux	continue	0-3 , 4 et plus	%	IMS Canada
7. Incidence de cinq MADO entériques	continue	1-6, 7-14, 15-100	Nouveaux cas / 10 ⁵ personnes / 1998	MADO, DSP
8. Tabagisme	continue	moins de 26, plus de 26	%	MSSS
9. Autochtones	continue	0-0,5, 0,6-3,5, 3,6 et plus	%	MSSS, Statistiques Canada
10. Statut socio- économique	discrète catégorique	Indice de défavorisation 1, 2, 3	-	Atlas du MSSS

*Les différentes sources des variables sont définies à la section 3.1.7

3.3.2 Définition des variables principales

Les variables indépendantes principales énumérées au tableau IV de la page 47 sont détaillées dans cette section. Il est à noter que les variables ont été obtenues pour chaque CLSC du Québec.

i) Immigration

Cette variable est la proportion de personnes immigrantes par rapport à la population totale exprimée en pourcentage de la population totale. Cette variable a été construite à partir des données collectées lors du recensement canadien de 2001 et fournies par le MSSS. Ces données provenaient initialement de Statistiques Canada (voir section 3.1.7). La population totale utilisée est celle de la province du Québec pour l'année 1996. Les catégories élaborées par la méthode des quantiles et utilisées pour les analyses statistiques sont : 0-10 % et 11-62 % d'immigrants pour la population totale de chaque CLSC.

ii) Ascendance juive

Cette variable est la proportion d'individus d'ascendance juive par rapport à la population totale exprimée en pourcentage de la population totale. Cette variable a été construite à partir des données collectées lors du recensement canadien de 2001 et fournies par le MSSS. Ces données provenaient initialement de Statistiques Canada (voir section 3.1.7). Les catégories élaborées par la méthode des quantiles et utilisées pour les analyses statistiques sont : 0-1%, 2-19% et 20-59% d'individus d'ascendance juive pour la population totale de chaque CLSC.

iii) Population autochtone

Cette variable est la proportion d'individus déclarés autochtones⁸ (Indien d'Amérique, Métis ou Inuit) par rapport à la population totale exprimée en pourcentage de la population totale. Cette variable a été construite à partir des données collectées

⁸ ***Identité autochtone**: personne ayant déclaré appartenir à au moins un groupe autochtone, c'est-à-dire Indien de l'Amérique du Nord, Métis ou Inuit (Esquimau) et/ou personne ayant déclaré être un Indien des traités ou un Indien inscrit tel que défini par la *Loi sur les Indiens* du Canada et/ou personne ayant déclaré appartenir à une bande indienne ou à une première nation. Au recensement de 1996, Statistique Canada a ajouté une question sur la propre perception du recensé face à son identité autochtone. La question du recensement de 2001 est la même que celle de 1996, c'est-à-dire : « Cette personne est-elle un Autochtone, c'est-à-dire un Indien de l'Amérique du Nord, un Métis ou un Inuit (Esquimau) ? ». (Source : Statistique Canada (2002) *Dictionnaire du recensement de 2001*. Éditeur : Ministre de l'industrie, page 42.)

lors du recensement canadien de 2001 et fournies par le MSSS. Ces données provenaient initialement de Statistiques Canada (voir section 3.1.7). Les catégories élaborées par la méthode des quantiles et utilisées pour les analyses statistiques sont : 0-0,5 %, 0,6-3,5 % et 3,6 % et plus d'individus déclarés autochtones pour la population totale de chaque CLSC.

iv) Contraceptifs oraux

Cette variable est l'estimé du pourcentage du nombre d'ordonnances de contraceptifs oraux de type 28 jours / femme âgée entre 15 et 40 ans / année. Cette variable a été construite à partir des données fournies par IMS Canada (voir section 3.1.7) et elle est une extrapolation des données de 2000-2006 à notre période à l'étude, soit 1995-2000. En effet, les données de consommation de contraceptifs oraux (CO) étaient disponibles seulement pour la période 2000-2006.

Afin d'établir s'il nous était possible d'extrapoler ces données pour la période 1995-2000, nous nous sommes assurés qu'il n'existait pas de tendance linéaire entre les années 2000-2006 en faisant une régression linéaire. Les résultats de l'analyse de régression ont été non significatifs ($p = 0,1748$) et ainsi les données ont pu être extrapolées à notre période à l'étude.

Pour être en mesure de bâtir le modèle de régression linéaire, nous avons combiné trois variables: les estimés des ordonnances totales dispensées par les pharmacies du Québec (septembre 2000 à 2006), les estimés du nombre de nouvelles ordonnances dispensées par les pharmacies du Québec (septembre 2000 à 2006) et les estimés du nombre d'ordonnances renouvelées dispensées par les pharmacies du Québec (septembre 2000 à 2006).

Par ailleurs, les catégories élaborées par la méthode des quantiles et utilisées pour les analyses statistiques sont : 0-3 et plus de 4 ordonnances de type 28 jours / année pour chaque CLSC. Les données utilisées incluent toutes les sortes de molécules sur le marché. Il est à noter que les données du CLSC Montréal Centre-ville (code : 6094) n'ont pas été utilisées puisque la consommation semble être surévaluée. On peut assumer que les femmes remplissent leurs ordonnances dans les pharmacies près de leur lieu de travail qui font partie d'un CLSC différent de leur lieu de résidence.

v) Urbanité

Cette variable est la classification des CLSC selon qu'ils sont urbains ou ruraux. Un CLSC est classifié « urbain » lorsque la proportion de codes postaux urbains tend vers 100%, tandis qu'il est classifié « rural » lorsque la proportion de codes postaux urbains tend vers 0. Cette variable a été construite en utilisant les données de codes postaux fournies par le MSSS (voir section 3.1.7).

Lorsqu'un code postal couvre une zone rurale, Poste Canada lui assigne un « 0 » à la deuxième position du code. Par le fait même, tous les codes postaux dont la deuxième position du code est un chiffre autre que « 0 » couvrent une zone urbaine. Pour chaque CLSC, nous avons fait la somme de tous les codes postaux urbains que nous avons rapporté sur le nombre total de codes postaux compris dans chaque CLSC. Cela nous a fourni une proportion de codes postaux urbains par CLSC. Nous avons établi qu'un CLSC est considéré urbain s'il comprenait une majorité de codes postaux ayant un deuxième identifiant différent de 0. Un CLSC est considéré rural s'il a une majorité de codes postaux ayant un « 0 » en deuxième position du code.

Par ailleurs, aucun seuil n'a été utilisé pour trancher entre « urbain » et « rural » puisque pour chaque CLSC, la proportion de codes postaux urbains tendait clairement soit vers 100% soit vers 0%. En effet, les pourcentages les plus éloignés des extrêmes étant 89% urbain et 24% rural (interprétés respectivement comme urbain et rural).

vi) Incidence de cinq maladies entériques à déclaration obligatoire

Cette variable est la moyenne de l'incidence moyenne de cinq maladies entériques à déclaration obligatoire exprimée par 10^5 personnes pour les années 1995 à 2000. Cette variable a été construite à partir des données d'incidence de quatre entérites bactériennes à déclaration obligatoire, soit la campylobactériose, la yersiniose, la salmonellose et la shigellose et d'une infection parasitaire, la giardiase.

Au départ, nous avons obtenu pour chaque CLSC et pour chaque maladie le nombre de cas recensés par année pour la période 1995-2000. Nous avons par la suite calculé l'incidence de chacune des cinq MADO pour chaque CLSC pour les années 1995 à 2000 à partir du nombre de cas répertoriés par année. Le dénominateur employé pour le calcul des taux est la population de chaque CLSC fournie par le recensement de 2001.

Par la suite, la moyenne de l'incidence de chaque MADO de 1995 à 2000 a été calculée, fournissant une incidence moyenne attribuable à l'année 1998 (mi-période). Finalement, la moyenne des cinq moyennes d'incidence pour la période 1995-2000 a été obtenue.

Ces données proviennent initialement de la banque de données MADO (voir section 3.1.7) et nous ont été fournies par le Secteur Vigie et Protection de la DSP de Montréal. Par ailleurs, les catégories élaborées par la méthode des quantiles et utilisées pour les analyses statistiques sont : 1-6, 7-14 et 15-100 nouveaux cas des cinq MADO entériques / 10^5 personnes / 1998.

vii) Tabagisme

Cette variable est l'estimé du pourcentage de tabagisme de la population de chaque CLSC, standardisé pour l'âge et le sexe. Cette variable a été construite à partir de données de tabagisme de régions socio-sanitaires (RSS) pour l'année 2003 qui ont été appliquées aux CLSC correspondants à chaque RSS puisque chaque RSS regroupe un certain nombre de CLSC. La correspondance entre CLSC et RSS est donnée en annexe 7.

En effet, les données de tabagisme pour 2003, soit le pourcentage de tabagisme pour différents groupes d'âge et par sexe fournies par le MSSS (voir section 3.1.7) étaient seulement disponibles pour 15 des 18 RSS pour l'année 2003 et non pas par région de CLSC. Ainsi, les pourcentages de tabagisme de 2003 des RSS ont été appliqués à la population des 12 à 65 ans des CLSC de 1996. Cela a permis d'*estimer* le nombre d'individus fumeurs par âge et selon le sexe pour chaque CLSC et cette valeur a servi à calculer le taux de tabagisme par CLSC. La méthode de calcul utilisée pour estimer le taux de tabagisme par CLSC, standardisé pour l'âge et le sexe selon les effectifs de population de 1996 de la province de Québec, est présentée en annexe 3. Le taux de tabagisme a par la suite été transposé en pourcentage de tabagisme. Les catégories élaborées par la méthode des quantiles et utilisées pour les analyses statistiques sont : moins de 26 % et plus de 26 % de tabagisme pour chaque CLSC.

viii) Statut socio-économique - indice de défavorisation

Il existe au Québec un indice qui reflète les dimensions matérielle et sociale de la défavorisation et qui fait usage des données du recensement et de certains outils

développés par Statistique Canada (133). La réalisation de cet indice se base a été fait à partir des travaux de Peter Townsend (134) pour qui la défavorisation peut être perçue comme un « état observable et démontrable de désavantage relatif face à la communauté locale ou à l'ensemble de la société à laquelle appartient l'individu, la famille ou le groupe » (traduction de l'anglais). Dans le présent ouvrage, l'indice de défavorisation est une variable ordinale qui reflète le statut socio-économique moyen des CLSC. Cet indice a été construit à partir des données fournies par l'Atlas de santé du MSSS (voir section 3.1.7). Selon le MSSS, la défavorisation peut être matérielle et/ou sociale. En effet, la défavorisation matérielle reflète la privation de biens et de commodités de la vie courante et la défavorisation sociale renvoie plutôt à la fragilité du réseau social, tant à l'échelle de la famille que de la communauté. Dans cet ouvrage, les données démographiques recueillies lors du recensement de Statistique Canada ont permis l'attribution de marqueurs de défavorisation aux différentes aires de diffusion du recensement pour lesquelles le territoire du Québec a été divisé. Ainsi, chaque région de CLSC regroupe un nombre plus ou moins élevé d'aires de diffusion du recensement, dépendamment de la taille de leur population.

L'indice de défavorisation utilisé dans cet ouvrage est donc la combinaison des quintiles matériels et des quintiles sociaux correspondant à la variation provinciale de la défavorisation pour chaque CLSC. L'indice de défavorisation comprend donc une composante matérielle et sociale. Le tableau présenté en annexe 4 est inspiré d'un tableau du MSSS et a servi à scinder l'indice de défavorisation en trois niveaux. Les catégories utilisées lors des analyses statistiques sont donc :

- 1: favorisé matériellement et socialement, incluant le niveau intermédiaire
- 2: défavorisé ou très défavorisé matériellement OU socialement
- 3: défavorisé ou très défavorisé matériellement ET socialement

3.4 Modélisation des données

3.4.1 Modèle général de régression

L'étude de la relation entre le taux d'incidence d'une maladie et la géographie d'un territoire ou une période de temps, implique un ajustement pour la confusion

potentiellement engendrée par l'utilisation de variables multiples. Ainsi, il est utile de modéliser les relations exposition-réponse de manière quantitative (60). Dans cet ouvrage, nous avons modélisé le taux d'incidence de la MC en lien avec certaines variables d'intérêt.

Tout d'abord, le nombre de nouveaux cas de MC qui surviennent dans la population de chaque CLSC pendant la période à l'étude est la variable prédite qui est un compte d'événements rares. L'objectif d'une analyse de régression dans un tel cas est donc de modéliser la variable dépendante comme étant un estimé des résultats prédits selon certaines variables explicatives (135). Par ailleurs, l'étude de l'apparition de cas d'une maladie dans une population implique que le compte des événements fournit toujours des valeurs entières positives. Dans le cas d'événements rares, la **distribution de Poisson** est la plus appropriée puisque la moyenne de cette distribution est supérieure à 0. Ainsi, le logarithme (log) de la variable prédite est relié à une fonction linéaire de variables explicatives (60).

Le modèle typique de régression de Poisson exprime donc le log de l'événement d'intérêt comme étant une fonction linéaire d'une séquence de variables prédictives (135). Il peut d'ailleurs être appliqué à des données dans lesquelles le temps et l'espace sont des facteurs centraux qui définissent les unités d'observation (60).

Le modèle de régression de Poisson est celui de l'équation 8 :

$$\text{Log}(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (129)$$

Équation 8 : Loi de Poisson

Où

y = variable dépendante prédite

β_0 = constante

$\beta_1 x_1$ = coefficient de la caractéristique indépendante x_1

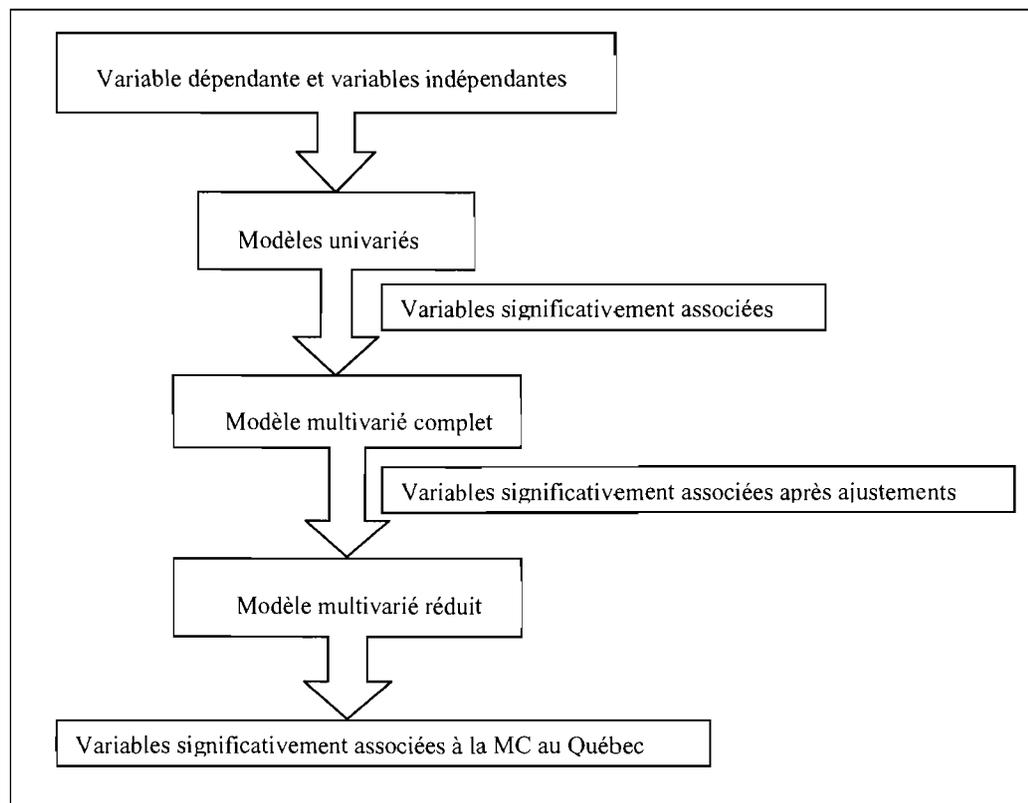
$\beta_n x_n$ = coefficient de la caractéristique indépendante x_n (n variables indépendantes)

Ainsi, les analyses statistiques permettent de déterminer un estimé des constantes β associées aux variables indépendantes. L'exponentielle de la valeur des estimés de β permet d'obtenir un estimé du ratio de risque (135).

3.4.2 Stratégie générale de modélisation

Chaque variable indépendante a été analysée dans un modèle de régression de Poisson univarié. Par la suite, les variables associées de manière statistiquement significatives à la variable dépendante dans les modèles univariés (valeur de $p < 0,05$, voir section 3.6) ont été introduites dans un modèle de régression multivarié qualifié de « complet ». Les variables associées de manière statistiquement significatives à la variable dépendante dans ce modèle multivarié ont servi à bâtir le modèle multivarié final qualifié de « réduit » (voir figure 5, p. 54). Chacune des étapes de modélisation est détaillée dans les sections suivantes.

Figure 5: Stratégie d'introduction des variables dans les modèles de régression



3.4.3 Modèles de régression : analyses univariées

La première étape de l'analyse est la sélection des variables qui pourront éventuellement faire partie d'un modèle à plusieurs variables. Ainsi, il est utile de bâtir autant de modèles qu'il y a de variables indépendantes afin de tester lesquelles seront associées de manière significatives à la variable dépendante (136). En effectuant ces analyses univariées, la puissance d'association entre les facteurs de risque et l'apparition de la maladie est mesurée par le IRR (121;121).

Le **IRR** se calcule en plaçant au numérateur le taux d'incidence des individus exposés et au dénominateur, le taux d'incidence des non-exposés. Le groupe des non-exposés est donc le groupe référence.

Taux d'incidence chez les sujets exposés

Taux d'incidence chez les sujets non-exposés

Équation 9 : Calcul du ratio de taux d'incidence

Par ailleurs, le taux d'incidence brut de la MC a été modélisé à l'univarié avec les variables « âge » et « sexe » afin d'identifier si ces variables avaient un effet sur la variable prédite. Par la suite, le taux d'incidence de la MC a été standardisé pour l'âge et le sexe de la province de Québec en 1996 (voir section 3.2.3) et est devenu la variable dépendante continue des modèles utilisés pour les analyses impliquant les autres variables indépendantes.

Ainsi, pour chaque territoire de CLSC, les variables ont été mises en relation une à une avec le taux d'incidence en les modélisant selon la régression de Poisson tel que:

$$\text{Log}(y) = \beta_n x_n$$

Équation 10: Relation entre la variable indépendante et dépendante – modèle univarié

Variable dépendante (y) : cas de MC

Variables indépendantes (x_n) : variables populationnelles (10)

3.4.4 Modèles de régression : analyses multivariées

Par la suite, le modèle de régression de Poisson a été utilisé à nouveau afin de tester quelles variables sont associées à la MC lorsqu'un ajustement mutuel entre ces variables est effectué. Ainsi, les variables ayant été préalablement associées de manière significative à la MC lors des analyses univariées ont été incluses dans un premier modèle multivarié qualifié de « complet ». La structure générale d'un modèle multivarié complet à « n » variables indépendantes est celui de l'équation 11:

$$y = x_1 + x_2 + x_3 + x_4 \dots x_n$$

Équation 11 : Relation entre les variables dépendantes et indépendantes - modèle multivarié

Il est à noter que les variables « âge » et « sexe » ont été exclues du modèle multivarié puisque les taux d'incidence employés lors des tests statistiques ont été standardisés pour ces deux facteurs.

Par la suite, les variables dont les tests statistiques se sont révélés significatifs (valeur de $p < 0,05$, voir section 3.6) dans le modèle complet ont été incluses dans la construction d'un modèle multivarié qualifié de « réduit » afin de déterminer lesquelles pouvaient être associées de façon statistiquement significative avec la MC.

3.4.5 Colinéarité entre les variables

Afin de vérifier si une ou plusieurs variables résultent de la combinaison linéaire d'autres variables indépendantes, un test de colinéarité entre les variables des modèles multivariés a été effectué. Trois modèles de régression ont été testés :

- Modèle complet incluant toutes les variables en continu du modèle multivarié
- Modèle réduit incluant les variables en continu du modèle multivarié
- Modèle réduit incluant les variables catégorisées du modèle multivarié

Le facteur d'inflation de la variance (VIF) et la valeur de tolérance (TOL) ont été utilisés lors de l'estimation de la colinéarité. Le VIF mesure l'inflation de la variance du

paramètre étudié. S'il y a inflation, celle-ci est due à la présence de colinéarité parmi les variables indépendantes du modèle de régression (137). Un VIF ≥ 10 est interprété comme étant synonyme de colinéarité. Par ailleurs, la TOL ($1/\text{VIF}$) est employée afin de déterminer le degré de colinéarité si tel est le cas. Une valeur de TOL $< 0,1$ est comparable à un VIF de 10 (138).

3.4.6 Corrélation entre les variables

Un test de corrélation permet de vérifier si les variables des modèles multivariés sont interreliées, c'est-à-dire évaluer à quel niveau une variable peut en prédire une autre. La corrélation est mesurée par le coefficient de corrélation (R) dont la valeur peut se retrouver entre -1 et $+1$. Ainsi, une valeur de R près de 1 reflète une corrélation très positive entre deux variables et une valeur de R près de -1 reflète une corrélation très négative. Dans le cas où $R = 0$, cela signifie qu'il n'y a pas d'association linéaire entre deux variables (139).

La corrélation a été mesurée par le « coefficient de corrélation de Pearson » auquel une valeur statistique de « p » est attribuée. L'interprétation d'un estimé de corrélation peut se faire à deux niveaux : selon que le coefficient de corrélation est statistiquement significatif et selon que la valeur de ce coefficient se trouve au-dessus ou au-dessous d'un seuil établi selon les buts de l'étude. En effet, il peut être souhaitable, dépendamment des hypothèses de recherche, d'établir quelle valeur du coefficient de corrélation représente le seuil en-dessous duquel un coefficient de corrélation ne sera pas interprété comme reflétant une forte corrélation.

3.4.7 Modélisation des mesures répétées et des données corrélées dans l'espace

Les mesures écologiques utilisées pour créer les variables à l'étude reflètent la réalité d'individus qui sont regroupés en territoires de CLSC. Il est donc théoriquement possible que des individus partagent des caractéristiques communes entre eux de par le fait qu'ils habitent un même territoire. Les données populationnelles sont ainsi soupçonnées d'être corrélées entre elles. Par ailleurs, les variables « âge » et « sexe » sont différentes des autres variables en ce sens qu'elles représentent des données recensées au niveau individuel.

En effet, la variable « âge » est constituée de cinq groupes et la variable « sexe » en compte deux. Ainsi, chaque CLSC a été mesuré afin de connaître combien de cas se retrouvent dans chaque groupe d'âge et combien de cas des deux sexes sont identifiés. Cela crée une répétition de mesures pour chaque CLSC et des données prises dans un même territoire auront tendance à être plus similaires que des données collectées entre différentes régions. Enfin, les mesures répétées sont de type longitudinal (140).

Il est donc important que les modèles statistiques choisis tiennent compte du phénomène de répétition des mesures et d'une possible corrélation entre les observations recensées dans un même CLSC (140). C'est pourquoi la procédure GENMOD du logiciel SAS 9.1 a été employée lors des analyses statistiques de régression puisqu'elle permet d'analyser des données répétées pouvant être corrélées en les modélisant selon des équations généralisées d'estimés (*generalized estimating equations* (GEE)). La méthodologie de GEE a été introduite par Liang et Zeger en 1986 et permet l'analyse de données corrélées (données longitudinales ou mesures faites sur des individus provenant d'un même territoire) qui seraient sinon modélisées selon un modèle linéaire général (les détails de la procédure sont présentés en annexe 6) (141).

3.4.8 Calcul du pseudo-R² du modèle de régression de Poisson

L'utilisation du coefficient de détermination (R²) est largement utilisé dans les analyses de régression classiques. Il se définit comme la proportion de variance « expliquée » par un modèle de régression. Cette propriété rend le R² utile comme mesure du succès de prédiction d'une variable dépendante par des variables indépendantes. Toutefois, dans le cas du modèle de régression de Poisson, le concept de variance résiduelle est plus difficilement mesurable. C'est pourquoi il a été démontré que la définition du R² peut être élargie aux modèles généraux afin de calculer un pseudo-R² (142). La formule mathématique suivante permet d'obtenir le pseudo- R² :

$$\text{Pseudo- } R^2 = 1 - \text{EXP} [(1/n) * (\text{dev}\beta - \text{dev}0)] \quad (142)$$

Équation 12 : Calcul du pseudo- R² pour un modèle régression de Poisson

Où

EXP : exponentielle

n : taille de l'échantillon

Dev β : déviance du modèle de régression

Dev0 : déviance du modèle vide ou nul

3.5 Tests statistiques des analyses

Une analyse univariée des variables ainsi que des tests d'interaction ont été effectués avec une analyse Chi-Carré pour la comparaison des variables discrètes. Les tests de colinéarité ont été effectués avec une analyse de T de Student. Les tests de corrélation ont été mesurés à l'aide du coefficient de corrélation de Pearson.

3.6 Seuil alpha

Un seuil alpha significatif de 0,05 a été utilisé pour tous les tests statistiques de ce travail.

3.7 Logiciels utilisés

Les données correspondant aux variables principales et aux sous-variables ont été organisées et entreposées à l'aide du logiciel Excel de Microsoft Office 2003. Les manipulations des données (proportions, moyennes, calculs d'incidence) ainsi que les analyses descriptives ont été effectuées à l'aide de la procédure GENMOD du logiciel SAS version 9.1

PARTIE 4 : Résultats - présentation et analyse

4.1 Épidémiologie de la maladie de Crohn au Québec – analyses descriptives

4.1.1 Nombre de cas de maladie de Crohn identifiés

Un total de **10 065** cas a été identifié pour la période 1995-2000. Lors d'une étude précédente dont la période à l'étude était 1993-2002, **21 172** cas avaient été identifiés (13).

4.1.2 Validation et rendement de la définition opérationnelle de cas

Un total de 69 cas cliniques a été recruté et **55** d'entre eux ont pu être inclus dans la validation puisqu'ils avaient été diagnostiqués entre 1993 et 2002. De ces 55 individus, 32 avaient un diagnostic de MC et 23 avaient un diagnostic de CU. Nous avons par ailleurs choisi de valider les **21 172** cas identifiés lors d'une étude précédente pour la période 1993-2002 (13). Ainsi, le fait d'utiliser 55 cas cliniques a permis une validation de **0.26%** des données totales.

Tableau V : Tableau 2X2 - résultats de la validation de la définition de cas

		DIAGNOSTIC CLINIQUE		
		MC (+)	Autres (-)	
DEFINITION DE CAS MED-ECHO (RAMQ)	MC (+)	31 a	4 b	35
	Autres (-)	1 c	19 d	20
		32	23	55

À l'aide des chiffres contenus dans les cellules du tableau V en page 60, différents calculs ont pu être faits. Ainsi, la sensibilité de la définition de cas a été évaluée à **97%**, et la spécificité, à **83%**. En ce qui a trait au rendement, la valeur prédictive positive a été estimée à **89%** et la valeur négative, **95%**. Les calculs sont présentés ci-bas :

Sensibilité: $a / (a+c)$

$$31 / (31+1) = (31/32) * 100 = 97\%$$

Spécificité: $d / (b+d)$

$$19 / (4+19) = (19/23) * 100 = 83\%$$

VP+: $a / (a+b)$

$$31 / (31+4) = (31/35) * 100 = 89\%$$

VP-: $d / (c+d)$

$$19 / (1+19) = (19/20) * 100 = 95\%$$

4.1.3 Incidence moyenne de la maladie de Crohn au Québec

Pour la période 1995-2000, l'incidence moyenne de la MC standardisée pour la population du Québec de 1996 est de **22,4** nouveaux cas / 10^5 personnes / année. Les taux d'incidence varient selon les CLSC de 0 à 44,1 nouveaux cas / 10^5 personnes / année. Par ailleurs, la population québécoise estimée par le recensement canadien de 1996 était de 7 138 800 personnes (125). Cela signifie qu'environ 1600 individus auraient développé la MC pour l'année 1996 au Québec.

De plus, l'incidence moyenne de la MC au Québec standardisée selon la population combinée de cinq provinces canadiennes (excluant le Québec) de 1996 est de 22,2 nouveaux cas / 10^5 personnes / année. La province canadienne qui a le plus haut taux d'incidence de la MC est le Québec (voir tableau VI, p. 62).

Tableau VI: Incidence de la MC dans six provinces canadiennes⁹

PROVINCE	TAUX D'INCIDENCE ^Δ (/10 ⁵ PERS/AN)
Colombie-Britannique*(CB)	8,8
Alberta*(AB)	16,5
Saskatchewan*(SK)	13,5
Manitoba*(MB)	15,4
Québec** (QC)	22,2
Nouvelle-Écosse*(NÉ)	20,2

* période de l'étude : 1998-2000

** période de l'étude : 1995-2000

^Δ Incidence standardisée pour l'âge et le sexe des populations combinées de CB, AB, SK, MB, NÉ

4.1.4 Visualisation des données d'incidence

Il existe une variation du taux d'incidence de la MC entre les 166 CLSC du Québec. Les figures 6 et 7 en pages 64 et 65 ont permis de visualiser cette variation entre les différentes régions de la province. Par ailleurs, les onze CLSC ayant les taux d'incidence standardisés les **plus élevés** (32,3 – 44,1 nouveaux cas / 10⁵ personnes / année) sont énumérés au tableau VII en page 63:

⁹ Tableau de résultats adapté de : Bernstein, CN. *The epidemiology of inflammatory bowel disease in Canada: a population-based study*. Am J Gastro. 2006(101)1559-1568 (4)

Tableau VII : CLSC ayant les taux d'incidence standardisés les plus élevés au Québec

CLSC (NOM)	CLSC (CODE 2005)	INCIDENCE STD.* (/10 ⁵ PERS/AN)	URBAIN (U) RURAL (RU)
Pabok	11021	44,1	Ru
Côte-St-Luc	6081	41,4	U
Cowansville Farnham- Bedford	16101	41,0	U
Lac-St-Jean est	2061	40,7	U
Des Chenaux	4022	37,7	Ru
Bonaventure	11011	35,3	Ru
Desjardins	12021	34,9	U
Joliette	14013	34,8	U
Haut Saint-François	5031	33,6	Ru
Cap-de-la-Madeleine	4052	32,9	U
Montréal Centre-ville	6094	32,3	U

* Incidence standardisée selon la population du Québec de 1996

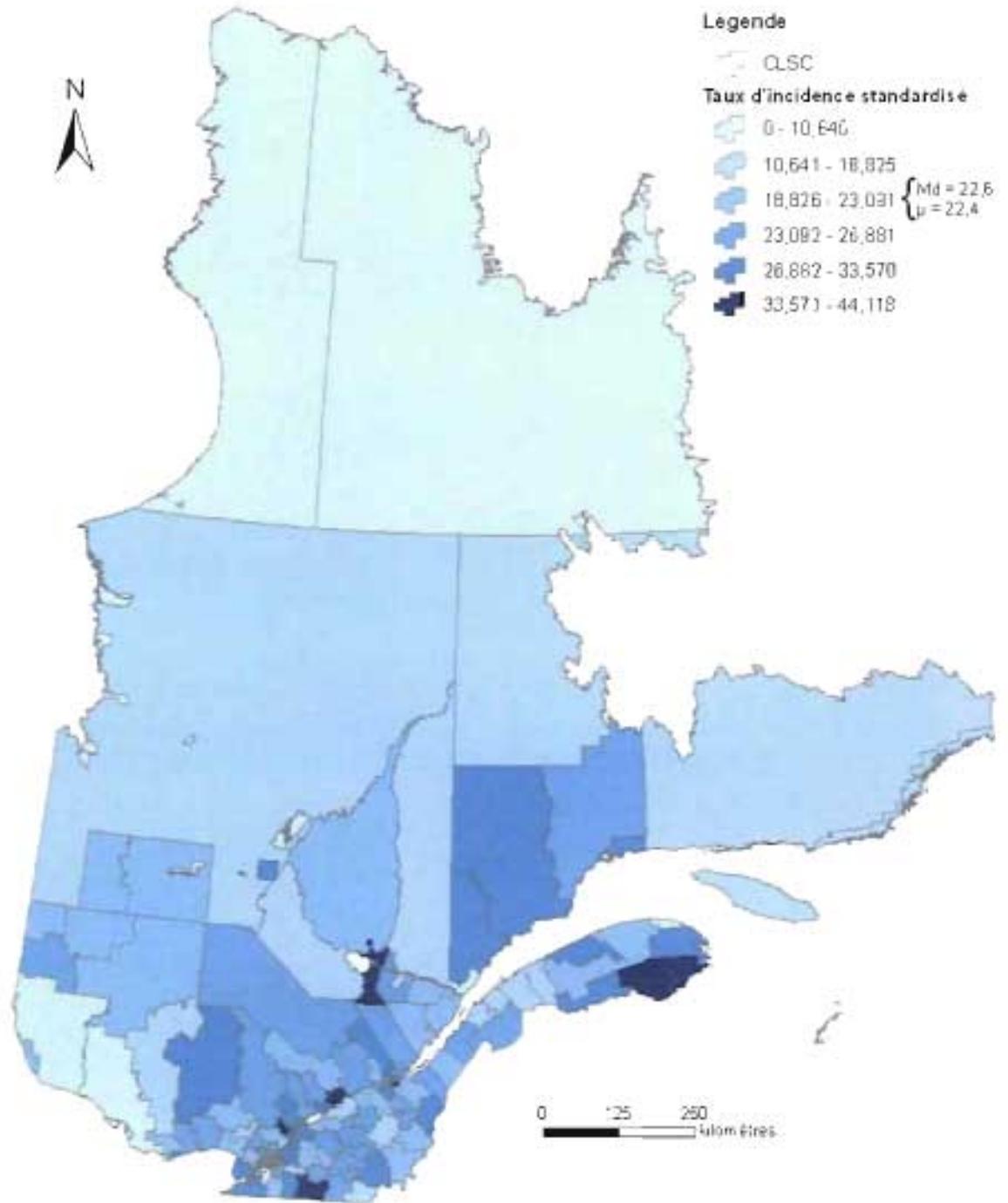
De plus, les huit CLSC ayant les taux d'incidence standardisés les **plus faibles** (0 – 9,8 nouveaux cas / 10⁵ personnes / année) sont énumérés au tableau VIII en page 63 :

Tableau VIII : CLSC ayant les taux d'incidence standardisés les plus faibles au Québec

CLSC (NOM)	CLSC (CODE 2005)	INCIDENCE STD.* (/ 10 ⁵ PERS/AN)	URBAIN (U) RURAL (RU)
Territoire Naskapi	9081	0	Ru
Baie d'Hudson	17101	0	Ru
Ungava	17102	0	Ru
Territoire Cri	18101	0	Ru
Pontiac	7021	5,8	Ru
Châteauguay-Mercier	16041	6,6	U
Grande-Vallée	11032	9,5	Ru
Ville-Marie	8021	9,8	U

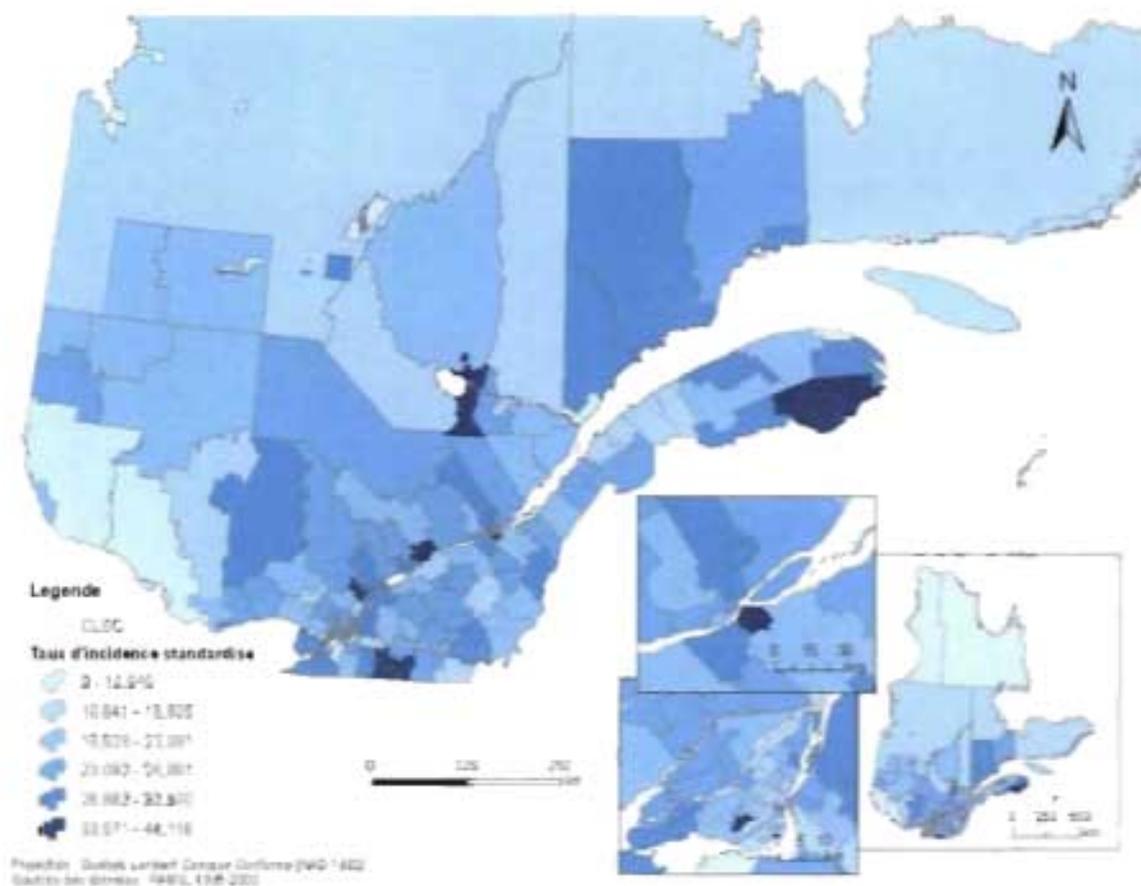
* Incidence standardisée selon la population du Québec de 1996

Figure 6 : Répartition géographique des taux d'incidence de la maladie de Crohn par régions de CLSC, province de Québec (1995-2000)



Projection : Québec Lambert Conique Conforme (NAD 1983)
Sources des données : RAMQ, 1995-2000

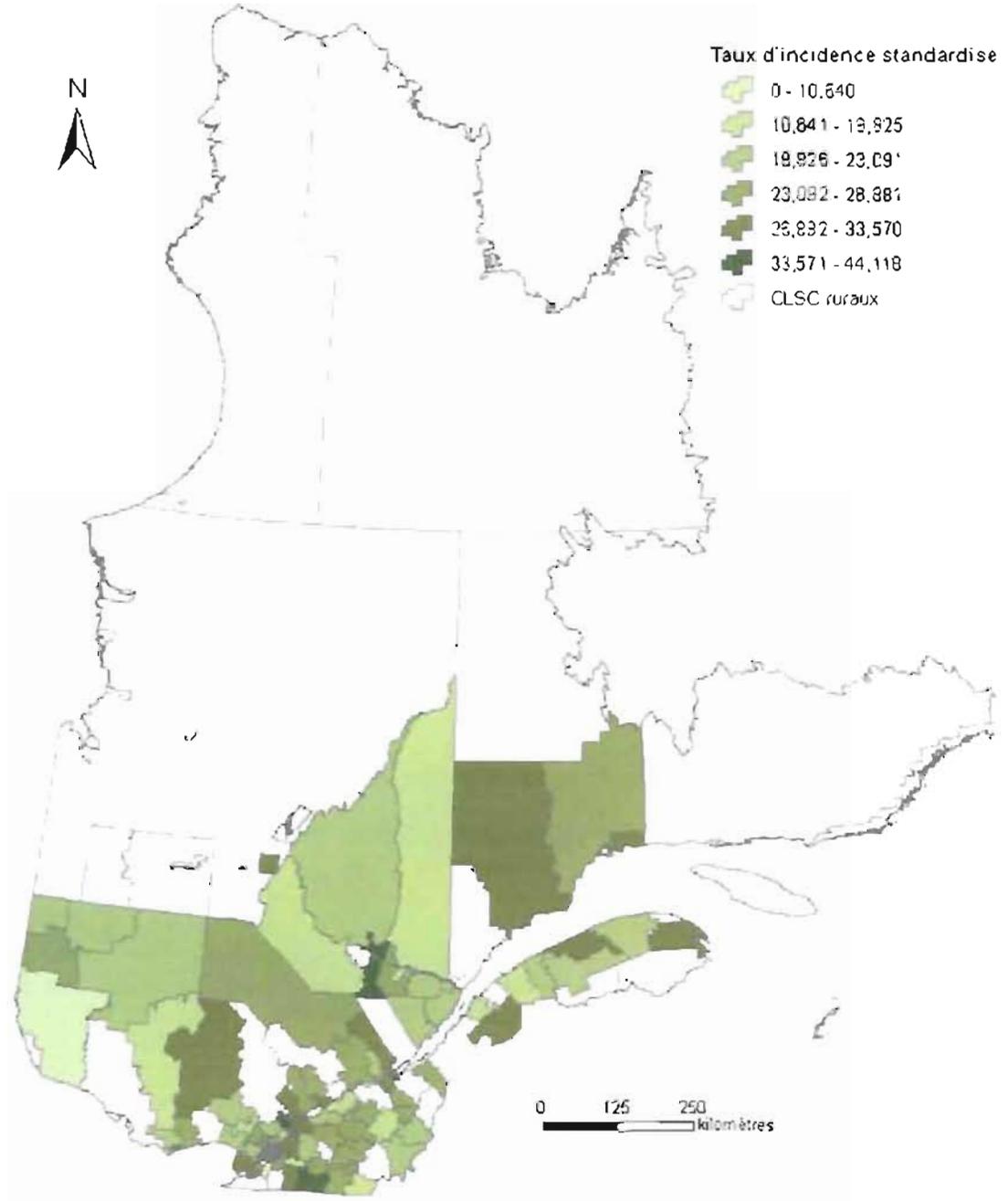
Figure 7 : Répartition géographique des taux d'incidence de la maladie de Crohn par régions de CLSC, excluant les CLSC du nord. zoom sur l'île de Montréal, province de Québec (1995-2000)



Tous les CLSC ont été identifiés comme étant urbain (U) ou rural (Ru). Dans le tableau VII en page 63, il est possible d'observer que parmi les onze CLSC où l'on retrouve un plus haut taux d'incidence de la MC, sept d'entre eux ont été identifiés comme étant urbains. De plus, dans le tableau VIII en page 63, on observe que parmi les huit CLSC où l'on retrouve les taux d'incidence de la MC les plus bas, six d'entre eux ont été identifiés comme étant ruraux.

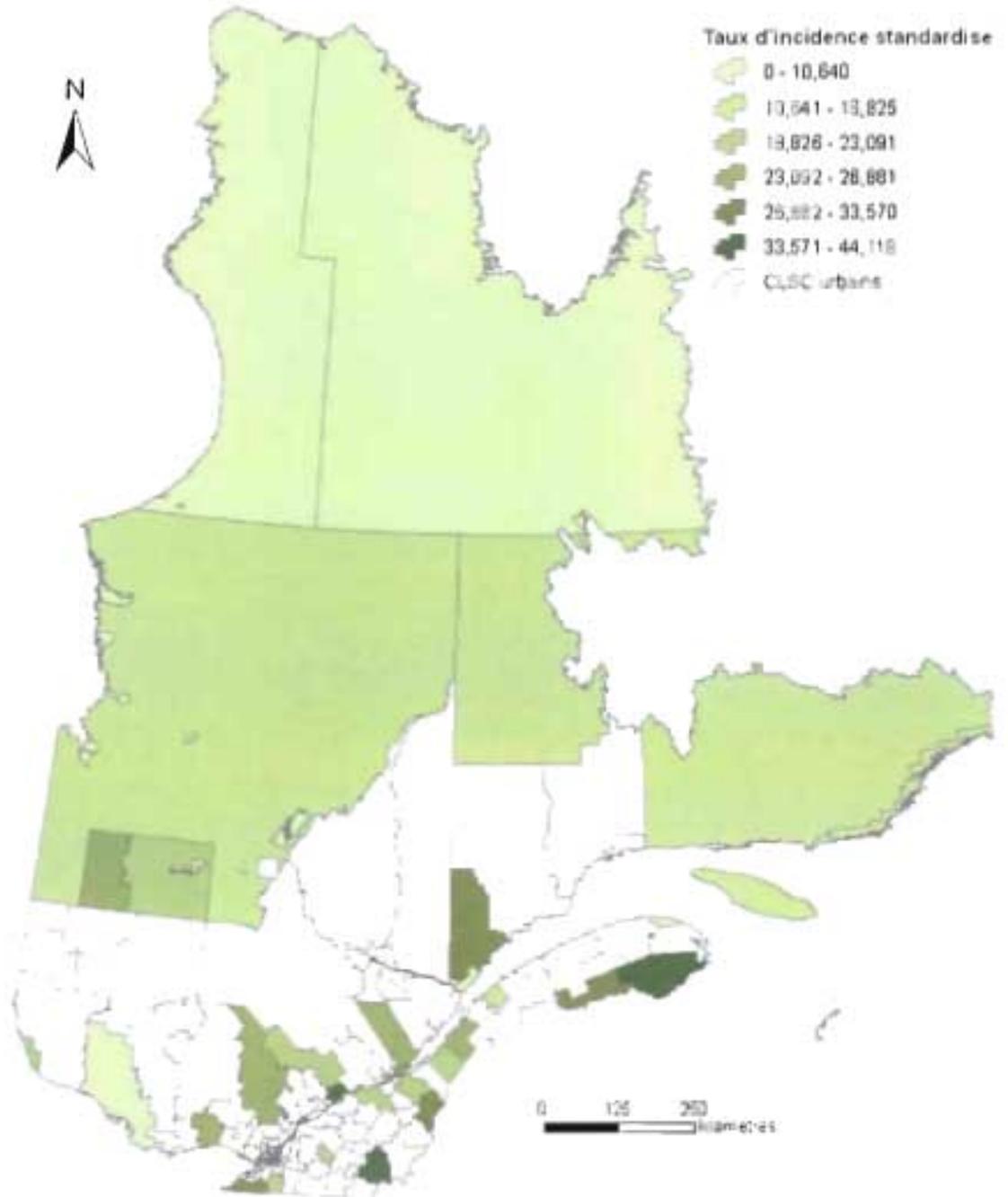
Afin de faire un contraste entre les taux d'incidence de la MC attribués aux régions de CLSC selon qu'ils sont classés urbains ou ruraux, les figures 8 et 9 en pages 67 et 68 ont été générées. Des 166 CLSC totaux, 34 sont ruraux et 132 sont urbains.

Figure 8 : Répartition géographique des taux d'incidence de la maladie de Crohn par régions de CLSC urbains, province de Québec (1995-2000)



Projection : Québec Lambert Conique Conforme (NAD 1983)
Sources des données : RAMQ 1995-2000

Figure 9 : Répartition géographique des taux d'incidence de la maladie de Crohn par régions de CLSC ruraux, province de Québec (1995-2000)



Projecteur : Québec Lambert Coque Contour (NAD 1983)
Sources des données : RAMQ 1995-2000

4.2 Analyses associatives – les modèles de régression

4.2.1 Analyses univariées

En premier lieu, l'analyse univariée de chacune des variables indépendantes a permis d'identifier lesquelles d'entre elles sont associées de façon statistiquement significative avec la MC. Les résultats des estimés bruts d'IRR sont présentés dans les tableaux IX et X en pages 69 et 70. Brièvement, on observe que les individus faisant partie du groupe d'âge des 20-29 ans ont un taux d'incidence de la MC 2,64 fois plus élevé que le groupe de référence des 0-19 ans. Chez les femmes, le taux d'incidence de la MC est presque deux fois plus élevé que chez les hommes (voir tableau IX, p. 69). Il est à noter qu'il n'y a pas d'interaction entre les variables « âge » et « sexe » puisqu'une valeur non significative de $p = 0,1488$ a été obtenue lors d'une analyse d'interaction.

Tableau IX : IRR bruts des modèles univariés incluant les variables « âge » et « sexe »

Variables prédictives	Analyses			
	Inc. Moy.*	IRR bruts ^Δ	IC 95%	
Âge				
0-19 ans (Réf.)	13,9		Réf.**	
20-29 ans	35,5	2,64	2,34	2,93
30-39 ans	24,3	1,62	1,44	1,82
40-54 ans	21,7	1,52	1,39	1,66
> 55 ans	20,6	1,5	1,36	1,64
Sexe				
Masculin (Réf.)	18,5		Réf.	
Féminin	27,9	1,95	1,68	2,27

* Incidence moyenne de la MC, par 10^5 personnes / année

** Intervalle de référence dans le calcul du IRR

^Δ IRR calculés en utilisant comme variable dépendante l'incidence brute de la MC

Par ailleurs, les variables populationnelles qui sont associées à la MC dans les modèles de régression univariés sont les suivantes : habiter en région urbaine, une forte proportion d'immigrants, une forte proportion d'individus de descendance juive, l'utilisation de contraceptifs oraux, une forte incidence de cinq MADO entériques, le statut autochtone et le statut socio-économique élevé.

Tableau X : IRR bruts et ajustés de modèles univariés et multivariés -- association de variables à la maladie de Crohn

Variables prédictives	ANALYSES					
	CLSC n =	Inc. Moy. *	IRR bruts ^Δ	IRR ajustés [¥]	IC 95%	
Urbanité - résidence						
Rural (Réf.)	34	20,2		Réf.		
Urbain	132	22,9	1,14	x	x	x
Immigration (%)						
0-10 % (Réf.)	127	23		Réf.		
11-62 %	39	20,1	0,87	0,78	0,70	0,87
Descendance juive (%)						
0-1% (Réf.)	152	22,3		Réf.		
2-19%	11	21,2	0,95	1,01	0,86	1,19x
20-59%	3	28,7	1,28	1,65	1,27	2,14
Contraceptifs oraux (ord. 28 jrs/an)						
0-3 (Réf.)	28	21,6		Réf.		
> 4	101	24	1,11	1,05	0,95	1,17
Incidence 5 MADO entériques (10 ⁵ pers/an)						
1 - 6 (Réf.)	15	19		Réf.		
7 - 14	60	22,6	1,19	1,23	1,04	1,46
15-100	91	23	1,21	1,20	1,01	1,42
Tabagisme (%)						
<26% (Réf.)	85	22,9		Réf.		
>26%	81	21,8	0,96	x	x	x
Autochtones (%)						
> 3.6% (Réf.)	21	15,4		Réf.		
0.6-3.5%	43	22,9	1,49	1,02	0,85	1,20
0-0.5%	102	23,6	1,53	0,95	0,79	1,15
Statut socio-économique (SES)						
1. favorisé matériellement et socialement et niveau intermédiaire	56	24,4	1,21	1,03	0,69	1,55
2. défavorisé ou très défavorisé matériellement ou socialement	103	21,4	1,07	0,96	0,79	1,17
3. défavorisé ou très défavorisé matériellement et socialement (Réf.)	7	20,2		Réf.		

* Incidence de la MC standardisée selon l'âge et le sexe de la population du QC 1996, par 10⁵ personnes / an

^Δ IRR de modèles univariés dont la variable dépendante est l'incidence de la MC standardisée selon l'âge et le sexe de la population du QC de 1996

[¥] IRR du modèle multivarié complet dont la variable dépendante est l'incidence de la MC standardisée selon l'âge et le sexe de la population du QC de 1996, ajusté pour les variables: immigration, juifs, incidence 5 MADO, autochtones et SES

Les données des tableaux IX et X en pages 69 et 70 permettent de comparer les estimés bruts d'IRR de chaque variable avec la moyenne de l'incidence standardisée pour la période de l'étude qui est de 22,4 nouveaux cas / 10⁵ personnes / année. Ainsi, l'incidence maximale de la MC est retrouvée chez les individus de la tranche d'âge 20-

29 ans et elle est de **35,5** nouveaux cas / 10^5 personnes / année. De plus, l'incidence de la MC est de 28,7 nouveaux cas / 10^5 personnes / année chez les populations comptant de 20 à 59% d'individus de descendance juive et elle est de 27,9 nouveaux cas / 10^5 personnes / année chez les femmes. À l'opposé, on observe une incidence de la MC plus faible que la moyenne lorsque la proportion d'immigrants est forte (de 11 à 62%), c'est-à-dire 20,1 nouveaux cas / 10^5 personnes / année ainsi qu'aux endroits où la population compte 3,6% et plus d'individus autochtones, c'est-à-dire **15,4** nouveaux cas / 10^5 personnes / année.

4.2.2 Analyses multivariées

i) modèle de régression multivarié complet

Après avoir identifié quelles variables étaient associées à la MC dans les modèles univariés, un premier modèle de régression multivarié a été bâti en n'incluant que ces variables. Toutefois, la variable « urbanité » a volontairement été exclue du modèle de régression multivarié complet bien qu'elle ait été associée de manière significative à la variable dépendante à l'univarié. Cette décision repose sur les résultats de tests statistiques qui ont démontré la présence d'interaction entre la variable « urbanité » et les variables « immigration », « descendance juive », « incidence des cinq MADO entériques » et « statut autochtone » (valeur de $p < 0,05$ pour chaque estimé). De plus, la variable « urbanité » est corrélée positivement avec la variable « immigration » et négativement avec « statut autochtone » et « indice de défavorisation socio-économique » (valeur de $p < 0,05$ pour chaque estimé). Comme cette variable ajoutait trop de distorsion à nos estimés globaux, nous avons choisi de ne pas la retenir dans l'analyse multivariée. Le modèle de régression multivarié complet est le suivant :

$$\text{Log}(y) = \beta_0 + \beta_1 \text{imm} + \beta_2 \text{juifs} + \beta_3 \text{CO} + \beta_4 \text{inc5mado} + \beta_5 \text{autocht} + \beta_6 \text{indiceSES}$$

Équation 13 : Modèle de régression multivarié complet

Où

- imm* : proportion d'immigrants dans la population
juifs : proportion d'individus de descendance juive dans la population
CO : pourcentage de femmes qui font usage de contraceptifs oraux

inc5mado : incidence des cinq MADO entériques
autocht : proportion de la population ayant le statut autochtone
indiceSES : indice de défavorisation socio-économique

Par ailleurs, des associations significatives ont été maintenues entre la MC et certaines variables catégorisées après ajustement entre les variables du modèle de régression multivarié complet, soit :

- Forte proportion d'immigrants dans la population (IRR = 0,78, IC 95% : 0,70 – 0,87)
- Forte proportion d'individus de descendance juive dans la population (IRR = 1,65, IC 95% : 1,27 – 2,14)
- Forte incidence des cinq MADO entériques (IRR = 1,20, IC 95% : 1,01 – 1,42)
- Incidence intermédiaire des cinq MADO entériques (IRR = 1,23, IC 95% : 1,04 – 1,46)

Les résultats complets sont indiqués au tableau X en page 70 dans la colonne « IRR ajustés ».

ii) modèle de régression multivarié réduit

Finalement, un modèle de régression multivarié réduit a été bâti en n'incluant que les variables indépendantes associées à la MC dans le modèle complet. Le modèle réduit est le suivant :

$$\text{Log}(y) = \beta_0 + \beta_1 \text{imm} + \beta_2 \text{juifs} + \beta_3 \text{inc5mado}$$

Équation 14 : Modèle de régression multivarié réduit

Les estimés ajustés des IRR de ce modèle sont présentés au tableau XI en page 73.

Tableau XI : IRR ajustés du modèle multivarié réduit qui inclut les variables significatives du modèle complet

Valeurs prédictives	Analyses		
	IRR ajustés*	IC 95%	
Immigration (11-62%)	0,81	0,73	0,89
Juifs (20-59%)	1,75	1,38	2,23
Incidence 5 MADO entériques (7-14)	1,27	1,11	1,45
Incidence 5 MADO entériques (15-100)	1,26	1,11	1,44

* Incidence standardisée selon l'âge et le sexe de la population du QC 1996; valeurs de référence pour chaque variable tel qu'indiqué au tableau X en p. 70

Il est à noter qu'il n'y a pas d'interaction entre les variables « immigration », « juifs » et « incidence de cinq MADO entériques » puisqu'une valeur de $p > 0,05$ a été obtenue lors du test statistique. De plus, un test de colinéarité a été effectué entre ces variables et aucune colinéarité n'a été détectée. En effet, pour toutes les variables, aucune valeur de VIF n'a dépassé 10 (la valeur la plus élevée de VIF était 1,58). Du coup, la valeur de tolérance de toutes les variables était plus élevée que 0,1 (la valeur la plus faible était 0,63).

4.2.3 Pseudo- R^2 du modèle final de régression de Poisson

Pour ce modèle de régression de Poisson, la valeur calculée du pseudo- R^2 est de **0,21**. Le calcul du pseudo- R^2 est le suivant :

$$\text{Pseudo-} R^2 = 1 - \text{EXP} [(1/n) * (\text{dev}\beta - \text{dev}0)] \quad (142)$$

$$\text{Pseudo-} R^2 = 1 - \text{EXP} [(1/166) * (431,56 - 470,98)]$$

$$\text{Pseudo-} R^2 = 0,21$$

PARTIE 5 : Discussion

La MC est une maladie inflammatoire des intestins dont l'étiologie est peu connue. Certains facteurs de risque ont été associés à cette maladie qui affecte principalement les individus des pays développés. Au Canada, les caractéristiques épidémiologiques de la MC ont été étudiées dans cinq provinces, soit la Colombie-Britannique, l'Alberta, la Saskatchewan, le Manitoba et la Nouvelle-Écosse. Dans cet ouvrage, nous avons effectué une caractérisation de la MC pour la province de Québec, en s'intéressant à la variation du taux d'incidence entre les territoires de CLSC ainsi qu'aux possibles associations existant entre la MC et certaines variables populationnelles.

Validation de la définition opérationnelle de cas de MC

La sensibilité et la spécificité constituent les deux mesures de la validité d'un test de dépistage (121). Dans cet ouvrage, nous avons validé non pas un test de dépistage, mais plutôt les cas identifiés par la définition opérationnelle de cas adaptée aux banques de données médicales du Québec. L'exercice de validation nous a permis de déterminer que la sensibilité de la définition de cas utilisée est très élevée. Cela suggère que parmi les cas identifiés selon les critères cliniques, seul un faible pourcentage d'entre eux n'a pas été décelé par la définition opérationnelle de cas. De plus, la spécificité de la définition opérationnelle de cas est élevée, tout en étant légèrement inférieure à la sensibilité. Ainsi, le taux de faux-positifs générés par la définition de cas est faible. Par ailleurs, la sensibilité mesurée dans le présent ouvrage est supérieure à celle obtenue par Bernstein et ses collaborateurs au Manitoba. Par contre, notre spécificité est légèrement inférieure à la leur. (12).

Le rendement de la définition opérationnelle de cas a aussi été estimé par la valeur prédictive. Cette valeur mesure la présence réelle ou non de la maladie chez les individus par rapport aux résultats de la définition de cas. La VP+ estime donc que parmi les individus identifiés comme des cas par la définition opérationnelle, presque tous sont réellement atteints par la MC. Par ailleurs, la VP- mesurée estime que presque tous les individus réellement non affectés par la MC (mais pouvant être atteints de CU dans le cas de cet exercice de validation) n'ont pas été captés par la définition opérationnelle (121). Ainsi, les résultats de rendement de cette étude démontrent que

pratiquement tous les individus décelés par la définition opérationnelle de cas sont des cas réels de MC, et que ceux qui n'ont pas été identifiés comme étant des cas ne le sont pas non plus dans la réalité. Ces valeurs s'ajoutent à la sensibilité et à la spécificité et confirment la validité de la définition opérationnelle de cas, et par le fait même, la précision de l'incidence mesurée au Québec.

Incidence de la MC au Québec

Nous avons déterminé le taux d'incidence moyen de la MC au Québec standardisé pour l'âge et le sexe selon la population combinée de cinq provinces (excluant le Québec) pour la période à l'étude. Nous avons utilisé cette population comme référence lors de la standardisation afin d'être en mesure de comparer de la manière la plus juste possible notre résultat d'incidence avec ceux obtenus pour la Colombie-Britannique, l'Alberta, la Saskatchewan, le Manitoba et la Nouvelle-Écosse. En effet, dans une étude précédente, l'âge et le sexe de la population combinée de ces cinq provinces avaient servi à standardiser leur taux d'incidence respectif (4).

En comparant le taux d'incidence de la MC du Québec à celui des autres provinces, nous observons que le taux d'incidence de la MC du Québec serait non seulement **le plus élevé au Canada**, mais à notre connaissance, il serait l'un des plus élevés **au monde**. De plus, le nombre d'individus qui auraient développé la MC pour l'année 1996 au Québec s'avère être considérablement élevé.

Variation géographique de l'incidence de la MC au Canada

Nous croyons que le taux d'incidence moyen de la MC au Québec obtenu dans cet ouvrage est vraisemblable puisqu'il s'insère dans la séquence croissante d'ouest en est des taux d'incidence calculés pour les autres provinces canadiennes. Selon ces résultats, la Colombie-Britannique a le taux d'incidence le plus faible (4) et le Québec, le plus élevé. Il existe donc une réelle variation des taux d'incidence de la MC entre les différentes provinces du Canada. À noter que nous nous exprimons en terme de « variation » et non de « gradient » puisqu'aucune analyse spatiale rigoureuse n'a été faite sur ces résultats. D'ailleurs, il est possible que cet effet de « gradient géographique » soulevé par d'autres auteurs serait atténué, voire même éliminé, si les données d'incidence étaient corrigées pour certaines variables potentiellement réparties

de façon inégale entre les provinces et associées à la MC. Par exemple, il existe peut-être une hétérogénéité dans la dispersion des individus de descendance juive sur le territoire du Canada. Il est aussi possible que la proportion d'immigrants et d'individus autochtones varie entre les populations. A ce sujet, Bernstein et ses collaborateurs mentionnent justement que « *les résultats d'incidence obtenus pour la Colombie-Britannique ont parfois été exclus des calculs d'incidence et de prévalence puisqu'il est connu que la composition ethnique de la population de cette province est différente des autres provinces* ¹⁰» (4).

Par ailleurs, on peut supposer que le nombre approximatif de Canadiens affectés par une MII estimé par Bernstein et ses collaborateurs serait potentiellement une sous-estimation de la réalité puisque le Québec ne figurait pas parmi les provinces incluses dans leurs analyses. Le taux d'incidence de la MC au Québec étant le plus élevé au Canada, le fait d'inclure cette valeur ferait probablement gonfler l'estimé du nombre d'individus affectés qui était initialement estimé à 170 000 (4).

Association entre la MC et l'âge des cas

Les informations collectées dans la banque de données de la RAMQ nous ont permis d'étudier deux variables construites à partir de données individuelles, c'est-à-dire l'influence de l'âge et du sexe sur le développement de la maladie. Nous avons déterminé que le taux d'incidence de la MC le plus élevé au Québec est retrouvé chez les individus âgés entre 20-29 ans. Ce taux est plus élevé que ce qui a été observé au Manitoba pour le même groupe d'âge (85). Une association a été observée entre ce groupe d'âge et la MC et elle est la plus forte obtenue dans cet ouvrage. De tels résultats concordent par ailleurs avec ce qui est retrouvé dans la littérature (12;42).

On peut soupçonner que les effets associés à l'âge pourraient en réalité représenter une exposition accrue à un autre facteur de risque (exogène ou endogène) qui affecterait différemment les individus selon leur âge. De plus, il est fort possible que les effets associés à l'âge puissent aussi refléter une susceptibilité changeante de l'hôte à la maladie étudiée (ex. diminution du système immunitaire) et finalement, une exposition

¹⁰ Traduit de l'anglais : « *Because BC proved to be an outlier, particularly for CD, and because its population ethnic make-up is somewhat different from the rest of Canada, overall incidence rates and prevalence are extrapolated to the rest of Canada including and excluding BC data.*» (4)

clé peut avoir des effets immédiats ou différés sur un individu, se révélant seulement à un âge plus avancé (118).

Les résultats des analyses des modèles univariés incluant la variable âge ont aussi démontré l'importance de standardiser l'incidence brute de la MC pour ce facteur. Ces informations correspondent d'ailleurs à ce qui est retrouvé dans la littérature. En utilisant l'incidence de la MC standardisée comme variable dépendante dans les autres modèles de régression, il a ainsi été possible de déterminer si d'autres variables indépendantes y étaient associées.

Association entre la MC et le sexe des cas

Il est aussi intéressant de noter que le taux d'incidence de la MC au Québec est deux fois plus élevé chez les **femmes** que chez les hommes. Ce résultat semble être supérieur à ce qui est observé dans la littérature. En effet, en Nouvelle-Zélande, il a été établi qu'un peu plus de la moitié des patients atteints de MC sont de sexe féminin (39). En Italie, il ne semble pas y avoir de différence entre les femmes et les hommes (42) tandis qu'au Danemark, le ratio se rapproche plus de celui observé pour le Québec (24). Ainsi, par rapport aux résultats rapportés dans la littérature, il semble y avoir une plus grande différence au Québec entre le nombre de femmes et d'hommes affectés par la MC que dans les autres pays.

De plus, le taux d'incidence de la MC chez les femmes pour la période à l'étude est plus de deux fois celui observé chez les femmes du Danemark pour une période semblable (24). Cette différence pourrait par contre être attribuée en partie au fait que les taux d'incidence de la MC calculés dans l'étude danoise ont été ajustés pour l'âge, tandis que dans le présent ouvrage, ce sont des taux d'incidence brut.

Par ailleurs, les résultats des analyses du modèle univarié incluant la variable sexe supportent aussi l'importance de standardiser l'incidence brute de la MC pour ce facteur. Ces informations correspondent d'ailleurs à ce qui est retrouvé dans la littérature. En utilisant l'incidence de la MC standardisée pour l'âge et le sexe de la population du Québec de 1996 comme variable dépendante dans les autres modèles de régression, il a ainsi été possible de déterminer si d'autres variables indépendantes y

étaient associées en éliminant un possible biais de confusion dû aux différentes structures en âge et en sexe de la population.

Variation géographique de l'incidence de la MC au Québec et populations autochtones

La variation géographique nord-sud de l'incidence de la MC observée dans d'autres continents est aussi retrouvée dans la province de Québec. En effet, les taux d'incidence de la MC les plus faibles sont observés dans les CLSC se trouvant plus au nord de la province. Ces CLSC sont caractérisés par la présence de communautés autochtones, c'est pourquoi le modèle multivarié complet a été ajusté pour cette variable. Une association protectrice significative est observée entre la MC et le statut autochtone lors de l'analyse univariée, mais elle n'a pas été maintenue lorsqu'un ajustement pour d'autres variables associées de manière significative à la MC est fait. Ainsi, la probabilité de développer la MC dans des populations ayant une faible proportion d'individus autochtones serait plus élevée qu'ailleurs.

Bien qu'il ne soit pas juste d'inférer des résultats populationnels à l'échelle individuelle, il est pourtant tentant de croire qu'il y aurait un effet protecteur à la MC chez les autochtones du Québec. Selon la littérature, on peut supposer que cet effet pourrait possiblement provenir de leur bagage génétique. En effet, un phénomène similaire est retrouvé chez les Maoris de Nouvelle-Zélande. La MC n'est pas une maladie commune chez cette population aborigène du Pacifique sud et il a été démontré qu'il existe une différence marquée dans la fréquence allélique des mutations du gène *NOD2/CARD15* entre les Maoris et les groupes caucasiens de Nouvelle-Zélande. Une corrélation linéaire a de plus été observée entre la diminution de la fréquence allélique d'une variante génétique retrouvée en association avec les polymorphismes de *NOD2/CARD15* et l'augmentation du degré d'ancêtres maoris (91).

Variation de l'incidence de la MC entre les régions urbaines et rurales du Québec

L'étude de la distribution géographique de l'incidence de la MC a permis d'identifier que la majorité des taux d'incidence les plus faibles au Québec (75%) sont observés dans les régions rurales, tandis qu'un peu plus de la moitié des taux d'incidence les plus élevés (63,6%) sont retrouvés dans les régions urbaines. Les cartes

géographiques ont permis de visualiser ces résultats. Cependant, vu la répartition inégale de la population du Québec entre les régions rurales et urbaines, il est possible que les taux d'incidence extrêmes ne soient en réalité que le reflet de densités de population différentes. Par exemple, les CLSC qui ont un taux d'incidence nul tels Territoire Naskapi ou Baie-d'Hudson représentent des territoires très larges sans grands bassins de population. Par ailleurs, une association entre la variable « urbanité » et la MC a été observée lors de l'analyse du modèle de régression univarié et indique que le taux d'incidence de la MC est plus élevé en région urbaine qu'en rurale. Lorsque l'on compare ces observations avec celles retrouvées dans la littérature, l'estimé brut d'IRR obtenu dans le présent ouvrage pour la variable « urbanité » s'insère vraisemblablement parmi les résultats obtenus pour les autres provinces canadiennes puisque la moyenne est de 1,05 et les IRR varient entre 0,85 à 1,52 (4). Considérant l'idée véhiculée par l'hypothèse hygiénique selon laquelle un environnement grandement aseptisé favoriserait le développement de maladies inflammatoires, il serait donc probable que des particularités propres au mode de vie urbain influencent l'apparition de la MC.

Association entre le statut d'immigrant et la MC

Il a été observé que la MC n'est pas répartie uniformément parmi les différents groupes ethniques. Dans cet ouvrage, une forte proportion d'**immigrants** dans une population est associée à de plus faibles taux d'incidence de la MC. Il est possible que ce phénomène reflète en fait une répartition particulière de prédispositions génétiques dans certains groupes qui, jadis, évoluaient dans des territoires restreints, et qui se retrouvent désormais en mouvance. En effet, depuis le début des années 60, le pays d'origine de la majorité des nouveaux immigrants au Canada est passé de l'Europe de l'ouest à l'Asie et l'Afrique (143).

De plus, il a été démontré que certains groupes ethniques possèdent des prédispositions génétiques à la MC, tandis que d'autres populations sont moins susceptibles. Certains SNPs dans le gène *NOD2/CARD15* associés à la MC sont absents des populations africaines et asiatiques en apparence résistantes à la MC, mais sont présentes chez les Caucasiens, chez qui l'incidence de la MC est croissante. À ce sujet, il a d'ailleurs été proposé que l'absence ou la présence de ces SNP chez certaines populations suggère que ces mutations seraient apparues après l'expansion hors de l'Afrique de *Homo sapiens* (95). Cependant, le fait que les MII ne soient pas communément retrouvées

chez certaines populations lorsque celles-ci se retrouvent dans leur environnement d'origine contraste avec l'émergence de ces maladies chez certaines communautés émigrées. Cela supporte l'idée qu'un facteur environnemental pourrait être l'élément déclencheur de l'inflammation typique de la MC chez un hôte génétiquement prédisposé.

Par ailleurs, il existe dans la littérature une hypothèse qui implique la vaccination au bacille de Calmette-Guérin (BCG) contre la tuberculose et le développement de la MC. Deux camps s'opposent à ce sujet, c'est-à-dire ceux pour qui la vaccination au BCG en bas âge favorise le bon développement du système immunitaire, protégeant donc la santé des individus à long terme (25) et ceux qui ont observé une association entre la MC et la vaccination au BCG à l'enfance (144), suggérant ainsi le vaccin BCG comme un facteur de risque potentiel au développement d'une IBD. En effet, bien que dans les pays développés la vaccination au BCG ne se fasse plus systématiquement elle est toujours en vigueur dans les régions du monde où la tuberculose demeure un problème de santé publique important. Rua et ses collaborateurs évoquent donc la question suivante : *serait-il possible que l'immunoprotection induite par le vaccin BCG dans la prévention contre la tuberculose pourrait aussi être active dans la prévention de la MC?*¹¹. Dans le but de répondre à cette question, une révision des dossiers médicaux d'un hôpital en Argentine a été faite, où 100% des nouveau-nés reçoivent une dose de BCG. Une forte association entre la vaccination à un bas âge et l'absence de MC dans la population a ainsi été observée (25). Cette hypothèse pourrait peut-être en partie expliquer les plus faibles taux d'incidence de la MC dans des territoires de CLSC où la population compte une plus forte proportion d'individus immigrants. De plus, l'association protectrice significative observée entre la MC et le statut autochtone lors de l'analyse univariée pourrait potentiellement être expliquée par ce même phénomène puisqu'il est connu que la prévalence de la tuberculose chez les communautés autochtones et inuites du Québec est plus élevée que pour l'ensemble de la province. Ainsi, la vaccination au BCG est toujours en vigueur chez la majorité de ces groupes à risque. Il est aussi intéressant de noter que la plus forte incidence de la MC observée au Québec pour la période à l'étude est retrouvée dans le groupe d'âge des 20-29 ans et les individus de cette tranche d'âge

¹¹ Traduit de l'anglais : "Is it possible that the immunoprotection induced by BCG vaccination to prevent tuberculosis might also be active in preventing CD?" (25)

correspondent aux premières cohortes du Québec qui n'ont pas reçu le vaccin BCG à la naissance. En effet, ce type de vaccination y a été arrêtée en 1976.

Association entre la descendance juive ashkénaze et la MC

Il est reconnu que l'incidence de la MC est plus élevée chez les Juifs ashkénazes par rapport à n'importe quel groupe de référence. En effet, leur bagage génétique les prédisposerait au développement de la MC. Par exemple, environ la moitié d'entre eux est porteuse des mutations du gène *NOD2/CARD15* associées à la MC, ce qui représente presque le double de ce qui est observé chez les Juifs sépharades (93). Par ailleurs, il est reconnu que la prévalence de la MC est plus forte chez les Juifs d'Amérique du Nord par rapport aux Juifs d'Israël (94) soutenant aussi l'hypothèse qu'un facteur environnemental pourrait être l'élément déclencheur de l'inflammation typique de la MC chez un hôte génétiquement prédisposé.

Les résultats obtenus dans le présent ouvrage vont dans la même direction que ceux des autres études puisqu'ils démontrent qu'une forte proportion d'individus de descendance juive dans la population québécoise est associée à de hauts taux d'incidence de la MC. En effet, les taux d'incidence de la MC observés dans les territoires où les communautés juives représentent une forte proportion de la population sont presque deux fois plus élevés que ceux des territoires comptant peu ou pas d'individus juifs. Cette observation concorde d'ailleurs avec ce qui est retrouvé dans la littérature, puisqu'il a été démontré que les taux d'incidence et de prévalence des MII seraient de deux à quatre fois plus élevés chez les individus juifs ashkénazes par rapport aux Caucasiens non-juifs (85;93) De plus, cette association positive est maintenue après ajustement pour les variables « immigration » et « incidence de cinq MADO entériques ».

Il est de plus intéressant de noter que dans la ville de Montréal, les Juifs ashkénazes et sépharades sont principalement établis dans les quartiers Côte-Saint-Luc, Hampstead ainsi que Côte-des-Neiges (145). Dans le présent ouvrage, il s'avère que le CLSC Côte-Saint-Luc est le CLSC qui a le deuxième taux d'incidence le plus élevé au Québec et ce CLSC regroupe justement les quartiers Côte-Saint-Luc et Hampstead (146). Ainsi, la forte incidence de MC dans ce CLSC peut être expliquée par la présence marquée d'individus juifs ashkénazes.

Association entre l'incidence de cinq maladies entériques et la MC

Une association positive entre la MC et l'incidence cumulée de cinq MADO entériques a été observée, soit la combinaison des données d'incidence de la campylobactériose, yersiniose, giardiase, shigellose et salmonellose. Cette association a été observée lors de l'analyse univariée et maintenue après les ajustements pour d'autres variables des modèles multivariés.

En effet, nous avons observé que les CLSC du Québec où l'on retrouve des taux d'incidence intermédiaires et élevés de ces cinq MADO entériques sont associées à un haut taux d'incidence de la MC. Cependant, l'association positive entre cette variable et la MC obtenue dans cet ouvrage va à l'encontre des résultats obtenus au Manitoba par Bernstein et ses collaborateurs. Ceux-ci avaient en effet observé une association négative puisque le taux d'incidence de la MC est plus élevé dans les régions de cette province où l'incidence des maladies entériques est la plus basse (85). À ce propos, ils mentionnaient que leurs résultats correspondaient à la théorie de l'« hypothèse hygiénique » (69;70). Cependant, il est important de mentionner que l'hypothèse hygiénique fait référence à l'hygiène associée à l'enfance et les mesures écologiques du statut socio-économique et des taux d'incidence des maladies entériques de leur étude ont été créées avec des estimés de 1996-2001. Ainsi, assumer que la mesure de l'exposition de la population à ces deux facteurs de risque est équivalente à l'exposition à l'enfance signifie que les individus ne se déplacent pas avec les années (78).

Dans cet ouvrage, il a été démontré que l'incidence de la MC est plus élevée là où l'incidence de cinq MADO entériques est de niveau moyen ou élevé, et ce, même lorsqu'un ajustement pour d'autres variables est fait. Cela suggère que la MC serait associée avec la présence de MADO entériques, peu importe si l'incidence est de niveau intermédiaire ou élevé. Il est par ailleurs primordial de préciser que l'exposition à ces maladies entériques est mesurée dans un passé récent (et non à l'enfance), ainsi nous postulons que la réalité capturée par cette variable est celle de la contamination bactérienne générale de l'environnement. En effet, il va de soi d'assumer que les maladies infectieuses à déclaration obligatoire mentionnées dans cet ouvrage dépendent de la prévalence des bactéries ou parasites qui les causent. Ainsi, une incidence plus élevée de MADO entériques dans un territoire donné reflète probablement une forte prévalence des microorganismes qui les causent.

Explication de la variance de la MC par le modèle de régression final

La valeur de pseudo- R^2 du modèle de régression de Poisson multivarié réduit ajusté pour les variables indépendantes « immigration », « descendance juive » et « incidence des cinq MADO entériques » a été obtenue. Ce coefficient nous indique que le modèle final n'explique qu'en partie la variance de la variable dépendante prédite, laissant sous-entendre qu'un ou plusieurs autres facteurs non mesurés dans cette étude seraient associés au développement de la MC au Québec. Nous suggérons donc que la variance du taux d'incidence de la MC non-expliquée par le modèle de régression pourrait être due à la présence variable d'une bactérie sur le territoire du Québec, soit *Mycobacterium avium* sp. *paratuberculosis*. Ceci est supporté entre autres par l'observation que l'augmentation rapide dans l'incidence de la MC au sein de différentes populations et groupes ethniques suggère l'implication d'un facteur environnemental important dans le développement de cette maladie (147)

Biais écologique

En choisissant de mener une étude de type écologique, il est essentiel de démontrer une vigilance par rapport au biais écologique. En effet, le « piège écologique » à éviter est de projeter une découverte faite à l'échelle populationnelle à un niveau individuel. Dans ce type de devis, la maladie est représentée par une fréquence populationnelle qui inclut les exposés et les non-exposés. Il serait donc faux d'inférer les résultats obtenus dans cette étude au niveau individuel (118). Tout au long de cet ouvrage, un effort a ainsi été fait afin de garder les interprétations au niveau populationnel.

Limitations de l'étude

Certaines limitations ont été identifiées dans l'élaboration de cette étude. Tout d'abord, la génération des variables peut avoir introduit des biais. En effet, la majorité des variables ont été bâties à partir de données provenant de diverses sources. Ainsi, pour certaines d'entre elles (ex. contraceptifs oraux, tabagisme), les données obtenues sont des estimés régionaux que nous avons par la suite appliqués à la population des CLSC. Il peut donc y avoir eu une perte de précision lors de ce transfert de données d'un large territoire à un plus restreint.

De plus, toutes les variables incluses dans l'étude sont influencées par le fait qu'elles représentent la réalité d'exposition pour la période à l'étude et non pour l'enfance ou au cours d'une vie. En prenant en compte la possibilité que le degré d'exposition à certains facteurs puisse varier entre les CLSC et le fait que certains facteurs puissent avoir eu un effet différé sur les individus, il se peut qu'un biais se soit glissé dans l'étude. Aussi, la période à l'étude étant relativement longue (six ans), il est possible que les individus aient déménagé pendant ces années. Par exemple, un individu pourrait avoir été exposé à divers facteurs de risque au cours de sa vie en vivant dans un premier territoire de CLSC et avoir reçu un diagnostic de MC après être déménagé dans un autre territoire de CLSC. Le degré d'exposition aux diverses variables associé à ce cas serait donc erroné. Par ailleurs, le fait d'avoir utilisé les territoires de CLSC comme aire de diffusion a, d'une part, permis de maximiser l'homogénéité intra-groupe pour les variables à l'étude. En effet, ces territoires sont les plus petites aires de diffusion reliées à la santé et les variables à l'étude ont plus de chances d'être distribuées uniformément à l'intérieur de ces territoires. D'autre part, cette caractéristique a pu provoquer l'insertion d'un biais, plus spécifiquement lors de la comparaison des CLSC entre eux. En effet, il est possible qu'un effet de grappe soit présent dans les données, c'est-à-dire que des personnes semblables auront tendance à se regrouper dans un même territoire pour vivre, ce qui peut influencer les analyses. Lors de l'analyse des variables « âge » et « sexe », il a toutefois été possible d'intercepter cet effet lors des manipulations informatiques puisque ces variables comportent des mesures répétées pour chaque CLSC. Par contre, lors des analyses des modèles impliquant les autres variables indépendantes, cet effet n'a pas pu être corrigé dû au design de l'étude.

De plus, les deux variables « statut autochtone » et « descendance juive » sont en général distribuées de manière relativement homogène entre les CLSC du Québec, à l'exception de quelques CLSC minoritaires pour lesquels on observe une grappe de ces individus. Les niveaux les plus élevés de ces variables ne correspondent donc qu'à un très faible nombre de CLSC. En effet, seulement 21 CLSC sur 166 comptent une proportion plus élevée d'individus autochtones et seulement 3 CLSC sur 166 ont une forte proportion d'individus de descendance juive.

Il est finalement important de noter que l'issue des tests statistiques dépend entre autres de la taille de l'échantillon. Une taille d'échantillon élevée pourra mettre en évidence

une relation statistique, aussi faible soit-elle (118) et cela pourrait être le cas dans cet ouvrage puisque la taille de l'échantillon des cas pour la période à l'étude est large.

Ouverture vers le futur

Afin d'élargir le spectre de connaissances qui découlent des données obtenues dans le présent ouvrage, il pourrait être intéressant lors de recherches futures d'effectuer l'analyse spatiale de la répartition géographique de l'incidence de la MC au Québec. En effet, une variation du taux d'incidence entre les différents territoires de CLSC a été observée grâce à la visualisation des données par cartographie. Il serait alors primordial d'établir si cette répartition inégale de l'incidence de la MC reflète bel et bien une réelle hétérogénéité, c'est-à-dire que les résultats obtenus diffèrent de ceux que l'on observerait par chance seulement.

De plus, étant donné l'hypothèse selon laquelle une mycobactérie jouerait un rôle dans le développement de la MC, deux thèmes de recherche pourraient éventuellement être abordés. Tout d'abord, il serait pertinent de tenter d'isoler MAP à partir de l'environnement afin d'établir quelle est sa prévalence sur le territoire du Québec. De plus, d'une perspective de santé publique, il incombe aux microbiologistes et aux scientifiques environnementaux de déterminer si des organismes de MAP viables peuvent être cultivés à partir du lait de consommation, de l'eau municipale, de l'eau de puits et de la viande (101). De cette manière, d'autres études épidémiologiques au niveau populationnel pourraient évaluer si une association peut être faite entre ce microorganisme et l'incidence de la MC. Par ailleurs, nous croyons que l'hypothèse impliquant une association inverse entre la vaccination au BCG et le développement de la MC (25) pourrait aussi se refléter dans le cas d'infections naturelles par le microorganisme *Mycobacterium tuberculosis*. Au Canada, il est connu que le taux d'incidence de la tuberculose est plus élevé dans les population immigrantes et chez les autochtones par rapport au reste de la population (143). Il s'avère qu'une association négative a été observée dans cet ouvrage entre ces deux populations et la MC; c'est pourquoi il serait donc intéressant d'étudier s'il existe une association inverse entre l'exposition au *Mycobacterium tuberculosis* et la MC.

Finalement, puisqu'il a été démontré que certains marqueurs de susceptibilité génétique sont impliqués dans l'étiologie de la MC, il serait enrichissant d'établir quelle est la

répartition géographique des individus prédisposés génétiquement à cette maladie au Québec. En effet, au Manitoba, le risque attribuable à la population de développer la MC pour la présence de mutations du gène *NOD2/CARD15* a été identifié (148). De telles données pourraient apporter des précisions quant aux résultats obtenus dans cet ouvrage concernant les variables reliées à l'ethnicité puisqu'il a été démontré que les allèles des différentes mutations de *NOD2/CARD15* varient de façon significative parmi les sous-groupes ethniques (89). Cela pourrait suggérer que ces deux variables sont en fait des marqueurs du bagage génétique de certains groupes d'individus.

La compréhension actuelle de ce qu'est la MC ne permet pas de différencier si la maladie est réellement une entité ou plutôt un amalgame de plusieurs sous-types de MII tel que le suggèrent les découvertes en génomique (34). Plusieurs évidences soutiennent toutefois la possibilité que la MC représenterait un syndrome ayant différentes étiologies (147). Il est donc possible que du bruit de fond soit introduit dès que l'on tente d'associer la maladie à des facteurs de risque, amoindrissant ainsi toute association. Afin d'obtenir des réponses plus précises concernant l'association de certains facteurs à la MC, il serait probablement préférable de stratifier les études selon l'âge des cas de MC et de poser des questions plus spécifiques¹².

¹² Suggestion de Dre Irina Costea (LZA-ST), communication personnelle de Dr Pascal Michel (LZA-SSH)

Conclusion

Pratiquement tous les individus décelés par la définition opérationnelle de cas sont des cas réels de MC et ceux qui n'ont pas été identifiés comme étant des cas ne le sont pas non plus dans la réalité. Cela augmente la fiabilité des résultats obtenus dans cette étude. L'incidence de la MC au Québec est la plus élevée au Canada et selon notre connaissance de la littérature, la plus élevée au monde. De plus, l'incidence de la MC est plus élevée chez les femmes que chez les hommes ainsi que chez les jeunes dans la vingtaine. Finalement, la MC est associée positivement avec une forte proportion d'individus juifs dans la population ainsi qu'avec une forte incidence de cinq MADO entériques. Elle est aussi associée négativement avec une forte proportion d'immigrants dans la population. Une étude cas-témoin pourrait éventuellement être menée afin de capter au niveau individuel les informations concernant les variables à l'étude.

Sources documentaires

Reference List

- (1) Van Limbergen J, Russel R, Nimmo E, Satsangi J. The genetics of inflammatory bowel disease. *Am J Gastroenterol* 102, 1-12. 2007.
Ref Type: Generic
- (2) Torrence ME. Understanding epidemiology. Mosby's biomedical science series , 133-151. 1997.
Ref Type: Generic
- (3) Casanelli J. Maladie de Crohn: première description en Côte-d'Ivoire. *Med Trop* 64, 384-386. 2004.
Ref Type: Generic
- (4) Bernstein C, Wajda A, Svenson L et al. The epidemiology of inflammatory bowel disease in Canada: a population-based study. *Am J Gastroenterol* 101, 1559-1568. 2006.
Ref Type: Generic
- (5) Vermeire S, Wild G, Kocher K, Cousineau J, Dufresne L, Bitton A. CARD15 genetic variation in a Quebec population :prevalence, genotype-phenotype relationship, and haplotype structure. *Am.J.Hum.Gen* 71, 74-83. 2002.
Ref Type: Generic
- (6) Bernstein C, Rawsthorne P, Cheang M, Blanchard J. A population-based case control study of potential risk factors for IBD. Blanchard JF, ed. *Am J Gastroenterol* 101, 993-1002. 2006.
Ref Type: Generic
- (7) Bernstein C, Kraut A, Blanchard J, Rawsthorne P, Yu N, Walld R. The relationship between inflammatory bowel disease and socioeconomic variables. *Am J Gastroenterol* 96[7], 2117-2125. 2001.
Ref Type: Generic
- (8) Parkes M, Barrett J, Prescott N et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature Genetics* 39[7], 830-832. 2007.
Ref Type: Generic
- (9) Gibbs P. Epidemiological evidence for *Mycobacterium avium* sp paratuberculosis as a cause of Crohn's disease. *Epidemiol Infect* , 1-12. 2007.
Ref Type: Generic
- (10) Saxegaard F. Isolation of *Mycobacterium paratuberculosis* from intestinal mucosa and mesenteric lymph nodes of goats by use of selective Dubos medium. *J Clin*

- Microbio 22, 312-313. 1985.
Ref Type: Generic
- (11) Dalziel T. Chronic interstitial enteritis. *Br Med J* 2, 1068-1070. 1913.
Ref Type: Generic
- (12) Bernstein C, Blanchard J, Rawsthorne P, Wajda A. Epidemiology of Crohn's disease and ulcerative colitis in a central canadian province: a population-based study. *Am J Epidemiol* 149[10], 916-924. 1999.
Ref Type: Generic
- (13) Roy P-O. *Epidemiologie de la maladie de Crohn au Quebec*. 2007.
Ref Type: Generic
- (14) Picco MF. New studies on the role of smoking in inflammatory bowel disease: are we not beating a dead horse? *Am J Gastroenterol* 102, 598-600. 2007.
Ref Type: Generic
- (15) Crohn B. Regional ileitis. A pathological and clinical entity. *JAMA* 1932;251(1):73-9.
- (16) Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology* 126, 1504-1517. 2004.
Ref Type: Generic
- (17) Wakefield AJ. Pathogenesis of Crohn's disease: multifocal gastrointestinal infarction. *The Lancet* 1, 1057-1062. 1989.
Ref Type: Generic
- (18) Gasche C. A simple classification of Crohn's disease: report of the working party for the world congresses of gastroenterology. *Inflamm Bowel Dis* 6[1], 8-15. 2000. Vienne 1998.
Ref Type: Generic
- (19) Ekbohm A. The epidemiology of inflammatory bowel disease: a large, population-based study in Sweden. *Gastroenterology* 100[2], 350-358. 1991.
Ref Type: Generic
- (20) Sartor R. Microbial influences in inflammatory bowel disease: role in pathogenesis and clinical implications. Sartor RB, ed. *Kirsner's inflammatory bowel disease*, 139-162. 2004.
Ref Type: Generic
- (21) Bloom S. Should we attack the environment in IBD? Programme - Inflammatory Bowel Disease 2007, Congress of the European Crohn's and Colitis Organisation, 23. 3-2-2007.
Ref Type: Generic

- (22) Pinsk V, Lemberg D, Grewal K, Barker C, Schreiber R, Jacobson K. Inflammatory bowel disease in the south asian pediatric population of British-Columbia. *Am J Gastroenterol* 102, 1077-1083. 2007.
Ref Type: Generic
- (23) Rajan T. Maladaptation of the vigilant genotype in a hyper-clean world? *Perspectives in Biology and Medicine* 49[2], 171-177. 2006.
Ref Type: Generic
- (24) Jacobsen B, Fallingborg J, Rasmussen H, Nielsen K, Sorensen H. Increase in incidence and prevalence of inflammatory bowel disease in northern Denmark: a population-based study, 1978-2002. *Eur J Gastroenterol Hepatol* 18, 601-606. 2006.
Ref Type: Generic
- (25) Rua E, Blacarce N, Drut R. Early Bacille de Calmette-Guérin vaccination and Crohn's disease: possible relationship? *Pediatric and Developmental Pathology* 3, 501-503. 2000.
Ref Type: Generic
- (26) Ouyang Q, Tandon R, Goh K, Pan G, Fock K, Xiao S. Management consensus of inflammatory bowel disease for the Asia-Pacific region. *Journal of Gastroenterology and hepatology* 21, 1772-1782. 2006.
Ref Type: Generic
- (27) Loftus EV Jr. Update on the incidence and prevalence of Crohn's disease and Ulcerative colitis in Olmsted County, Minnesota. *Inflamm Bowel Dis* 13[March (3)], 254-261. 2007.
Ref Type: Generic
- (28) Nerich V, Monnet E, Etienne A, Louafi S, Ramée C, Rican S. Geographical variations of inflammatory bowel disease in France: a study based on national health insurance data. *Inflamm Bowel Dis* 12[3], 218-226. 2006.
Ref Type: Generic
- (29) Baumgart D, Carding S. Inflammatory bowel disease: cause and immunobiology. *The Lancet* 369, 1627-1640. 2007.
Ref Type: Generic
- (30) Arnott I, Nimmo E, Drummond H, Fennell J, Smith B, Satsangi J. NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? *Genes and Immunity* 5, 417-425. 2004.
Ref Type: Generic
- (31) Desreumeaux P. The role of fat, adipocytes and fat wrapping. Programme - Inflammatory Bowel Disease 2007, Congress of the European Crohn's and Colitis Organisation, 17. 3-1-2007.
Ref Type: Generic

- (32) Hugot J. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 379, 821-823. 1996.
Ref Type: Generic
- (33) Negoro K. Analysis of the *IBD5* locus and potential gene-gene interactions in Crohn's disease. *Gut* 52[4], 541-546. 2003.
Ref Type: Generic
- (34) Rioux J, Xavier R, Taylor K et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genetics* 39[5], 596-604. 2007.
Ref Type: Generic
- (35) Hugot J, Chamaillard M, Zouali H, Lesage S, Cézard J, Belaiche J. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411, 599-603. 2001.
Ref Type: Generic
- (36) Ogura Y, Bonen D, Inohara N, Nicolae D, Chen F, Ramos R. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411, 603-606. 2001.
Ref Type: Generic
- (37) Hildebrand H, Finkel Y, Grhnquist L, Lindholm J, Ekbohm A, Askling J. Changing pattern of paediatric inflammatory bowel disease in northern Stockholm 1990-2001. *Gut* 52, 1432-1434. 2003.
Ref Type: Generic
- (38) Munkholm P, Langholz E, Nielsen O, Kreiner S, Binder V. Incidence and prevalence of Crohn's disease in the county of Copenhagen, 1962-1987: a sixfold increase in incidence. *Scand J Gastroenterol* 7, 609-614. 1992.
Ref Type: Generic
- (39) Geary R, Richardson A, Frampton C et al. High incidence of Crohn's disease in Canterbury, New Zealand: results of an epidemiologic study. *Inflamm Bowel Dis* 12[10], 936-943. 2006.
Ref Type: Generic
- (40) Morita N. Incidence and prevalence of inflammatory bowel disease in Japan: nationwide epidemiological survey during the year 1991. *J.Gastroenterol.* 30[8], 1-4. 1995.
Ref Type: Generic
- (41) Leong R. The epidemiology and phenotype of Crohn's disease in the Chinese population. *Inflamm Bowel Dis* 10, 646-651. 2004.
Ref Type: Generic
- (42) Tragnone A. Incidence of inflammatory bowel disease in Italy: a nationwide population-based study. *Cruppo Italiano per lo Studio del Colon e del Retto*

- (GISC). *Int J Epidemiol* 25[5], 1044-1052. 1996.
Ref Type: Generic
- (43) Lesage S. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Genet* 70, 845-857. 2002.
Ref Type: Generic
- (44) Hodgson H. Pathogenesis of Crohn's disease. *Bailliere's Clin Gastroenterol* 12, 1-17. 1998.
Ref Type: Generic
- (45) Desai HG, Gupte PA. Increasing incidence of Crohn's disease in India: Is it related to improved sanitation? *Indian J Gastroenterol* 24, 23-24. 2007.
Ref Type: Generic
- (46) Cosnes J. Tobacco and IBD: relevance in the understanding of disease mechanisms and clinical practice. Best practices and research *Clinical Gastroenterology* 18[3], 481-496. 2004.
Ref Type: Generic
- (47) Calkins BM. A meta-analysis of the role of smoking in inflammatory bowel disease. *Digestive diseases and Science* 34, 1841-1854. 1989.
Ref Type: Generic
- (48) Corrao G, Tragnone A, Caprilli R et al. Risk of inflammatory bowel disease attributable to smoking, oral contraception and breastfeeding in Italy: a nationwide case-control study. *International Journal of Epidemiology* 27, 397-404. 1998.
Ref Type: Generic
- (49) Regueiro M. Cigarette smoking and age at diagnosis of inflammatory bowel disease. *Inflamm Bowel Dis* 11[1], 42-47. 2005.
Ref Type: Generic
- (50) Mahid S, Minor K, Stromberg A, Galandiuk S. Active and passive smoking in childhood is related to the development of inflammatory bowel disease. *Inflamm Bowel Dis* 13[4], 431-438. 2007.
Ref Type: Generic
- (51) Jang J, Kim H, Jung J, Chae M, Kim N. The role of smoking as a risk factor in inflammatory bowel diseases: single center study in Korea. *Korean J Gastroenterol* 47[3], 198-204. 2006.
Ref Type: Generic
- (52) Birrenbach T. Inflammatory bowel disease and smoking: A review of epidemiology, pathophysiology and therapeutic implications. *Inflamm Bowel Dis* 10, 848-859. 2007.
Ref Type: Generic

- (53) Bridger S, Lee J, Bjarnason I, Lennard Jones J, Macpherson A. In sibilings with similar genetic susceptibility for inflammatory bowel disease, smokers tend to develop Crohn's disease and non-smokers develop ulcerative colitis. *Gut* 51, 21-25. 2002.
Ref Type: Generic
- (54) Guo X, Ko J, Mei Q, Cho C. Aggravating effect of cigarette smoke exposure on experimental colitis is associated with leukotriene B4 and reactive oxygen metabolites. *Digestion* 63, 180-187. 2001.
Ref Type: Generic
- (55) Holmes EW. Glutathione content of colonic mucosa: evidence for oxidative damage in active ulcerative colitis. *Dig Dis Sci* 43, 1088-1095. 1998.
Ref Type: Generic
- (56) Godet PG. Meta-analysis of the role of oral contraceptive agents in inflammatory bowel disease. *Gut* 37, 669-673. 1005.
Ref Type: Generic
- (57) Sicilia B. Environmental risk factors and Crohn's disease: a population-based, case-control study in Spain. *Dig Liver Dis* 33[9], 762-767. 2001.
Ref Type: Generic
- (58) Wakefield A, Sawyerr A, Hudson M, Dhillon A, Pounder R. Smoking, the oral contraceptive pill, and Crohn's disease. *Dig Dis Sci* 36, 1147-1150. 1991.
Ref Type: Generic
- (59) Boyko E, Theis M, Vaughan T, Nicol-Blades B. Increased risk of inflammatory bowel disease associated with oral contraceptive use. *Am J Epidemiol* 140[3], 268-278. 1994.
Ref Type: Generic
- (60) Rothman K. *Modern epidemiology*. 570-575. 2007.
Ref Type: Generic
- (61) Rogers BHG. The epidemiologic and demographic characteristics of inflammatory bowel disease: an analysis of a computerized file of 1400 patients. *J Chron Dis* 24, 743-773. 1971.
Ref Type: Generic
- (62) Pinchbeck BR. Effect of religious affiliation and education status on the prevalence of inflammatory bowel disease in northern Alberta. *Can J Gastroenterol* 2 suppl A, 95-100A. 1988.
Ref Type: Generic
- (63) Bonnevie O. A socio-economic study of patients with ulcerative colitis. *Scand J Gastroenterol* 2, 129-136. 1967.
Ref Type: Generic

- (64) Keighley A. The demographic and social characteristics of patients with Crohn's disease in the Nottingham area. *Scand J Gastroenterol* 11, 293-296. 1976.
Ref Type: Generic
- (65) Sorensen VZ. Life prospects and quality of life in patients with Crohn's disease. *Gut* 28, 382-385. 1987.
Ref Type: Generic
- (66) Jantchou P, Monnet E, Carbonnel F. Environmental risk factors in Crohn's disease and ulcerative colitis (excluding tobacco and appendectomy). *Gastroenterol Clin Biol.* 30[6-7], 859-867. 2006.
Ref Type: Abstract
- (67) Devendra K. Investigating the hygiene hypothesis as a risk factor in pediatric onset Crohn's disease: a case-control study. *Am J Gastroenterol* 101, 1005-1011. 2006.
Ref Type: Generic
- (68) Wells RW. The increasing prevalence of Crohn's disease in industrialized societies: the price of progress? *Can J Gastroenterol* 19, 89-95. 2005.
Ref Type: Generic
- (69) Strachan D. Hay fever, hygiene, and household size. *BMJ* 299[6710], 1259-1260. 1989.
Ref Type: Generic
- (70) Gent A, Ellier M, Grace R, Swarbrick E, Coggon D. Inflammatory bowel disease and domestic hygiene in infancy. *The Lancet* 343, 766-767. 1994.
Ref Type: Generic
- (71) Macpherson A. Interactions between commensal intestinal bacteria and the immune system. *Nat Rev Immunol* 4, 478-485. 2004.
Ref Type: Generic
- (72) Shanahan F. Microbial-epithelial interaction in IBD. Programme - Inflammatory Bowel Disease 2007, Congress of the European Crohn's and Colitis Organisation, 16. 3-1-2007.
Ref Type: Generic
- (73) Crabbé PA. Immunohistochemical observations on lymphoid tissues from conventional and germ-free mice. *Lab Invest* 22, 448-457. 1970.
Ref Type: Generic
- (74) Rousseau M. Potential health effects from non-specific stimulation of the immune function in early age: the example of BCG vaccination. *Pediatr Allergy Immunol* . 2007.
Ref Type: Generic
- (75) Wakefield A, Murch S, Anthony A. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 351,

637-641. 1998.

Ref Type: Generic

- (76) Seagroatt V. MMR vaccine and Crohn's disease: ecological study of hospital admissions in England, 1991 to 2002. *BMJ* 330, 1120-1121. 2005.
Ref Type: Generic
- (77) Wurzelmann J. Childhood infections and the risk of inflammatory bowel disease. *Dig Dis Scie* 39, 555-560. 1994.
Ref Type: Generic
- (78) Moayyedi P. Invited commentary: clues to the etiology of inflammatory bowel disease - a return to John Snow? *Am J Epidemiol* 164, 624-626. 2006.
Ref Type: Generic
- (79) Sands B. Inflammatory bowel disease: past, present, and future. *J Gastroenterol* 42, 16-25. 2007.
Ref Type: Generic
- (80) Garcia Rodriguez LA. Acute gastroenteritis is followed by an increased risk of inflammatory bowel disease. *Gastroenterology* 130, 1588-1594. 2006.
Ref Type: Generic
- (81) Marteau P. Review article: gut flora and inflammatory bowel disease. *Aliment Pharmacol Ther* 20[4], 18-23. 2004.
Ref Type: Generic
- (82) Seksik P. Alterations of the dominant faecal bacterial groups in patients with Crohn's diseases of the colon. *Gut* 52, 237-242. 2003.
Ref Type: Generic
- (83) Guarner F. Mechanisms of disease: the hygiene hypothesis revisited. *Nature* 3[5], 275-284. 2006.
Ref Type: Generic
- (84) Karlinger K, Gyorke T, Mako E, Mester A, Tarjan Z. The epidemiology and the pathogenesis of inflammatory bowel disease. *Eur J Radiol* 35[3], 154-167. 2000.
Ref Type: Generic
- (85) Green C, Elliott L, Beaudoin C, Bernstein C. A population-based ecologic study of inflammatory bowel disease: searching for etiologic clues. *Am J Epidemiol* 164[7], 615-623. 2006.
Ref Type: Generic
- (86) Nguyen GC. Inflammatory bowel disease characteristics among African Americans, Hispanics, and non-Hispanic Whites: characterization of a large North American cohort. *Am J Gastroenterol* 101, 1012-1023. 2006.
Ref Type: Generic

- (87) Finlay D, Basu D, Sellin J. Effect of race and ethnicity on perceptions of inflammatory bowel disease. *Inflamm Bowel Dis* 12[6], 503-507. 2006.
Ref Type: Generic
- (88) Basu D, Lopez I, Kulkarni A, Sellin J. Impact of race and ethnicity on inflammatory bowel disease. *Am J Gastroenterol* 100[10], 2254-2261. 2005.
Ref Type: Generic
- (89) Economou M. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am J Gastroenterol* 99, 2393-2404. 2004.
Ref Type: Generic
- (90) Inoue N. Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology* 123, 86-91. 2002.
Ref Type: Generic
- (91) Gearry R, Lea R, Roberts R, Chambers G, Barclay M, Kennedy M. CARD15 allele frequency differences in New Zealand Maori: ancestry specific susceptibility to Crohn's disease in New Zealand? *Gut* 55, 580. 2006.
Ref Type: Generic
- (92) Congilosi S, Rosendale D, Herman D. Crohn's disease - a rare disorder in American indians. *West J Med* 157[6], 682. 1992.
Ref Type: Generic
- (93) Karban A, Waterman M, Panhuysen C, Dresner Pollak R, Neshet S, Datta L. NOD2/CARD15 genotype and phenotype differences between Ashkenazi and Sephardic Jews with Crohn's disease. *Am J Gastroenterol* 99[6], 1134-1140. 2004.
Ref Type: Generic
- (94) Karban A. The relation between NOD2/CARD15 mutations and the prevalence and phenotypic heterogeneity of Crohn's disease: lessons from the Israeli Arab Crohn's disease cohort. *Digestive Diseases and Sciences* 50[9], 1692-1697. 2005.
Ref Type: Generic
- (95) Stockton J, Howson J, Awomoyi A, McAdam K, Blackwell J, Newport M. Polymorphism in NOD2, Crohn's disease, and susceptibility to pulmonary tuberculosis. *FEMS Immunology and Medical Microbiology* 41, 157-160. 2004.
Ref Type: Generic
- (96) Montgomery S. Asian ethnic origin and the risk of inflammatory bowel disease. *Eur J Gastroenterol Hepatol* 11[5], 543-546. 1999.
Ref Type: Generic
- (97) Tsironi E, Feakins R, Probert C, Rampton D, Phil D. Incidence of inflammatory bowel disease is rising and abdominal tuberculosis is falling in Bangladeshis in East London, United Kingdom. *Am J Gastroenterol* 99[9], 1749-1755. 2004.
Ref Type: Generic

- (98) Freeman H, Hershfield N. Anticipation in an Indo-Canadian family with Crohn's disease. *Can J Gastroenterol* 15[10], 695-698. 2001.
Ref Type: Generic
- (99) Joossens M, Simoens M, Vermeire S, Bossuyt X, Geboes K, Rutgeerts P. Contribution of genetic and environmental factors in the pathogenesis of Crohn's disease in a large family with multiple cases. *Inflamm Bowel Dis* 13[5], 580-584. 2007.
Ref Type: Generic
- (100) Straus W, Eisen G, Sandler R, Murray S, Sessions J. Crohn's disease: does race matter? *Am J Gastroenterol* 95[2], 479-483. 2000.
Ref Type: Generic
- (101) Behr M, Schurr E. Mycobacteria in Crohn's disease: a persistent hypothesis. *Inflamm Bowel Dis* 12[9], 1000-1004. 2006.
Ref Type: Generic
- (102) Manning E. Mycobacterium avium subspecies paratuberculosis: a review of current knowledge. *Journal of Zoo and Wildlife Medicine* 32[3], 293-304. 2001.
Ref Type: Generic
- (103) Chiodini R, Van Kruiningen H, Merkal R, Thayer W, Coutu J. Characteristics of an unclassified Mycobacterium species isolated from patients with Crohn's disease. *J Clin Microbio* 20[5], 966-971. 1984.
Ref Type: Generic
- (104) Jeyanathan M, Boutros-Tadros O, Radhi J, Semret M, Bitton A, Behr M. Visualization of Mycobacterium avium in Crohn's tissue by oil-immersion microscopy. *Microb Infect*. In press 2007.
- (105) Green E. Sequence and characteristics of IS900, an insertion element identified in a human Crohn's disease isolate of Mycobacterium paratuberculosis. *Nucleic Acids Res* 17, 9063-9073. 1989.
Ref Type: Generic
- (106) Semret M, Alexander D, Turenne C et al. Genomic polymorphisms for Mycobacterium avium subsp. paratuberculosis diagnostics. *J Clin Microbio* 43[8], 3704-3712. 2007.
Ref Type: Generic
- (107) Turenne C, Semret M, Cousins D, Collins D, Behr M. Sequencing of hsp65 distinguishes among subsets of the Mycobacterium avium complex. *J Clin Microbio* 44[2], 433-440. 2006.
Ref Type: Generic
- (108) Abubakar I. Detection of Mycobacterium avium subspecies paratuberculosis from patients with Crohn's disease using nucleic acid-based techniques: a systematic review and meta-analysis. *Inflamm Bowel Dis*. In press 2007.

- (109) Sechi L, Scanu A, Molicotti P et al. Detection and isolation of *Mycobacterium avium* subspecies paratuberculosis from intestinal mucosal biopsies of patients with and without Crohn's disease in Sardinia. *Am J Gastroenterol* 100, 1529-1536. 2005.
Ref Type: Generic
- (110) Tzen C. Detection of mycobacteria in Crohn's disease by a broad spectrum polymerase chain reaction. *J Formos Med Assoc* 105[4], 290-298. 2006.
Ref Type: Generic
- (111) Semret M. Insertion Sequence IS900 revisited. *J Clin Microbiol* 44[3], 1081-1083. 2006.
Ref Type: Generic
- (112) Behr M, Kapur V. The evidence for *Mycobacterium paratuberculosis* in Crohn's disease. Lowe A, ed. 2007.
Ref Type: Personal Communication
- (113) Chamberlin W, Ghobrial G, Chehtane M, Naser S. Successful treatment of a Crohn's disease patient infected with bacteremic *Mycobacterium paratuberculosis*. *Am J Gastroenterol* 102[3], 689-691. 2007.
Ref Type: Generic
- (114) Feller M. *Mycobacterium avium* subspecies paratuberculosis and Crohn's disease: a systematic review and meta-analysis. *The Lancet* 7, 607-613. 2007.
Ref Type: Generic
- (115) Whittington R. Survival of *Mycobacterium avium* subsp. paratuberculosis in dam water and sediment. *Appl Environ Microbiol* 71[9], 5304-5308. 2005.
Ref Type: Generic
- (116) Caribvet. Paratuberculose. Carribean animal health network 2007; Available from: URL: <http://www.caribvet.net/upload/Paratuberculose.pdf>
- (117) RAIZO. Revue d'épidémiologie animale du RAIZO - Paratuberculose. mapaq 2006;26. Available from: URL: http://www.mapaq.gouv.qc.ca/NR/rdonlyres/608F1531-B9B9-4FBE-88D2-D0E95411252D/0/Revue_RAIZO_bilan_2006.pdf
- (118) Philippe P. Introduction au cours d'épidémiologie MSO6001. Université de Montréal . 2006.
Ref Type: Generic
- (119) Statistique Canada. Classification canadienne des actes diagnostiques, thérapeutiques et chirurgicaux (CCADTC). Centre canadien d'information sur la santé , 200 pp. 1992.
Ref Type: Generic
- (120) Gouvernement du Québec - portail Québec. Renseignements qui figurent sur la carte d'assurance maladie avec photographie. Répertoire des programmes -

- services aux citoyens 2007; Available from: URL:
http://www.formulaire.gouv.qc.ca/cgi/affiche_doc.cgi?dossier=279&table=0
- (121) Hennekens C. *Épidémiologie en médecine*. 66-75. 1998.
 Ref Type: Generic
- (122) Gouvernement du Québec. Régie de l'assurance maladie du Québec: banques de données. Régie de l'assurance maladie du Québec 2003; Available from: URL:
<http://www.ramq.gouv.qc.ca/fr/statistiques/banques/index.shtml>
- (123) Vérificateur général du Québec. *Rapport du Vérificateur général sur les indicateurs de santé*. Vérificateur général du Québec 2007; Available from: URL:
http://www.vgq.qc.ca/publications/rapp_2002_2/Rapport/html/07.html
- (124) Gouvernement du Québec. Atlas de la santé et des services sociaux du Québec. Atlas-Statistiques-MSSS 2007; Available from: URL:
<http://www.msss.gouv.qc.ca/statistiques/atlas/atlas/index.php>
- (125) Statistique Canada. À propos de nous: notre travail. Statistique Canada 2007; Available from: URL: http://www.statcan.ca/francais/about/overview_f.htm
- (126) IMS Health incorporated. Tout sur IMS: présentation de l'entreprise. IMS - Intelligence applied 2007; Available from: URL:
http://www.imshealth.com/web/channel/0,3147,77303623_63872710_77770062,00.html
- (127) Gouvernement du Québec. Rapport annuel MADO 2002. Ministère de la Santé et des Services sociaux 2005; 2-25. Available from: URL:
<http://publications.msss.gouv.qc.ca/acrobat/f/documentation/2005/05-268-02W.pdf>
- (128) Gouvernement du Québec. Maladies à déclaration obligatoire. Ministère de la Santé et des Services sociaux 2007; Available from: URL:
<http://www.msss.gouv.qc.ca/sujets/santepub/mado.php>
- (129) Rothman K. *Modern epidemiology*. [second edition], 459-465. 2007.
 Ref Type: Generic
- (130) Hennekens C. *Épidémiologie en médecine*. 92. 1998.
 Ref Type: Generic
- (131) Ressources naturelles Canada. L'Atlas du Canada - symboles cartographiques. L'Atlas du Canada 2004; Available from: URL:
http://atlas.nrcan.gc.ca/site/francais/learningresources/cartocorner/map_content_carto_symbology.html/document_view
- (132) Ressources naturelles Canada. L'Atlas du Canada - les projections cartographiques. L'Atlas du Canada 2004; Available from: URL: (
<http://atlas.nrcan.gc.ca/site/francais/learningresources/cartocorner/index.html>)

- (133) Pampalon R. Un indice de défavorisation pour la planification de la santé et du bien-être au Québec. Raymond G, ed. *Maladies chroniques au Canada* 21[4]. 2000. Agence de la santé publique du Canada.
Ref Type: Generic
- (134) Townsend P. Deprivation. *J Soc Policy* 16[2], 125-146. 1987.
Ref Type: Generic
- (135) Oxford journals. chap 13 - Poisson regression analysis. Oxford university press 2007; Available from: URL:
http://www.oxfordjournals.org/tropej/online/ma_chap13.pdf
- (136) Labbe A. La Régression de Poisson. Département de mathématiques et statistique de l'Université Laval 2007; Available from: URL:
https://cours.mat.ulaval.ca/fileadmin/cours/20071_16569A_65623A/public/CHAP1TRE6.pdf
- (137) SAS Institute Inc. SAS help and documentation - Reg procedure. 2003.
Ref Type: Generic
- (138) UCLA Academic Technology Services. Chapter 2 - Regression diagnostics, 2.4 Tests for collinearity. UCLA Academic Technology Services 2007; Available from: URL: <http://www.ats.ucla.edu/stat/sas/webbooks/reg/chapter2/sasreg2.htm>
- (139) Kleinbaum D. *Applied regression analysis and other multivariable methods*. 2nd ed. Boston: PWS-KENT, 1988.
- (140) Matthews D. *The Analysis of Longitudinal Data*. In: Karger, ed. *Using and understanding medical statistics*. 3rd ed. Basel: 1996: 161-71.
- (141) SAS Institute Inc. **Generalized Estimating Equations (GEE)**. Support SAS - Statistics and Operations research 2007; Available from: URL:
<http://support.sas.com/rnd/app/da/new/dagee.html>
- (142) Nagelkerke N. A note on a general definition of the coefficient of determination. *Biometrika* 78, 691-692. 1991.
Ref Type: Generic
- (143) Long R. The Canadian lung association/Canadian thoracic society and tuberculosis prevention and control. *Can Respir J* 14[7], 427-431. 2007.
Ref Type: Generic
- (144) Baron S, Turck D, Leplat C, Colombel J. Environmental risk factors in paediatric inflammatory bowel disease: a population based case control study. *Gut* 54, 357-363. 2005.
Ref Type: Generic
- (145) Paré S. Le rôle des agents immobiliers dans la ségrégation résidentielle ethnique: processus et mécanismes de production. Université de Montréal 1998;180.

- (146) Blondeau M. Codes et noms des territoires et des installations CLSC par RSS, par RLS et identifiant le RUIS. Publications du MSSS . 2006.
Ref Type: Generic
- (147) Hulten K, Almashhrawi A, El-Zaatari F, Graham D. Antibacterial therapy for Crohn's disease: a review emphasizing therapy directed against Mycobacteria. *Dig Dis Sci* 45[3], 445-456. 2000.
Ref Type: Generic
- (148) Brant S. A population-based case-control study of CARD15 and other risk factors in Crohn's disease and ulcerative colitis. Bernstein CN, ed. *Am J Gastroenterol* 102[2], 313-323. 2007.
Ref Type: Generic

Annexes

Annexe 1 : Liste des variables principales et sous-variables dans la base de données

Identification des CLSC

Nom de variable : IDCLSC

La variable IDCLSC est l'identifiant numérique des CLSC. Cette variable est basée sur les CLSC de 2005 et correspond aussi à ceux de 2006. Les CLSC de 2005 inclut 166 CLSC. Par rapport aux CLSC de 2004, le CLSC 1306 fictif de 2004 (combinaison des CLSC 1302 et 1305) est devenu le CLSC 1061 (Temiscouata), le CLSC 2107 fictif de 2004 (combinaison des CLSC 2102 et 2106) est devenu le CLSC 2107 (Saguenay) et le CLSC 5111 fictif de 2004 (combinaison de 5107 et 5108) est devenu le CLSC 5071 (Sherbrooke). Les CLSC 2004 comptaient 169 CLSC. Cette variable est nominale.

Nom de variable : AIDCLSC

La variable AIDCLSC est l'identifiant alpha-numérique des CLSC, basée sur la variable IDCLSC. Cette variable est nominale.

Nom de variable : NOMCLSC

La variable NOMCLSC est le nom des CLSC, basée sur les CLSC de 2005, sans accent. Cette variable est nominale.

Nom de variable : CLSC04

La variable CLSC04 est l'identifiant CLSC pour 2004. Cette variable est nominale.

Nom de variable : CLSC05

La variable CLSC05 est l'identifiant CLSC pour 2005. Cette variable correspond également à l'identifiant CLSC pour 2006. Cette variable est nominale.

Nom de variable : RSS

La variable RSS est l'identifiant numérique des régions socio-sanitaires (RSS), basée sur les RSS de 2006, sans accent. Les RSS correspondent aux régions administratives au Québec. Cette variable est nominale.

Nom de variable : NOMRSS

La variable NOMRSS est le nom des RSS, basée sur les RSS de 2006, sans accent. Les RSS correspondent aux régions administratives au Québec. Cette variable est nominale.

Nombre brut de cas de la maladie de Crohn par âge et par sexe, de 1995 à 2000

Nom de variable : CASF18

La variable CASF18 est le nombre brut de cas de maladie de Crohn chez les femmes âgées de moins de 18 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : CASF1864

La variable CASF1864 est le nombre brut de cas de maladie de Crohn chez les femmes âgées entre 18 et 64 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : CASF64

La variable CASF64 est le nombre brut de cas de maladie de Crohn chez les femmes âgées de plus de 64 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : CASFTOT

La variable CASFTOT est le nombre brut de cas de maladie de Crohn chez les femmes de tous âges, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : CASM18

La variable CASM18 est le nombre brut de cas de maladie de Crohn chez les hommes âgés de moins de 18 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : CASM1864

La variable CASM1864 est le nombre brut de cas de maladie de Crohn chez les hommes âgés entre 18 et 64 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : CASM64

La variable CASM64 est le nombre brut de cas de maladie de Crohn chez les hommes âgés de plus de 64 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : CASMTOT

La variable CASMTOT est le nombre brut de cas de maladie de Crohn chez les hommes de tous âges, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : CASTOT

La variable CASTOT est le nombre brut de cas de maladie de Crohn total chez les femmes et les hommes de tous âges, entre 1995-2000, soit l'addition de CASFTOT et CASMTOT. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Population québécoise de 1996

Les données de population du Québec de 1996 ont été choisies car il s'agissait d'une année centrale pour les données de départ (entre 1993 et 2002).

Nom de variable : POPF

La variable POPF est le nombre de femmes présentes dans la population du Québec en 1996.

Nom de variable : POPM

La variable POPM est le nombre d'hommes présents dans la population du Québec en 1996.

Nom de variable : POPTOT

La variable POPTOT est le nombre d'individus présents dans la population du Québec en 1996. La population de 1996 représente le dénominateur de l'incidence (doit multiplier par 6 pour obtenir des personnes/années, en assumant que les 6 années ont eu des populations similaires).

Nom de variable : POPTOTX6

La variable POPTOTX6 est le nombre d'individus présents dans la population du Québec en 1996 multiplié par 6 pour obtenir des personnes/années. Cette variable représente le dénominateur de l'incidence.

Taux d'incidence 1995-2000 de la maladie de Crohn standardisé pour l'âge et le sexe de 1996

Nom de variable : TAUX10

La variable TAUX10 est le taux d'incidence de la maladie de Crohn, standardisé pour l'âge et le sexe, par 10 000 habitants.

Groupes d'âge

La variable prédictive « Groupes d'âge » fourni des valeurs d'incidence de la maladie de Crohn pour 5 groupes d'âge différents. Le groupe 0-19 ans a été choisi comme groupe de référence puisqu'il représente le groupe d'individus « non-exposés » au facteur de risque « âge ». Il est observé que la maladie se développe dans la vingtaine.

Nom de variable : Cas019

La variable CAS019 est le nombre brut de cas de maladie de Crohn chez les hommes et femmes âgés de moins de 19 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : Cas2029

La variable CAS2029 est le nombre brut de cas de maladie de Crohn chez les hommes et femmes âgés entre 20 et 29 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : Cas3039

La variable CAS3039 est le nombre brut de cas de maladie de Crohn chez les hommes et femmes âgés entre 30 et 39 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : Cas4054

La variable CAS4054 est le nombre brut de cas de maladie de Crohn chez les hommes et femmes âgés entre 40 et 54 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Nom de variable : Cas50+

La variable CAS50+ est le nombre brut de cas de maladie de Crohn chez les hommes et femmes âgés de plus de 50 ans, entre 1995-2000. Il d'agit du cumul des cas identifiés par Pierre-Olivier pour chaque CLSC.

Population québécoise de 1996 – pour groupes d'âge

Les données de population du Québec de 1996 ont été choisies car il s'agissait d'une année centrale pour les données de départ (entre 1993 et 2002).

Nom de variable : Pop019

La variable POP019 est le nombre de femmes et d'hommes âgés de moins de 19 ans présents dans la population du Québec en 1996.

Nom de variable : Pop2029

La variable POP2029 est le nombre de femmes et d'hommes âgés entre 20 et 29 ans présents dans la population du Québec en 1996.

Nom de variable : Pop3039

La variable POP3039 est le nombre de femmes et d'hommes âgés entre 30 et 39 ans présents dans la population du Québec en 1996.

Nom de variable : Pop4054

La variable POP4054 est le nombre de femmes et d'hommes âgés entre 40 et 54 ans présents dans la population du Québec en 1996.

Nom de variable : Pop55+

La variable POP55+ est le nombre de femmes et d'hommes âgés de plus de 55 ans présents dans la population du Québec en 1996.

Taux d'incidence 1995-2000 de la maladie de Crohn par groupes d'âge

Nom de variable : Taux019

La variable TAUX019 est le taux d'incidence de la maladie de Crohn pour les femmes et les hommes âgés de moins de 19 ans par 10 000 habitants. Calculé à partir des variables CAS019 et POP019.

Nom de variable : Taux2029

La variable TAUX2029 est le taux d'incidence de la maladie de Crohn pour les femmes et les hommes âgés entre 20 et 29 ans par 10 000 habitants. Calculé à partir des variables CAS2029 et POP2029.

Nom de variable : Taux3039

La variable TAUX3039 est le taux d'incidence de la maladie de Crohn pour les femmes et les hommes âgés entre 30 et 39 ans par 10 000 habitants. Calculé à partir des variables CAS3039 et POP3039.

Nom de variable : Taux4054

La variable TAUX4054 est le taux d'incidence de la maladie de Crohn pour les femmes et les hommes âgés entre 40 et 54 ans par 10 000 habitants. Calculé à partir des variables CAS4054 et POP4054.

Nom de variable : Taux55+

La variable TAUX55+ est le taux d'incidence de la maladie de Crohn pour les femmes et les hommes âgés de plus de 55 ans par 10 000 habitants. Calculé à partir des variables CAS55+ et POP55+.

Population juive

Les données de population juive proviennent du recensement de 2001 de Statistique Canada par l'entremise du MSSS. Des tableaux géocodés ont été achetés par le MSSS. Les statistiques ont été calculées selon les territoires de CLSC de 2004.

Nom de variable : POP01

La variable POP01 est la population totale québécoise lors de recensement de 2001.

Nom de variable : POPJUIV

La variable POPJUIV est la population déclarée juive selon la religion au Québec lors du recensement 2001. Les données de religion ont été utilisées puisqu'il a été supposé que dans le cas de la religion juive, celle-ci représente généralement bien l'ethnie.

*Religion : confession religieuse précise ou appartenance à un groupe ou à un organisme religieux, à une secte, à un culte ou à toute autre collectivité ayant adopté une religion ou un système de croyances quelconque.

Population autochtone

Les données de population autochtone proviennent du recensement de 2001 de Statistique Canada par l'entremise du MSSS.

Nom de variable : PPIND

La variable PPIND est la proportion de la population déclarée Indien d'Amérique par rapport à la population totale (non-autochtone + autochtone), en pourcentage de la population totale, lors du recensement 2001, par territoire de CLSC de 2003.

Nom de variable : PPMET

La variable PPMET est la proportion de la population déclarée Métis par rapport à la population totale (non-autochtone + autochtone), en pourcentage de la population, lors du recensement 2001, par territoire de CLSC de 2003.

Nom de variable : PPINU

La variable PPINU est la proportion de la population déclarée Inuit par rapport à la population totale (non-autochtone + autochtone), en pourcentage de la population, lors du recensement 2001, par territoire de CLSC de 2003.

Contraceptifs oraux

Nom de variable : CO0405

La variable CO0405 est le nombre total d'unités de contraceptifs oraux par CLSC de septembre 2004 à août 2005.

Nom de variable : CO0506

La variable CO0506 est le nombre total d'unités de contraceptifs oraux par CLSC de septembre 2005 à août 2006.

Nom de variable : CO0406

La variable CO0406 est la moyenne entre les variables CO0405 et CO0506, c'est-à-dire la moyenne entre le nombre total d'unités de contraceptifs oraux de septembre 2004 à août 2005 et le nombre total d'unités de contraceptifs oraux par CLSC de septembre 2005 à août 2006.

Nom de variable : POPF1540

La variable POPF1540 est la population de femmes âgées entre 15 et 40 ans qui sont définie comme étant en âge de procréer, et donc, de consommer la pilule contraceptive.

Nom de variable : COFA

La variable COFA est l'estimé du nombre moyen d'unité de contraceptifs oraux par femme âgées entre 15 et 40 ans par année définie par CO0406/ POPF1540.

Urbanité

Nom de variable : CPU

La variable CPU est la somme de tous les codes postaux urbains selon Poste Canada.

Nom de variable : CPTOT

La variable CPTOT est le nombre total de codes postaux (urbains et ruraux) compris dans chaque CLSC.

Nom de variable : PU

La variable PU est la proportion de codes postaux urbains, c'est-à-dire $PU = CPU / CPTOT$.

Maladies à déclaration obligatoire entériques

Nom de variable : CAMPY

La variable CAMPY est la moyenne des incidences pour la campylobacteriose pour les années 1995 à 2000 calculée à partir des incidences par 100 000 de population de chaque année.

Nom de variable : YERS

La variable YERSI est la moyenne des incidences pour la yersiniose pour les années 1995 à 2000 calculée à partir des incidences par 100 000 de population de chaque année.

Nom de variable : GIAR

La variable GIAR est la moyenne des incidences pour la giardiase pour les années 1995 à 2000 calculée à partir des incidences par 100 000 de population de chaque année.

Nom de variable : SALM

La variable SALM est la moyenne des incidences pour la salmonellose pour les années 1995 à 2000 calculée à partir des incidences par 100 000 de population de chaque année.

Nom de variable : SHIG

La variable SHIG est la moyenne des incidences pour la shigellose pour les années 1995 à 2000 calculée à partir des incidences par 100 000 de population de chaque année.

Tabagisme

Nom de variable : TAUTA

La variable TAUTA est le taux de tabagisme calculés, défini par le nombre de fumeurs par sexe, divisé par la population par sexe, combinés pour les deux sexes.

Annexe 2 : Calculs de la standardisation directe

Calculs de la standardisation directe pour l'âge et le sexe du taux d'incidence de la maladie de Crohn par CLSC pour la période 1995-2000 :

1. Regrouper les cas identifiés entre 1995 et 2000 par CLSC
2. Pour chaque CLSC : regrouper les cas selon le sexe et par groupes d'âge au moment du diagnostic (trois groupes d'âge : 0-19 ans, 20-64 ans, 65 ans et +) (C);
3. Pour chaque CLSC : identifier la population du Québec en 1996 par sexe et groupes d'âge (p) ;
4. Pour obtenir des personnes/années, faire le quotient des cas (C) et de la population du Québec en 1996 (p) multipliée par 6 (période de l'étude = 6ans) ce qui donne le taux d'incidence observé selon le sexe et l'âge pour chaque CLSC) (I);
5. Pour obtenir les diagnostics ajustés : appliquer le taux d'incidence observé à la population-type du Québec en 1996 ($E=IxP$);
6. Pour obtenir le taux d'incidence ajusté : faire la somme des diagnostics ajustés (E) et diviser par la population totale du Québec en 1996 qui représente ainsi la population à risque ($\sum E/ 7\ 138\ 800$).

Annexe 2 – Méthode de calcul du taux d'incidence de la maladie de Crohn par CLSC, standardisée pour l'âge et le sexe

SEXE	AGE AU DX	CUMUL CAS 1995-00 (C)	POP CLSC 1996 (P)	TX D'INCIDENCE PAR ($I=C/P \times 6$)	POP QUÉBEC 1996 (P)	DIAGNOSTICS AJUSTÉS ($E=IXP$)	INCIDENCE STDISEE AGE+SEXE
F	0-19				911 730		
	20-64				2 222 730		
	65+				510 960		
M	0-19				954 625		
	20-64				2 189 010		
	65+				349 745		

Total : 7 138 800 Tot: _____

Annexe 3 : Calculs taux de tabagisme

La méthode de calcul utilisée pour le taux de tabagisme, standardisé pour l'âge et le sexe est décrite ci-dessous :

- Les données de pourcentage de fumeurs par sexe et par groupe d'âge pour les différentes RSS (15 au lieu de 18) pour 2003 ont été fournies.
- Pour certains groupes d'âge, les données n'étaient pas fournies (notées « **n.p.» dans la banque de données originale) pour cause d'instabilité des chiffres. Nous avons décidé que dans ce cas, le pourcentage de fumeurs du groupe d'âge général « 12 ans et + » leur serait attribué. Sur les données utilisées, 23% d'entre elles ont été modifiées.
- La population de femmes et d'hommes de l'année 1996 par CLSC, pour 5 groupes d'âge, a été utilisée.
- La correspondance entre les RSS et les CLSC a été établie: chaque CLSC compris dans un RSS s'est fait attribuer les pourcentages de fumeurs selon le sexe du RSS, pour chaque groupe d'âge.
- Pour chaque CLSC, et pour chaque groupe d'âge, le pourcentage de fumeurs a été appliqué à la population du groupe d'âge afin d'obtenir le nombre de fumeurs estimé par groupe d'âge selon le sexe pour chaque CLSC.
- La somme des individus par sexe a été faite (de 12 à 65 ans et +) pour obtenir un dénominateur; la somme des fumeurs par sexe a été faite pour obtenir le numérateur; le taux de tabagisme par CLSC a été obtenu en divisant le nombre de fumeurs par sexe par la population à risque (X 100,000).

Annexe 3 – Exemple de calcul du taux de tabagisme par CLSC.

RSS	TABAGISME 2003					
	F 10-15	15-20	20-30	H 10-15	15-20	20-30
1.						
2.						
3.	15%	15%	30%			
18.						

Dans la RSS#3, 15% des femmes âgées entre 10-15 ans fument, ..., 30% des femmes âgées entre 20-30 ans fument.

Le CLSC#6 est dans la RSS#3.

Annexe 3 – Exemple de calcul du taux de tabagisme par CLSC (suite)

CLSC	POPULATION 1996		
	F 10-15 ans	15-20 ans	20-30 ans
6.	8, 000	10,000	10,000
166.			

Dans le CLSC#6, il y a 10,000 femmes âgées entre 20-30 ans. Pour savoir combien d'entre elles fument (en théorie), le % retrouvé chez les femmes de cet âge dans la RSS qui englobe le CLSC#6 est appliqué à ce CLSC : 30% de 10,000 = 3,000 fumeuses dans cette catégorie d'âge.

Les fumeuses dans toutes les catégories d'âge sont cumulées et divisées par le nombre de femmes total dans le CLSC#6.

La même chose a été faite pour les hommes.

Le résultat est le taux de tabagisme par sexe pour le CLSC.

Les données utilisées sont : homme, femme, groupes d'âge excluant le groupe 12 ans et + (donc 5 groupes d'âge par sexe) pour les 18 RSS.

Annexe 4 : Légende d'attribution de l'indice de défavorisation – statut socioéconomique

COMBINAISON	SOUS-INDICE	SIGNIFICATION	INDICE
11	1	Très favorisé mat. et soc.	1
12	2	Favorisé mat. et soc.	1
21	2		1
22	2		1
13	3	Intermédiaires	1
23	3		1
31	3		1
32	3		1
33	3		1
41	4	Défavorisé mat.	2
42	4		2
43	4		2
51	5	Très défavorisé mat.	2
52	5		2
53	5		2
14	6	Défavorisé soc.	2
24	6		2
34	6		2
15	7	Très défavorisé soc.	2
25	7		2
35	7		2
44	8	Défavorisé mat. et soc.	3
45	8		3
54	8		3
55	9	Très défavorisé mat. et soc.	3

- où
- 1 : favorisé matériellement et socialement à un niveau intermédiaire
 - 2: défavorisé ou très défavorisé matériellement ou socialement
 - 3: défavorisé et très défavorisé matériellement et socialement

Annexe 5 : Recodification des CLSC selon les modifications faites par le MSSS

Tableau généré par Anne-Marie Lowe

AVANT 2004	CLSC2004 (169)	NMCLSC2004	CLSC2005 (166)	NMCLSC2005
	1101	Rimouski-Neigette	1011	Rimouski-Neigette
	1102	La Mitis	1021	La Mitis
	1103	Matane	1031	Matane
	1105	La Matapédia	1041	La Matapédia
	1301	Les Basques	1051	Les Basques
	1302	Saint-Éleuthère	1061	Témiscouata
	1303	Rivière-du-Loup	1071	Rivière-du-Loup
	1304	Kamouraska	1081	Kamouraska
	1305	Cabano	1061	Témiscouata
	2101	Fjord	2011	Fjord
	2102	Saguenay	2021	Saguenay
	2103	Jonquière	2031	Jonquière
	2106	Chicoutimi	2021	Saguenay
	2202	Domaine-du-Roy	2041	Domaine-du-Roy
	2203	Maria-Chapdelaine	2051	Maria-Chapdelaine
	2204	Lac-Saint-Jean-Est	2061	Lac-Saint-Jean-Est
3501	3000	Portneuf	3011	Portneuf
3505	3101	Laurentien	3021	Laurentien
3508	3102	Sainte-Foy - Sillery	3022	Sainte-Foy - Sillery
3601	3201	Québec - Haute-Ville	3023	Québec - Haute-Ville
3602	3202	Québec - Basse-Ville	3024	Québec - Basse-Ville
3603	3203	Limoilou-Vanier	3025	Limoilou-Vanier
3605	3204	Duburger-Les Saules-Lebourgneuf	3026	Duburger-Les Saules-Lebourgneuf
3506	3300	Loretteville - Val-Bélair	3031	Loretteville - Val-Bélair
3302	3401	Beauport	3032	Beauport
3305	3402	Orléans	3033	Orléans
3301	3500	Charlesbourg	3034	Charlesbourg
3304	3701	Charlevoix-Est	3041	Charlevoix-Est
3306	3702	Charlevoix-Ouest	3042	Charlevoix-Ouest

4101	4401	Haut-Saint-Maurice	4011	Haut-Saint-Maurice
4102	4402	Mékinac	4021	Mékinac
4103	4403	Centre-de-la-Mauricie	4031	Centre-de-la-Mauricie
4301	4404	Maskinongé	4041	Maskinongé
4302	4405	Trois-Rivières	4051	Trois-Rivières
4303	4406	Des Chenaux	4022	Des Chenaux
4305	4407	Cap-de-la-Madeleine	4052	Cap-de-la-Madeleine
4304	4501	Nicolet-Yamaska	4061	Nicolet-Yamaska
4306	4502	Bécancour	4062	Bécancour
4202	4503	Drummond	4071	Drummond
4203	4504	Arthabaska	4081	Arthabaska
4204	4505	De l'Érable	4082	De l'Érable
	5101	Granit	5011	Granit
	5102	Asbestos	5021	Asbestos
	5103	Haut-Saint-François	5031	Haut-Saint-François
	5104	Val Saint-François	5041	Val Saint-François
	5105	Coaticook	5051	Coaticook
	5106	Memphrémagog	5061	Memphrémagog
	5107	Fleurimont - Lennoxville	5071	Sherbrooke
	5108	Sherbrooke	5071	Sherbrooke
	6101	Lac Saint-Louis	6011	Lac Saint-Louis
	6103	Pierrefonds	6012	Pierrefonds
	6104	Dollard-des-Ormeaux	6013	Dollard-des-Ormeaux
	6105	Lachine	6021	Lachine
	6201	Pointe-Saint-Charles	6031	Pointe-Saint-Charles
	6202	Verdun	6032	Verdun
	6204	Saint-Paul	6033	Saint-Paul
	6206	Lasalle	6022	Lasalle
	6301	Rivière-des-Prairies	6041	Rivière-des-Prairies
	6302	Pointe-aux-Trembles	6042	Pointe-aux-Trembles
	6303	Mercier-Est	6043	Mercier-Est
	6304	Mercier-Ouest	6051	Mercier-Ouest
	6305	Hochelaga-Maisonneuve	6052	Hochelaga-Maisonneuve
	6306	Rosemont	6053	Rosemont
	6308	Anjou	6044	Anjou
	6309	Saint-Léonard	6061	Saint-Léonard

	6401	Côte-des-Neiges	6071	Côte-des-Neiges
	6402	Snowdon	6072	Snowdon
	6403	Côte-Saint-Luc	6081	Côte-Saint-Luc
	6404	Mont-Royal	6073	Mont-Royal
	6501	Notre-Dame-de-Grâces - Montréal-Ouest	6082	Notre-Dame-de-Grâces - Montréal-Ouest
	6503	Métro	6074	Métro
	6504	Saint-Louis-du-Parc	6091	Saint-Louis-du-Parc
	6505	Saint-Henri	6034	Saint-Henri
	6601	Montréal-Nord	6131	Montréal-Nord
	6603	Saint-Michel	6062	Saint-Michel
	6605	Ahuntsic	6132	Ahuntsic
	6606	Bordeaux-Cartierville	6111	Bordeaux-Cartierville
	6608	Saint-Laurent	6112	Saint-Laurent
	6701	Montréal - Centre-Sud	6092	Montréal - Centre-Sud
	6702	Plateau-Mont-Royal	6093	Plateau-Mont-Royal
	6704	Parc-Extension	6075	Parc-Extension
	6705	Montréal - Centre-Ville	6094	Montréal - Centre-Ville
	6706	Villeray	6121	Villeray
	6707	Petite Patrie	6122	Petite Patrie
7101	7201	Hull	7011	Hull
7102	7202	Aylmer	7012	Aylmer
7103 et 7109	7300	Gatineau	7013	Gatineau
7107	7400	Pontiac	7021	Pontiac
7106	7500	Les Collines-de- l'Outaouais	7031	Les Collines-de-l'Outaouais
7104	7600	Des Forestiers	7041	Des Forestiers
7105	7701	Vallée-de-la-Lièvre	7051	Vallée-de-la-Lièvre
7110	7702	Petite-Nation	7052	Petite-Nation
	8101	Témiscaming	8011	Témiscaming
	8102	Ville-Marie	8021	Ville-Marie
	8103	Rouyn-Noranda	8031	Rouyn-Noranda
	8104	Abitibi-Ouest	8041	Abitibi-Ouest
	8105	Abitibi	8051	Abitibi
	8106	Vallée-de-l'Or	8061	Vallée-de-l'Or
	9101	Les Escoumins	9011	Les Escoumins
	9102	Forestville	9012	Forestville
	9103	Manicouagan	9021	Manicouagan
	9105	Port-Cartier	9031	Port-Cartier

	9106	Sept-Îles	9041	Sept-Îles
9112	9107	Caniapiscau	9051	Caniapiscau
	9109	Minganie	9061	Minganie
	9110	Basse Côte-Nord	9071	Basse Côte-Nord
	9112	Territoire Naskapi	9081	Territoire Naskapi
	10101	Chibougamau-Chapais	10101	Chibougamau-Chapais
	10102	Lebel-sur-Quévillon	10102	Lebel-sur-Quévillon
	10103	Matagami	10103	Matagami
	10104	Baie-James	10104	Baie-James
	11201	Bonaventure	11011	Bonaventure
	11203	Pabok	11021	Pabok
	11204	Gaspé	11031	Gaspé
	11205	Grande-Vallée	11032	Grande-Vallée
	11206	Îles-de-la-Madeleine	11041	Îles-de-la-Madeleine
	11207	Murdochville	11033	Murdochville
	11208	Denis-Riverin	11051	Denis-Riverin
	11209	Avignon	11012	Avignon
	12101	Lac-Etchemin	12011	Lac-Etchemin
	12102	La Nouvelle-Beauce	12025	La Nouvelle-Beauce
	12103	Beauce-Sartigan	12031	Beauce-Sartigan
	12104	Robert-Cliche	12032	Robert-Cliche
	12105	L'Amiante	12041	L'Amiante
	12401	Desjardins	12021	Desjardins
	12402	Chaudière	12022	Chaudière
	12403	Bellechasse	12023	Bellechasse
	12404	Lotbinière	12024	Lotbinière
12701 et 12703	12702	L'Islet	12051	L'Islet
	12704	Montmagny	12052	Montmagny
	13801	Duvernay	13011	Duvernay
	13803	Chomedey	13012	Chomedey
	13805	Pont-Viau	13013	Pont-Viau
	13807	Sainte-Rose-de-Laval	13014	Sainte-Rose-de-Laval
	14201	D'Autray	14011	D'Autray
	14202	Matawinie	14012	Matawinie
	14203	Joliette	14013	Joliette
	14204	Montcalm	14014	Montcalm
	14205	Les Moulins	14021	Les Moulins
	14206	L'Assomption	14022	L'Assomption

	15101	Deux-Montagnes - Mirabel	15011	Deux-Montagnes - Mirabel
	15102	Thérèse-De Blainville	15031	Thérèse-De Blainville
	15103	Antoine-Labelle	15041	Antoine-Labelle
	15104	La Rivière-du-Nord - Mirabel	15021	La Rivière-du-Nord - Mirabel
	15105	Les Pays-d'en-Haut	15051	Les Pays-d'en-Haut
	15106	Les Laurentides	15061	Les Laurentides
	15107	Argenteuil	15071	Argenteuil
16103	16001	Vaudreuil-Soulanges	16121	Vaudreuil-Soulanges
16102	16002	Haut-Saint-Laurent	16021	Haut-Saint-Laurent
16104	16003	Valleyfield-Beauharnois	16031	Valleyfield-Beauharnois
	16004	Châteauguay-Mercier	16041	Châteauguay-Mercier
16405	16005	Les Jardins-de-Napierville	16042	Les Jardins-de-Napierville
16305	16006	Saint-Constant - La Prairie	16043	Saint-Constant - La Prairie
16304	16007	Brossard - Saint-Lambert	16052	Brossard - Saint-Lambert
16308	16008	Longueuil-Ouest	16061	Longueuil-Ouest
16307	16009	Longueuil-Est	16062	Longueuil-Est
16306	16010	Saint-Hubert	16051	Saint-Hubert
16301	16011	Lajemmerais	16063	Lajemmerais
16406	16012	Saint-Jean-sur-Richelieu - Saint-Luc	16071	Saint-Jean-sur-Richelieu - Saint-Luc
16204	16013	Saint-Bruno - Beloeil - Saint-Hilaire	16081	Saint-Bruno - Beloeil - Saint-Hilaire
16402	16014	Chambly-Carignan-Marieville	16072	Chambly-Carignan-Marieville
16201	16015	Bas Richelieu	16091	Bas Richelieu
16203	16016	Les Maskoutains	16082	Les Maskoutains
16401	16017	Cowansville-Farnham-Bedford	16101	Cowansville-Farnham-Bedford
16206	16018	Granby-Shefford-Bromont	16111	Granby-Shefford-Bromont
16205	16019	Acton	16083	Acton
	17101	Baie D'Hudson	17101	Baie D'Hudson
	17102	Ungava	17102	Ungava

	18101	Territoire Cri	18101	Territoire Cri
--	-------	----------------	-------	----------------

Annexe 6 : Procédure GENMOD de SAS

Référence: SAS Help and documentation

The GENMOD Procedure

The GENMOD procedure fits generalized linear models, as defined by Nelder and Wedderburn (1972). The class of generalized linear models is an extension of traditional linear models that allows the mean of a population to depend on a *linear predictor* through a nonlinear *link function* and allows the response probability distribution to be any member of an exponential family of distributions. Many widely used statistical models are generalized linear models. These include classical linear models with normal errors, logistic and probit models for binary data, and log-linear models for multinomial data. Many other useful statistical models can be formulated as generalized linear models by the selection of an appropriate link function and response probability distribution. Refer to McCullagh and Nelder (1989) for a discussion of statistical modeling using generalized linear models. The books by Aitkin, Anderson, Francis, and Hinde (1989) and Dobson (1990) are also excellent references with many examples of applications of generalized linear models. Firth (1991) provides an overview of generalized linear models.

The analysis of correlated data arising from repeated measurements when the measurements are assumed to be multivariate normal has been studied extensively. However, the normality assumption may not always be reasonable; for example, different methodology must be used in the data analysis when the responses are discrete and correlated. Generalized Estimating Equations (GEEs) provide a practical method with reasonable statistical efficiency to analyze such data.

Liang and Zeger (1986) introduced GEEs as a method of dealing with correlated data when, except for the correlation among responses, the data can be modeled as a generalized linear model. For example, correlated binary and count data in many cases can be modeled in this way.

The GENMOD procedure can fit models to correlated responses by the GEE method. You can use PROC GENMOD to fit models with most of the correlation structures from Liang and Zeger (1986) using GEEs. Refer to Liang and Zeger (1986), Diggle, Liang, and Zeger (1994), and Lipsitz, Fitzmaurice, Orav, and Laird (1994) for more details on GEEs.

The GENMOD procedure fits a generalized linear model to the data by maximum likelihood estimation of the parameter vector β . There is, in general, no closed form solution for the maximum likelihood estimates of the parameters. The GENMOD procedure estimates the parameters of the model numerically through an iterative fitting process. The dispersion parameter ϕ is also estimated by maximum likelihood or, optionally, by the residual deviance or by Pearson's chi-square divided by the degrees of freedom.

Covariances, standard errors, and are computed for the estimated parameters based on the asymptotic normality of maximum likelihood estimators.

A number of popular link functions and probability distributions are available in the GENMOD procedure. The built-in link functions are

- identity: $g(\mu) = \mu$
- logit: $g(\mu) = \log(\mu/(1 - \mu))$
- probit: $g(\mu) = \Phi^{-1}(\mu)$, where Φ is the standard normal cumulative distribution function
- power: $g(\mu) = \begin{cases} \mu^\lambda & \text{if } \lambda \neq 0 \\ \log(\mu) & \text{if } \lambda = 0 \end{cases}$
- log: $g(\mu) = \log(\mu)$
- complementary log-log: $g(\mu) = \log(-\log(1 - \mu))$

The available distributions and associated variance functions are

- normal: $V(\mu) = 1$
- binomial (proportion): $V(\mu) = \mu(1 - \mu)$
- **Poisson**: $V(\mu) = \mu$
- gamma: $V(\mu) = \mu^2$
- inverse Gaussian: $V(\mu) = \mu^3$
- negative binomial: $V(\mu) = \mu + k\mu^2$
- multinomial

The negative binomial is a distribution with an additional parameter k in the variance function. PROC GENMOD estimates k by maximum likelihood, or you can optionally set it to a constant value. Refer to McCullagh and Nelder (1989, Chapter 11), Hilbe (1994), or Lawless (1987) for discussions of the negative binomial distribution.

- estimated values, standard errors, and confidence limits for user-defined contrasts and least-squares means
- ability to create a SAS data set corresponding to most tables displayed by the procedure (see [Table 31.3](#))
- confidence intervals for model parameters based on either the profile likelihood function or asymptotic normality
- syntax similar to that of PROC GLM for the specification of the response and model effects, including interaction terms and automatic coding of classification variables
- **ability to fit GEE models for clustered response data**

Annexe 7 : Correspondance entre RSS et CLSC

Données utilisées lors de la création de la variable tabagisme

En ombrage gris : RSS sans données de tabagisme

CLSC 2005	RSS	NOMRSS
1011	1	Bas-Saint-Laurent
1021	1	Bas-Saint-Laurent
1031	1	Bas-Saint-Laurent
1041	1	Bas-Saint-Laurent
1051	1	Bas-Saint-Laurent
1061	1	Bas-Saint-Laurent
1071	1	Bas-Saint-Laurent
1081	1	Bas-Saint-Laurent
2011	2	Saguenay-Lac-Saint-Jean
2021	2	Saguenay-Lac-Saint-Jean
2031	2	Saguenay-Lac-Saint-Jean
2041	2	Saguenay-Lac-Saint-Jean
2051	2	Saguenay-Lac-Saint-Jean
2061	2	Saguenay-Lac-Saint-Jean
3011	3	Capitale-Nationale
3021	3	Capitale-Nationale
3022	3	Capitale-Nationale
3023	3	Capitale-Nationale
3024	3	Capitale-Nationale
3025	3	Capitale-Nationale
3026	3	Capitale-Nationale
3031	3	Capitale-Nationale
3032	3	Capitale-Nationale
3033	3	Capitale-Nationale
3034	3	Capitale-Nationale
3041	3	Capitale-Nationale
3042	3	Capitale-Nationale
4011	4	Mauricie et Centre-du-Québec
4021	4	Mauricie et Centre-du-Québec
4022	4	Mauricie et Centre-du-Québec
4031	4	Mauricie et Centre-du-Québec
4041	4	Mauricie et Centre-du-Québec
4051	4	Mauricie et Centre-du-Québec
4052	4	Mauricie et Centre-du-Québec
4061	4	Mauricie et Centre-du-Québec
4062	4	Mauricie et Centre-du-Québec
4071	4	Mauricie et Centre-du-Québec
4081	4	Mauricie et Centre-du-Québec
4082	4	Mauricie et Centre-du-Québec
5011	5	Estrie
5021	5	Estrie
5031	5	Estrie

5041	5	Estrie
5051	5	Estrie
5061	5	Estrie
5071	5	Estrie
6011	6	Montreal
6012	6	Montreal
6013	6	Montreal
6021	6	Montreal
6022	6	Montreal
6031	6	Montreal
6032	6	Montreal
6033	6	Montreal
6034	6	Montreal
6041	6	Montreal
6042	6	Montreal
6043	6	Montreal
6044	6	Montreal
6051	6	Montreal
6052	6	Montreal
6053	6	Montreal
6061	6	Montreal
6062	6	Montreal
6071	6	Montreal
6072	6	Montreal
6073	6	Montreal
6074	6	Montreal
6075	6	Montreal
6081	6	Montreal
6082	6	Montreal
6091	6	Montreal
6092	6	Montreal
6093	6	Montreal
6094	6	Montreal
6111	6	Montreal
6112	6	Montreal
6121	6	Montreal
6122	6	Montreal
6131	6	Montreal
6132	6	Montreal
7011	7	Outaouais
7012	7	Outaouais
7013	7	Outaouais
7021	7	Outaouais
7031	7	Outaouais
7041	7	Outaouais
7051	7	Outaouais
7052	7	Outaouais
8011	8	Abitibi-Témiscamingue
8021	8	Abitibi-Témiscamingue

8031	8	Abitibi-Témiscamingue
8041	8	Abitibi-Témiscamingue
8051	8	Abitibi-Témiscamingue
8061	8	Abitibi-Témiscamingue
9011	9	Cote-Nord
9012	9	Cote-Nord
9021	9	Cote-Nord
9031	9	Cote-Nord
9041	9	Cote-Nord
9051	9	Cote-Nord
9061	9	Cote-Nord
9071	9	Cote-Nord
9081	9	Cote-Nord
10101	10	Nord-du-Québec
10102	10	Nord-du-Québec
10103	10	Nord-du-Québec
10104	10	Nord-du-Québec
11011	11	Gaspésie-Iles-de-la-Madeleine
11012	11	Gaspésie-Iles-de-la-Madeleine
11021	11	Gaspésie-Iles-de-la-Madeleine
11031	11	Gaspésie-Iles-de-la-Madeleine
11032	11	Gaspésie-Iles-de-la-Madeleine
11033	11	Gaspésie-Iles-de-la-Madeleine
11041	11	Gaspésie-Iles-de-la-Madeleine
11051	11	Gaspésie-Iles-de-la-Madeleine
12011	12	Chaudière-Appalaches
12021	12	Chaudière-Appalaches
12022	12	Chaudière-Appalaches
12023	12	Chaudière-Appalaches
12024	12	Chaudière-Appalaches
12025	12	Chaudière-Appalaches
12031	12	Chaudière-Appalaches
12032	12	Chaudière-Appalaches
12041	12	Chaudière-Appalaches
12051	12	Chaudière-Appalaches
12052	12	Chaudière-Appalaches
13011	13	Laval
13012	13	Laval
13013	13	Laval
13014	13	Laval
14011	14	Lanaudiere
14012	14	Lanaudiere
14013	14	Lanaudiere
14014	14	Lanaudiere
14021	14	Lanaudiere
14022	14	Lanaudiere
15011	15	Laurentides
15021	15	Laurentides
15031	15	Laurentides

15041	15	Laurentides
15051	15	Laurentides
15061	15	Laurentides
15071	15	Laurentides
16021	16	Monteregie
16031	16	Monteregie
16041	16	Monteregie
16042	16	Monteregie
16043	16	Monteregie
16051	16	Monteregie
16052	16	Monteregie
16061	16	Monteregie
16062	16	Monteregie
16063	16	Monteregie
16071	16	Monteregie
16072	16	Monteregie
16081	16	Monteregie
16082	16	Monteregie
16083	16	Monteregie
16091	16	Monteregie
16101	16	Monteregie
16111	16	Monteregie
16121	16	Monteregie
17101	17	Nunavik
17102	17	Nunavik
18101	18	Terres-Cries-de-la-Baie-James