

Institut national de la recherche scientifique,  
Centre Armand-Frappier Santé Biotechnologie

**Vaccination au bacille Calmette-Guérin (BCG) et développement de maladies  
inflammatoires chroniques d'origine auto-immune et/ou métabolique**

Par

Philippe Corsenac

Thèse par articles présentée pour l'obtention du grade de  
de Philosophiae Doctor (Ph.D.) sur mesure  
en Épidémiologie et Immunologie

**Jury d'évaluation**

Présidente du jury et  
examinatrice interne

Pre Maritza Jaramillo  
Centre Armand-Frappier Santé Biotechnologie,  
Institut national de la recherche scientifique

Examinatrice externe

Pre Caroline Sirois  
Faculté de Pharmacie, Université Laval

Examinatrice externe

Pre Helen Trottier  
École de santé publique, Université de Montréal

Directrice de recherche

Pre Marie-Claude Rousseau  
Centre Armand-Frappier Santé Biotechnologie,  
Institut national de la recherche scientifique

Co-directrices de recherche

Pre Simona Stäger  
Pre Marie-Élise Parent  
Centre Armand-Frappier Santé Biotechnologie,  
Institut national de la recherche scientifique

## Résumé

Contexte :

L'expérimentation animale démontre que la vaccination au bacille Calmette-Guérin (BCG) favorise une stimulation de l'immunité régulatrice et un changement de métabolisme cellulaire capable d'enrayer l'auto-immunité et l'hyperglycémie chronique. Ce vaccin pourrait donc potentiellement prévenir la survenue de sclérose en plaques (SEP) et de diabète. À ce jour, les études épidémiologiques n'ont pas montré d'association. Elles présentent toutefois des limites méthodologiques, ne distinguent pas les formes cliniques de SEP, et seul le diabète de type 1 dans l'enfance a été étudié.

Objectifs :

Ce doctorat comprend trois articles visant respectivement à estimer l'association entre la vaccination au BCG et :

- 1) l'incidence de SEP considérant toutes les formes cliniques, puis la forme récurrente-rémittente ;
- 2) l'incidence de diabète de type 1 à l'adolescence ;
- 3) l'incidence des diabètes de type 1, 2, et auto-immun latent (LADA) à l'âge adulte.

Méthodes :

Une cohorte établie par jumelage de bases de données administratives, dont le registre provincial de vaccination au BCG, a été utilisée. Elle est composée de 400,611 personnes nées au Québec entre 1970 et 1974, suivies pour l'utilisation des services de santé jusqu'en 2014. Des algorithmes validés ont permis d'identifier les cas de SEP et diabète. Les remboursements de médicaments par l'assurance médicaments publique ont été utilisés pour distinguer les formes cliniques.

Résultats :

Un risque accru de SEP a été observé chez les sujets vaccinés au BCG lorsque toutes les formes cliniques étaient considérées, mais aucune association n'a été mise en évidence avec la forme la plus fréquente, la SEP récurrente-rémittente (article 1).

La vaccination au BCG n'était pas associée au risque de diabète de type 1 à l'adolescence (article 2).

Aucune association n'a été observée entre la vaccination au BCG et le diabète de type 1 survenant avant l'âge de 30 ans. Après cet âge, une association inverse a été estimée. Les sujets vaccinés présentaient un risque plus faible de diabète de type 2 à l'âge adulte, comparativement aux non-vaccinés. Pour le diabète LADA, un risque accru, imprécis dû au faible nombre de cas identifiés, a été observé chez les sujets vaccinés (article 3).

Conclusions :

Ces études populationnelles, palliant les limites méthodologiques des études antérieures, montrent des associations entre le BCG et la survenue de SEP et diabètes. Elles mettent en évidence la nécessité de préciser les formes cliniques dans les études futures et de reproduire ces résultats à partir de données ou de registres cliniques.

**Mots Clés** : Association ; Bacille Calmette-Guérin (BCG) ; Diabète de type 1 ; Diabète de type 2 ; Diabète auto-immun latent à l'âge adulte (LADA) ; Données médico-administratives ; Études de cohorte; Incidence ; Sclérose en plaques ; Vaccination.

## **Abstract**

### **Background:**

Animal experiments demonstrated that *Bacillus Calmette-Guerin* (BCG) vaccination promotes the stimulation of regulatory immunity and a change in cell metabolism able to suppress autoimmunity and chronic hyperglycemia. This vaccine may therefore potentially prevent the onset of multiple sclerosis (MS) and diabetes. To date, epidemiological studies have not established associations. However, they had methodological limitations, did not distinguish between clinical forms of MS, and focused only on type 1 diabetes in childhood.

### **Aims:**

This doctorate includes three articles, respectively aiming at estimating the associations between BCG vaccination and:

- 1) the incidence of MS considering all clinical forms, and then the relapsing-remitting phenotype;
- 2) the incidence of type 1 diabetes in adolescence;
- 3) the incidence of type 1, 2, and latent autoimmune diabetes in adults (LADA).

### **Methods:**

A cohort established by probabilistic linkage of administrative health databases, including the provincial BCG vaccination registry, was used. It included 400,611 people born in Quebec between 1970 and 1974, and followed up for their use of health services until 2014. Validated definitions were used to identify incident cases of MS and diabetes. Data on drug reimbursements by the public pharmaceutical insurance were used to distinguish clinical forms.

### **Results:**

An increased risk of MS was observed among BCG vaccinated subjects when all clinical forms were considered, but no association was found with the most common phenotype, relapsing-remitting MS (article 1).

BCG vaccination was not associated with the risk of type 1 diabetes in adolescence (Article 2).

No association was observed between BCG vaccination and type 1 diabetes occurring before 30 years old. After this age, a protective association was estimated. BCG vaccinated subjects had a lower risk of type 2 diabetes in adulthood compared to unvaccinated subjects. For LADA, an increased risk, imprecise due to the small number of identified cases, was observed in vaccinated subjects (article 3).

Conclusions:

These population-based studies, overcoming the methodological limitations of previous studies, establish reliable associations between BCG and the onset of MS and diabetes. They highlight the need to specify clinical forms in future studies and to reproduce these results using clinical data or registers.

**Keywords:** Administrative health data; Association; Bacillus Calmette-Guerin (BCG); Cohort studies; Incidence; Latent autoimmune diabetes in adults (LADA); Multiple sclerosis (MS); Type 1 diabetes; Type 2 diabetes; Vaccination.

# TABLE DES MATIÈRES

Résumé.....	ii
Abstract.....	iv
Liste des tableaux.....	x
Liste des figures.....	xii
Liste des sigles et abréviations.....	xiii
Remerciements.....	xvii
Structure du doctorat sur mesure en épidémiologie et immunologie.....	xix
1 Introduction générale.....	1
2 Contexte.....	4
2.1 Quelques notions d'immunologie et de physiologie.....	4
2.2 La sclérose en plaques.....	6
2.2.1 Épidémiologie.....	6
2.2.2 Étiologie.....	7
2.2.2.1 Les facteurs génétiques et familiaux.....	7
2.2.2.2 Les facteurs infectieux de l'exposome.....	12
2.2.2.3 Les facteurs individuels et environnementaux de l'exposome.....	15
2.2.3 Physiopathologie de la sclérose en plaques.....	22
2.2.3.1 Les phénotypes cliniques.....	22
2.2.3.2 L'origine auto-immune ou virale de la sclérose en plaques.....	24
2.2.3.3 Description des processus immunitaires médiés par les lymphocytes menant à la destruction de la gaine de myéline.....	26
2.2.3.4 Description de la neurodégénérescence axonale.....	30
2.2.4 Les thérapeutiques médicamenteuses.....	32
2.2.4.1 Médicaments immunomodulateurs de première génération et de première intention.....	32
2.2.4.2 Traitements de seconde génération et de deuxième intention qui modifient l'évolution de la maladie.....	32
2.3 Le diabète.....	34
2.3.1 Les phénotypes cliniques.....	34
2.3.2 Épidémiologie.....	35

2.3.3	Physiopathologie des différents diabètes .....	37
2.3.4	Étiologie.....	43
2.3.4.1	Les facteurs familiaux, génétiques .....	43
2.3.4.2	Les facteurs infectieux de l'exposome .....	48
2.3.4.3	Les facteurs individuels et environnementaux de l'exposome .....	49
2.3.5	Les thérapeutiques médicamenteuses .....	59
2.3.5.1	L'insulinothérapie.....	59
2.3.5.2	Les antidiabétiques oraux (ADO).....	59
2.4	Le bacille Calmette-Guérin (BCG) .....	61
2.4.1	Historique.....	61
2.4.1.1	Son développement.....	61
2.4.1.2	Programme de vaccination BCG au Québec .....	62
2.4.1.3	Registre de vaccination au BCG.....	63
2.4.2	Protection contre les tuberculoses.....	65
2.4.3	Les protections hétérologues.....	66
2.4.3.1	Diminution de la mortalité infantile et protection contre les infections 66	
2.4.3.2	Protection contre le cancer .....	68
2.4.3.3	Protection contre l'auto-immunité.....	69
2.4.3.4	Changement de la phosphorylation oxydative vers la glycolyse aérobie 70	
3	Problématique .....	72
4	Objectifs de recherche et hypothèses.....	73
5	Méthodes épidémiologiques .....	75
5.1	Schéma, population d'étude et couplage probabiliste.....	75
5.2	Sources de données .....	76
5.3	Les expositions – vaccination au BCG et âge à la vaccination .....	79
5.4	Définition des maladies d'intérêt : sclérose en plaques et diabète.....	81
5.4.1	Sclérose en plaques .....	81
5.4.2	Les diabètes.....	83
5.5	L'analyse de survie et le modèle de régression proportionnel de Cox.....	87
5.5.1	Le modèle proportionnel de Cox .....	89
5.5.2	Les hypothèses du modèle de Cox et test de proportionnalité.....	90

5.5.3	Le modèle de Cox cause-spécifique .....	92
5.6	Les co-variables et les facteurs de confusion .....	92
5.6.1	Provenance des co-variables .....	92
5.6.2	Principe d'un graphe acyclique dirigé et de la confusion .....	94
5.6.2.1	Notations d'un graphe acyclique dirigé et confusion .....	94
5.6.2.2	Structure causale, origine des graphes acycliques dirigés et application dans l'étude .....	95
5.6.3	Choix des facteurs de confusion .....	96
5.6.4	Interaction avec le sexe .....	96
5.7	Les données manquantes et techniques d'imputation .....	96
5.7.1	Les mécanismes des données manquantes .....	97
5.7.2	Les principes de l'imputation « MICE » .....	98
5.8	Le redressement des échantillons .....	99
5.8.1	Utilisation du score de propension et de l'IPW pour le redressement .....	99
5.9	Résumé du plan d'analyse .....	101
5.10	Éthique .....	102
6	Article 1. Bacillus Calmette-Guerin vaccination and multiple sclerosis: a population-based birth cohort study in Quebec, Canada .....	103
6.1	Mise en contexte .....	105
6.2	Abstract .....	106
6.3	Introduction .....	107
6.4	Methods .....	108
6.5	Results .....	112
6.6	Discussion .....	115
6.7	Conclusions .....	120
6.8	Supporting information .....	133
7	Article 2. Association between Bacillus Calmette-Guerin vaccination and type 1 diabetes in adolescence: a population-based birth cohort study in Quebec, Canada .....	161
7.1	Mise en contexte .....	163
7.2	Abstract .....	164
7.3	Introduction .....	165
7.4	Methods .....	166
7.5	Results .....	169

7.6	Discussion .....	170
7.7	Conclusion.....	174
7.8	Supplementary file .....	180
8	Article 3. Early life Bacillus Calmette-Guerin vaccination and incidence of type 1, type 2, and latent autoimmune diabetes in adulthood.....	184
8.1	Mise en contexte.....	186
8.2	Abstract .....	187
8.3	Introduction .....	188
8.4	Material and Methods.....	189
8.5	Results .....	194
8.6	Discussion .....	195
8.7	Conclusions .....	200
8.8	Supplementary File .....	214
9	Discussion générale .....	241
9.1	Retour sur les résultats principaux .....	241
9.2	Validité interne.....	242
9.2.1	Qualité de l'information sur la vaccination au BCG .....	242
9.2.2	Identification des cas incidents pour les maladies d'intérêt.....	243
9.2.3	Confusion.....	245
9.2.4	Pertes au suivi .....	246
9.2.5	Impact de l'imputation des données .....	247
9.3	Validité externe .....	247
9.3.1	Représentativité de la cohorte.....	247
9.4	Études antérieures .....	248
9.5	Limites de nos trois études.....	249
9.6	Forces de nos trois études.....	252
9.7	Mécanismes immunologiques sous-jacents potentiels.....	254
10	Perspectives futures de recherche.....	257
11	Conclusion.....	258
12	Bibliographie .....	260

## Liste des tableaux

Tableau 2.1: Facteurs génétiques de risque ou de protection de la sclérose en plaques.....	9
Tableau 2.2 : Infections virales influençant l'apparition de la sclérose en plaques et leurs plausibles mécanismes immunologiques .....	13
Tableau 2.3: Facteurs de risque ou de protection individuels et environnementaux soupçonnés d'influencer la survenue de la sclérose en plaques .....	17
Tableau 2.4 : Facteurs de protection individuels et environnementaux, soupçonnés, qui influenceraient la survenue de la sclérose en plaques.....	20
Tableau 2.5: Facteurs familiaux et génétiques de risque ou de protection des diabètes de type 1, 2 et LADA.....	45
Tableau 2.6: Facteurs de risque ou de protection individuels, périnataux et de mode de vie qui influencent les survenues de diabètes de type 1, 2 et LADA .....	53
Tableau 5.1 : Variables utilisées de la Cohorte de naissance québécoise sur l'immunité et la santé et leur provenance en fonction des fichiers sources, Québec, Canada, entre 1970 et 2014.....	78
Tableau 5.2 : Résumé du plan d'analyse .....	101
Table 6.1: Description of the study population according to BCG vaccination status, Quebec Birth Cohort on Immunity and Health, Quebec, Canada.....	125
Table 6.2: Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident multiple sclerosis between January 1 <sup>st</sup> 1983 and December 31 <sup>st</sup> 2014, Quebec Birth Cohort on Immunity and Health (1983-2014), Quebec, Canada....	129
Table 6.3: Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident multiple sclerosis between January 1 <sup>st</sup> 1983 and December 31 <sup>st</sup> 2014, over two periods, Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada. ....	131

Table 7.1: Description of the study population, overall (N= 387,704) and according to BCG vaccination status, Quebec Birth Cohort on Immunity and Health (1985-1993), Quebec, Canada. ....	177
Table 7.2: Crude and adjusted hazard ratios and 95% confidence intervals for the association between BCG vaccination, and age at vaccination, and the onset of incident type 1 diabetes, Quebec Birth Cohort on Immunity and Health (1985-1993), Quebec, Canada.....	179
Table 8.1: Description of the study population, overall (N <sub>1</sub> =396,118) and according to BCG vaccination status, Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada. ....	205
Table 8.2: Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident type 1 and 2 diabetes mellitus, as defined by our pharmaceutical definitions among subjects with drug insurance (N <sub>2</sub> =260,440), Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada. ....	208
Table 8.3: Non-weighted prevalence and incidence of phenotypes of diabetes mellitus (DM), as defined by the Immunology of Diabetes Society and the Japanese Diabetes Society, among subjects with long-term drug insurance (N <sub>3</sub> =160,622), Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada. ....	210
Table 8.4: Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident type 1, type 2 diabetes mellitus (DM) and LADA, as defined by the Immunology of Diabetes Society and the Japanese Diabetes Society, among subjects with long-term drug insurance (N <sub>3</sub> =160,622), Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.....	211

## Liste des figures

Figure 2.1 : Les différents phénotypes cliniques de la sclérose en plaques.....	24
Figure 2.2 : Lymphocytes T CD4+ et CD8+ activés contre les peptides de la gaine de myéline dans le système nerveux central.....	28
Figure 2.3 : Dérèglement de la tolérance périphérique et rôle de la signalisation du récepteur des cellules T (TCR) dans la pathogenèse du diabète de type 1.....	39
Figure 2.4 : Étiologie du diabète LADA, et des diabètes de type 1 et 2.....	42
Figure 2.5 : Photos d'un certificat de vaccination au BCG vierge (conservé sous forme de microfilm 1926-1955 et sous forme papier 1956-1992) et des listes de vaccination en papier (1956-1992).....	64
Figure 5.1 : Périodes de couverture des différents registres et fichiers médico-administratifs et pharmaceutiques utilisés dans la Cohorte de naissance québécoise sur l'immunité et la santé, Québec, 1970-2014.....	77
Figure 6.1: Flowchart of participants' inclusions and exclusions in the analytical subset from the Quebec Birth Cohort on Immunity and Health (1983-2014), Quebec, Canada. ....	123
Figure 6.2: Directed acyclic graph of the association between BCG vaccination and multiple sclerosis. ....	124
Figure 7.1: Flowchart of participants' inclusions and exclusions in the analytical subset from the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada. ....	176
Figure 8.1: Flowchart of participants' inclusions and exclusions in the three analytical subsets from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.....	202
Figure 8.2: Diabetes mellitus (DM) definitions among participants included in the analytical subsets from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada. ....	203

## Liste des sigles et abréviations

ADP, Adénosine diphosphate

ALCAM, *Activated Leukocyte Cell Adhesion Molecule*, Molécule d'adhésion cellulaire des leucocytes activés

ARN, Acide ribonucléique

ATP, Adénosine Triphosphate

BCG, Bacille Calmette-Guérin

Ca<sup>2+</sup>, Ion calcium

CADRISQ, Centre d'accès aux données de recherche de l'Institut de la statistique du Québec

cAMP, *Cyclic Adenosine Monophosphate*, Adénosine monophosphate cyclique

CCDSS, *Canadian Chronic Disease Surveillance System*, Système canadien de surveillance des maladies chroniques

CD, *Cluster of Differentiation*, Molécules de surface cellulaire

CIM-9/10, 9<sup>ème</sup>/10<sup>ème</sup> Classification statistique internationale des maladies et des problèmes de santé connexes

CIQSS, Centre interuniversitaire québécois de statistiques sociales

CMH, Complexe majeur d'histocompatibilité

CPA, Cellule présentatrice d'antigènes

CMV, Cytomégalovirus

CO, Monoxydes de carbone

CTL, Cellules T CD8<sup>+</sup> activées, souvent appelées lymphocytes T cytotoxiques

CTLA4, *T-lymphocyte-associated antigen*

EAE, Encéphalomyélite auto-immune expérimentale

EBV, Epstein-Barr

FIPA, Fichier d'inscription des personnes assurés

GADA, *Glutamic Acid Decarboxylase Antibodies*, Anticorps anti-acide glutamique décarboxylase

GLP-1, Agonistes des récepteurs du peptide-1 apparenté au glucagon

GPCR30, *G Protein-Coupled Receptor 30*, Récepteur couplé à la protéines G 30

GWAS, *Genome Wide Association Study*, Étude d'association à l'échelle du génome

HbA1c, Hémoglobine glycosylée

HERVs, *Human Endogenous Retroviruses*, Rétrovirus endogènes humains

HLA, *Human Leucocyte antigen*

HR, *Hazard Ratio*, rapport des risques instantanés

HTLV-1, *Human Lymphotropic Virus Type 1*, Virus T lymphotrope humain de type 1

IAA, *Insulin AutoAntibodies*, Autoanticorps anti-insuline

IC, Intervalle de confiance

ICA, *Islet cells Antibodies*, Autoanticorps des cellules des îlots de Langerhans

ICAM, *InterCellular Adhesion Molecule*, Molécule d'adhésion intercellulaire

iDPP-4, Inhibiteurs de la dipeptidyl peptidase-4

IDS, *Immunology of Diabetes Society*, Société d'immunologie du diabète

IFN, Interféron

IL, Interleukines

Ig, Immunoglobuline (M ou G précisé)

IPW, *Inverse Probability Weights*, Poids ou pondérations de probabilité inverse

IMC, Indice de masse corporelle

INRS, Institut national de la recherche scientifique

IRM, Imagerie par résonance magnétique

iSGLT-2, Inhibiteurs du cotransporteur sodium-glucose de type 2

JDS, *Japan Diabetes Society*, Société japonaise du diabète

K+, Ion potassium

LADA, *Latent Autoimmune Diabetes in Adults*, Diabète auto-immun de l'adulte

LT, Lymphocyte T, dont le phénotype est CD4+ ou CD8+ (précisé)

LB, Lymphocyte B

MAR, *Missing at random*, Données manquantes de façon aléatoire

MCAR, *Missing completely at random*, Données manquantes de façon complètement aléatoire

Med-Echo, Maintenance et exploitation des données pour l'étude de la clientèle hospitalière

MNAR, *Missing not at random*, Données manquantes de façon non aléatoire

MICE, *Multiple Imputation by Chained Equations*, Imputation multiple par équations chaînées

MODY, *Maturity Onset Diabetes of the Young*, Diabète non insulino-dépendant chez les jeunes

MRR, *Mortality rate ratios*, Ratios de taux de mortalité

MVNI, *MultiVariate Normal Imputation*, Imputation normale multivariée

NA<sup>+</sup>, Ion sodium

NAM, Numéro d'assurance maladie

NO, Monoxydes d'azote

NOD, Non-Obese Diabetic mouse, Souris diabétique non obèse

OR, *Odds Ratio*, Rapport de cotes

PMM, *Predictive Mean Matching*

PPARG, *Peroxisome Proliferator Activated Receptor Gamma*

PTPN22, *Protein tyrosine phosphatase non-receptor type 22*

QBCIH, *Quebec Birth Cohort on Immunity and Health*, Cohorte de naissance québécoise sur l'immunité et la santé

RAMQ, Régie de l'assurance maladie du Québec

RD, *Region of Difference*

Reg, Régulateur ou régulation

ROS, *Reactive Oxygen Species*, Espèces réactives de l'oxygène

RR, Risque relatif

SEP, Sclérose en plaques

SCFAs, *Short chain fatty acids*, acides gras à chaîne courte

SEP-PP, Sclérose en plaques de forme progressive primaire

SEP-SP, Sclérose en plaques de forme progressive secondaire

SEP-RR, Sclérose en plaques de forme récurrente-rémittente

SIPW, *Stabilized Inverse Probability Weights*, Poids ou pondérations de probabilité inverse stabilisée

SNP, *Single Nucleotide Polymorphisms*, Polymorphisme mono-nucléotidique

SO, Solvants organiques

T1D, *Type 1 diabetes*, Diabète de type 1

T2D, *Type 2 diabetes*, Diabète de type 2

TCF7L2, *Transcription Factor 7 Like 2*

TCR, *T-Cell Receptor*, Récepteur des lymphocytes T

TFh, Lymphocyte folliculaire

Th, Lymphocyte de type *helper* dont le phénotype est CD4+ (orientation 1,2 ou Reg précisée)

TLR, *Toll-Like Receptor*, Récepteur analogue au récepteur Toll

TNF, *Tumor necrosis factor*, Facteur de nécrose tumorale

TZD, Thiazolidinediones

UPR, *Unfolded Protein Response*

UV, Ultraviolet

VCAM-1, *Vascular Cell Adhesion Molecule 1*, Molécule d'adhésion cellulaire vasculaire 1

VDR, *Vitamin D Receptor*, Récepteur de vitamine D

VLA-4, Intégrine alpha 4 récepteur de VCAM-1

VZV, *Varicella-Zoster Virus*, Virus de la Varicelle et du Zona

ZnT8, Autoanticorps contre l'isoforme 8 du transporteur de zinc spécifique des îlots Langerhans

## Remerciements

Mes premiers remerciements sont dédiés à mes trois directrices de thèse, Marie-Claude Rousseau, Simona Stäger et Marie-Élise Parent, sans qui ce projet ambitieux en deux sciences, le premier de son genre à l'INRS, n'aurait jamais vu le jour. Je les remercie personnellement pour leur implication, leur sollicitude, et leur patience face à un étudiant âgé parfois désenchanté.

Je remercie la direction de l'INRS pour avoir mis en place ce programme sur mesure en collaboration avec mes directrices de thèse ainsi que les Fonds de Recherche du Québec-Santé qui par l'attribution d'une bourse doctorale m'ont permis de me consacrer à temps plein à la réalisation de ce travail doctoral au cours des cinq dernières années.

Je tiens à remercier l'Institut de la statistique du Québec, la Régie de l'assurance maladie du Québec, et la Commission d'accès à l'information du Québec pour nous avoir donné accès aux données de ce projet. Il est également nécessaire de souligner le travail remarquable des analystes du Centre d'accès aux données de recherche (CADRISQ) hébergés dans les locaux du Centre interuniversitaire québécois de statistiques sociales (CIQSS) de l'université de Montréal, sans qui ce travail aurait été simplement impossible. Leurs abnégations et leurs heures de travail ont été particulièrement importantes dans la divulgation des résultats des trois articles du présent doctorat. Je remercie également le CIQSS pour les bourses de publications octroyées dans le cadre des conférences et de la publication de mes articles.

Je remercie Hughes Richard, pour son travail méticuleux dans la préparation des données et dans les échanges pour élaborer nos définitions médico-administratives. Je remercie Jennifer Yu, collaboratrice de notre équipe, pour son soutien amical ainsi que Daniela Filastre-Celidonio, Joanna Monnot et Françoise Lahrant qui ont transformé et accompagné ma vie.

J'aimerais particulièrement remercier ma mère, défunte durant mon travail doctoral, pour m'avoir transmis la résilience, le goût de l'effort et l'intime conviction de toujours croire en ses rêves. Elle garde une place importante dans mon cœur, mes pensées, et guide chacun de mes pas.

Enfin, mon entière reconnaissance va à mon mari, le Dr Alain Kaban, pour son soutien indéfectible durant ces 22 dernières années. Il a su sacrifier son cadre médical et personnel en Polynésie Française puis en Nouvelle-Calédonie, chers à nos cœurs, pour soutenir ce projet un peu fou de retour aux études à plus de 40 ans, dans un pays qui ne nous était pas familier.

## **Structure du doctorat sur mesure en épidémiologie et immunologie**

Un programme de doctorat sur mesure de l'INRS a pour but de répondre à un besoin de formation différent par rapport à un cursus traditionnel n'offrant qu'une seule science. Le présent programme a été élaboré en deux sciences, épidémiologie et immunologie. Les fondements de l'immunologie ont tout d'abord été apportés par des cours de niveaux maîtrise et doctorat afin d'acquérir un savoir fondamental théorique dans cette discipline. Il est opportun de noter que ces cours n'ont pas été accompagnés de mise en pratique en laboratoire. Des cours en épidémiologie de niveau doctoral ont complété le cursus. Cette base permet une réflexion et une compréhension immuno-épidémiologique guidant les hypothèses de recherche ou de travail, les approches méthodologiques et l'interprétation des résultats. Ce programme complète harmonieusement mes savoir-faire et expériences en tant qu'infirmier et épidémiologiste clinicien ayant préalablement évolué en santé publique dans le champ des maladies chroniques, infectieuses et vectorielles.

# 1 Introduction générale

---

La sclérose en plaques et le diabète sont des maladies complexes car toutes deux sont polymorphiques.

La sclérose en plaques se compose principalement de deux phénotypes prédominants : les formes récurrente-rémittente (85-90%) et progressive primaire (10-15%) (1-5). Le diabète se compose de nombreux phénotypes. Il comprend des formes auto-immunes, soit le diabète à progression lente appelé LADA (*Latent Autoimmune Diabetes in Adults*) et à progression rapide dit de type 1 (6). Il comprend également le diabète de type 2, d'origine métabolique, et le diabète gestationnel (7).

Les prévalences et incidences de la sclérose en plaques (8-10) et des diabètes de type 1 (11-13), de type 2 (14) et LADA (6) sont en constante augmentation dans le monde. Les causes des diabètes auto-immuns et du diabète de type 2 sont partiellement expliquées par la pandémie d'obésité et l'inflammation chronique qui en résulte (15, 16). Néanmoins, l'ensemble des causes de la sclérose en plaques et du diabète demeurent encore incomprises. Elles seraient attribuées à des composantes héréditaires et environnementales.

Pour la sclérose en plaques, la génétique n'explique que 30% des cas (17). Les facteurs environnementaux (18, 19) semblent importants, mais peu sont confirmés. L'étiologie de la sclérose en plaques demeure donc complexe et controversée. L'origine auto-immune reste la plus probable (20-22). Néanmoins, en l'absence de caractérisation d'auto-antigènes qui la déclencheraient en périphérie, certains auteurs suggèrent une origine virale (21). Même au niveau physiopathologique, si tous s'accordent sur une neurodégénérescence secondaire due à une inflammation chronique, les phénotypes principaux de sclérose en plaques récurrente-rémittente et progressive primaire pourraient être des processus inflammatoires partiellement distincts (23).

Pour les diabètes auto-immuns et de type 2, l'histoire familiale de diabète revêt une importance capitale. Les deux formes de diabètes partagent en effet de nombreux facteurs de risque génétiques (6). Néanmoins, seuls 50% des cas de diabète de type 1 s'expliquent par la génétique (24, 25). Les facteurs environnementaux sont donc fortement suspectés, à la fois pour les formes auto-immunes que pour le diabète de type 2. Néanmoins, ces facteurs créent une disparité surprenante entre le diabète de type 1 et le diabète LADA, deux formes cliniques auto-immunes, et le T2D

d'autre part, d'origine métabolique. En effet, aucune preuve formelle n'incrimine, à ce jour, des facteurs environnementaux autre que périnataux dans le risque de diabète de type 1 (26) alors que les diabètes LADA et de type 2 partagent de nombreux facteurs de risque environnementaux (27). La vaccination au bacille Calmette-Guérin (BCG) est utilisée depuis une centaine d'années (28), pour protéger les humains de la tuberculose. En plus de son action antituberculeuse, le vaccin au BCG agirait de façon non spécifique sur l'immunité innée et adaptative (28-39) et sur l'utilisation du glucose dans les cellules immunitaires (37, 39-43). Il pourrait ainsi limiter à la fois l'auto-immunité (28-39) et d'autre part l'hyperglycémie chronique mise en évidence par une réduction de l'hémoglobine glycosylée (HbA1c) (37, 39, 41). Dans les formes cliniques auto-immunes de sclérose en plaques et de diabète, des lymphocytes T et B autoréactifs échappent à la sélection centrale ou périphérique et s'attaquent à des organes spécifiques (la gaine de myéline, les oligodendrocytes puis les neurones pour la sclérose en plaques, les cellules  $\beta$  du pancréas pour les diabètes auto-immuns), sans qu'aucune réponse régulatrice ne puisse les contrebalancer. Or, de nombreuses preuves thérapeutiques ou expérimentales, à la fois chez l'humain et l'animal (37, 39, 44, 45) étayent le fait que le vaccin au BCG peut corriger ce déséquilibre. Il accroît en effet le nombre des cellules immunitaires, innées et adaptatives ainsi que de leurs cytokines associées, ayant des propriétés régulatrices (37, 39, 44, 45). Néanmoins ces résultats sont obtenus à très court terme dans un cadre thérapeutique ou expérimental, tous deux éloignés d'une vaccination conventionnelle au BCG, avec une seule injection en bas âge. Pour le diabète de type 2, une inflammation chronique due à l'obésité (15, 16), une insuffisance d'insuline et une résistance à cette dernière provoquent une hyperglycémie chronique qui permet la survenue de ce phénotype de diabète. Le vaccin au BCG pourrait donc endiguer les survenues de sclérose en plaques et de diabète, toutes formes confondues, en contrebalançant les processus inflammatoires et/ou en favorisant une utilisation accrue du glucose par les cellules. De plus, la vaccination au BCG pourrait avoir un effet différent selon le sexe (46) et l'âge à la vaccination (47, 48). En effet, administré très tôt dans l'enfance, ce vaccin permettrait une meilleure stimulation et une éducation hétérologue des cellules innées (47, 48) que lors d'une vaccination plus tardive. Les hommes et les femmes ont par ailleurs des régulations hormonales différentes qui induisent de fait des réponses immunitaires différentes (49-53).

Peu d'études épidémiologiques ont étudié l'association entre la vaccination au BCG et les survenues de sclérose en plaques (6 études (54-59), regroupées dans une méta-analyse (60)) et de

diabète (sept études, (61-67)). Aucune de ces études n'a rapporté d'association, mais toutes comportaient des limites méthodologiques importantes comme une taille d'échantillon ou une période de suivi toutes deux limitées et/ou l'utilisation de données auto-rapportées pour mesurer l'exposition principale. De plus, aucune ne distinguait l'ensemble des formes cliniques de ces deux maladies. Toutes les études portant sur le diabète s'intéressaient au seul diabète de type 1 survenant dans l'enfance (61, 62, 64-67), sauf une qui l'étudiait jusqu'à 20 ans (63). Les diabètes LADA et de type 2 de l'adulte n'ont donc jamais été étudiés en regard de la vaccination au BCG.

Le présent doctorat s'articule autour de trois objectifs répartis dans trois articles originaux. Le premier article palliera, du moins en partie, les limites méthodologiques des précédentes études portant sur l'association entre la vaccination au BCG et la survenue de sclérose en plaques. Il distinguera de plus les associations selon les formes cliniques de sclérose en plaques, en isolant le phénotype prépondérant, la forme récurrente-rémittente. Le second article palliera les limites méthodologiques des précédentes études portant sur l'association entre la vaccination au BCG et le diabète de type 1. Cette étude circonscrit les cas de diabète durant l'adolescence, isolant le pic d'incidence du diabète de type 1 (14 ans) (68, 69), doublant ainsi la puissance statistique d'une étude antérieure menée au Québec (63) et utilisant de plus une nouvelle définition médico-administrative validée pour la population québécoise (70). Enfin un troisième article consistera à estimer pour la première fois les associations entre la vaccination au BCG et les diabètes de type 1, type 2 et LADA chez l'adulte, entre 22 et 44 ans. Les articles 1 et 3 vont permettre, pour la première fois, d'utiliser les données pharmaco-épidémiologiques québécoises en complément des données médico-administratives pour préciser les phénotypes de sclérose en plaques et de diabète, ainsi que leurs âges de survenue.

La présente thèse est une thèse par articles. Elle débutera tout d'abord par une mise en contexte présentant les revues de la littérature épidémiologiques, immunologiques, étiologiques et physiopathologiques de nos deux maladies d'intérêt. Les éléments mécanistiques du vaccin au BCG pouvant potentiellement expliquer une protection vis-à-vis de ces maladies (notre hypothèse de recherche) seront également abordés. Une partie méthode viendra ensuite dans laquelle seront déclinés les objectifs. Enfin les trois articles seront présentés tels qu'acceptés dans des revues scientifiques (les trois sont déjà publiés). Le dernier chapitre présente la discussion et la conclusion générale de la thèse.

## 2 Contexte

---

Cette recension de la littérature débute par les aspects immunologiques et physiologiques des issues étudiées. Ensuite, l'épidémiologie, les facteurs étiologiques, la physiopathologie et les thérapies médicamenteuses seront décrites pour la sclérose en plaques et le diabète. Finalement, l'histoire de la vaccination au BCG au Québec sera brièvement décrite, ainsi que les mécanismes immunologiques et métaboliques engendrés par ce vaccin et les hypothèses de recherche qui en découlent.

### 2.1 Quelques notions d'immunologie et de physiologie

Les cellules innées sont responsables du déclenchement de notre première réponse à un agent étranger, l'immunité non spécifique. Lors d'un premier contact, les cellules innées signalent le danger immédiat (via des récepteurs non-spécifiques comme les *Toll* et *NOD Like Récepteurs*), et induisent l'inflammation (35) par de premiers messagers chimiques, cytokines et chimiokines, dites pro-inflammatoires.

Ces cellules innées comprennent une myriade de cellules dont certaines sont appelées cellules présentatrices d'antigènes professionnelles (CPA) (ex : macrophages, cellules dendritiques, monocytes, neutrophiles) (71, 72). Ces CPA sont spécialisées dans la présentation antigénique de peptides étrangers au système immunitaire adaptatif composé des lymphocytes CD8+, CD4+ et B naïfs afin de les activer spécifiquement contre cet antigène étranger et d'ainsi induire notre seconde réponse, la réponse adaptative (71, 72).

Il existe aussi des CPA non-professionnelles retrouvées au niveau des organes, comme le cerveau (73) ou le pancréas (74), qui peuvent également initier cette présentation antigénique (exemple : cellules épithéliales ou de l'endothélium vasculaire, cellules  $\beta$  des îlots de Langerhans du pancréas, fibroblastes, cellules gliales).

Les cellules innées disposent d'une mémoire immunologique appelée *trained immunity* (38), distincte de celle des cellules adaptatives lymphocytaires (T et B). Contrairement au dogme préétabli avant Netea et al. 2014 (31), l'immunité innée peut apprendre et mettre en mémoire l'apprentissage fait de précédentes infections et/ou vaccinations (31, 38, 39, 75). Une

reprogrammation épigénétiques de l'ADN dans les lignées myéloïdes mais surtout dans leurs progéniteurs hématopoïétiques permet une pérennité de cet apprentissage qui peut se traduire par une réponse immunitaire améliorée ou au contraire diminuée (tolérance) lors d'une seconde stimulation (31, 38, 39, 75).

Les lymphocytes CD8+, CD4+ et B activés par les CPA, attaquent spécifiquement le corps étranger présenté, de façon cytolytique pour les CD8+ (76) et par des anticorps spécifiques pour les cellules B (différenciées en plasmocytes) (77). Les T *helpers* (Th/CD4+), par leur plasticité phénotypique en fonction du contexte de cytokines, assurent la coordination vers trois réponses spécifiques : 1) cellulaire (Th1, Th17) ; 2) humorale (Th2, TFh) ; et 3) régulatrice ou anti-inflammatoire chargée d'éteindre ou de réduire les deux premières (T Reg, B Reg) (78).

Le système HLA ou *Human Leukocyte Antigen* a été découvert sur les cellules du même nom mais il existe sur l'ensemble des cellules du corps humain (79). Aujourd'hui appelé CMH pour complexe majeur d'histocompatibilité, il permet entre autres au système immunitaire de reconnaître le soi et le non-soi (79, 80). Chez l'humain, il existe 3 grandes familles, les CMH de classe I, II et III. Les deux familles I et II sont spécialisées dans la présentation antigénique (80). Les produits des gènes de classe III incluent des protéines du complément et les facteurs de nécrose tumorale (TNF- $\alpha$  et lymphotoxine- $\alpha$ ) (80, 81).

Le CMH de classe I présente au système immunitaire des peptides du soi afin de l'éduquer à la « tolérance ». Il se compose des molécules HLA-A, -B et -C (82). Néanmoins, lorsqu'une cellule est infectée par un pathogène étranger, elle remplace ce peptide du soi par un peptide de ce pathogène (80). Ce signalment au système immunitaire (LT CD8+) permet ainsi la lyse de la cellule infectée; c'est la voie endogène.

Le CMH de classe II se compose des molécules HLA-DR, -DQ et -DP (82). Il est utilisé par les CPA spécialisées dans la présentation antigénique de peptides étrangers au système immunitaire adaptatif (LT CD4+) afin de les activer spécifiquement contre cet antigène étranger; c'est la voie exogène (80).

Parfois dans un contexte inflammatoire, les CPA présentent des peptides du soi inoffensifs à des lymphocytes autoréactifs ayant échappé à la sélection centrale ou périphérique (83). Ce processus est appelé brèche de tolérance, car il permet l'activation de ces dits lymphocytes autoréactifs, naturellement présents dans nos répertoires lymphocytaires (83). Ils vont dès lors s'attaquer à des

peptides et organes spécifiques du corps humain. Cette auto-immunité est alors responsable du déclenchement de la sclérose en plaques (84) et des diabètes auto-immuns (diabète de type 1 et LADA, partiellement) (85), car de tels lymphocytes sont activés pour éliminer des peptides du soi, ceux de la gaine de myéline pour la sclérose en plaques ou ceux des cellules  $\beta$  des îlots de Langerhans, pour les diabètes auto-immuns.

Les mécanismes inflammatoires initiés par l'immunité innée (86) et amplifiés par l'immunité adaptative (87) permettent à la fois le débalancement de l'homéostasie immunologique (et ainsi la brèche de tolérance) mais aussi la chronicisation des processus auto-immuns (88). Cette inflammation chronique peut ainsi être entretenue, à cause d'agents étrangers (88), mais elle peut également provenir d'autres sources d'inflammation tel que l'obésité (15, 16, 88-90) (exemple : adipokines (91)), le manque de sport (88, 92, 93), l'alimentation (88, 94), etc., c'est-à-dire l'environnement ou l'exposome de l'individu (l'ensemble des facteurs environnementaux auxquels il est exposé) (95). Ces expositions peuvent ainsi majorer ou renforcer des processus inflammatoires métaboliques responsables de l'insulino-résistance présente dans les diabètes de type 2 et LADA, mais également les processus auto-immuns (15).

## **2.2 La sclérose en plaques**

### **2.2.1 Épidémiologie**

La sclérose en plaques est une maladie neurodégénérative chronique dont la prévalence et l'incidence ne cessent d'augmenter dans les pays industrialisés, notamment ceux de l'hémisphère Nord, incluant le Canada (8-10). La prévalence au Canada est parmi les plus élevées au monde, avec des estimations >50 cas pour 100 000 habitants en 1985 à Terre-Neuve et jusqu'à 300 cas pour 100 000 habitants dans la Saskatchewan en 2005 (8, 9, 96). Au début des années 2000, la prévalence globale était estimée à 180 cas/100,000 individus au Québec, alors qu'elle pouvait atteindre jusqu'à 350/100,000 dans les provinces de l'Atlantique et des Prairies (97).

Si les prévalences canadiennes de sclérose en plaques diffèrent grandement selon les provinces (96), c'est également le cas des incidences. Warren et al. ont estimé son incidence en Alberta à 20-25 cas pour 100,000 personnes-années, entre 1990 et 2004 (98). Marrie et al. l'ont établie en Nouvelle-Écosse aux alentours de 5 cas pour 100,000 personnes-années, entre 1990 et 2010 (99).

L'incidence de la sclérose en plaques est plus élevée chez les caucasiens, toute formes cliniques confondues, que chez les ethnies africaine, hispanique ou asiatique (100).

L'incidence de la forme récurrente-rémittente est 2 à 4 fois plus élevée chez les femmes que chez les hommes, alors qu'elle est égale selon le sexe pour la forme progressive primaire (3, 101-103).

## **2.2.2 Étiologie**

Dans cette section portant sur l'étiologie, les connaissances sont résumées dans une série de tableaux de synthèse, précisant le niveau d'évidence épidémiologique et/ou immunologique pour chacun des facteurs ou groupes de facteurs mentionnés. Ce niveau d'évidence a été élaboré à partir de classifications existantes, celle du *World Cancer Research Fund grading system 2007* (104), adaptée à partir d'autres échelles de niveau d'évidence présentées par Burns 2011 dont celles du *Canadian Task Force on the Periodic Health Examination's Levels of Evidence* et *Levels of Evidence from Sackett* (105). Cette échelle est croissante à quatre niveaux comme celle du *World Cancer Research Fund grading system 2007* (104) mais elle reprend l'idée des autres échelles présentées par Burns (105) afin d'y inclure tous les types de devis d'étude (essais randomisés et non randomisés, cas-témoins, cohortes, études transversales, séries temporelles, conférence de consensus) bien menées ou avec des lacunes perçues. Parfois les études immunologiques ne pouvaient répondre aux critères épidémiologiques de ces échelles. Cette classification prend donc en considération le nombre d'études, la méthodologie utilisée, la convergence ou divergence des résultats entre les études, l'existence de revues, revues systématiques ou méta-analyses de la littérature ainsi que la cohérence globale des résultats et la plausibilité biologique.

### **2.2.2.1 Les facteurs génétiques et familiaux**

#### L'histoire familiale de sclérose en plaques et l'épigénétique

La sclérose en plaques est une maladie multifactorielle complexe pour laquelle les facteurs génétiques, familiaux, infectieux et environnementaux s'entremêlent afin de déclencher sa survenue. Les prédispositions génétiques sont nombreuses, mais elles ne suffisent pas à elles seules pour déclencher la maladie (17). Néanmoins, plus les parents transmettent de matériel génétique

prédisposant, plus le risque de sclérose en plaques augmente (106). Le risque de survenue de la maladie dans la descendance est ainsi doublé lorsque deux parents sont atteints au lieu d'un (106). Chez les jumeaux monozygotes, qui partagent un matériel génétique identique à la naissance, des changements épigénétiques peuvent néanmoins modifier le risque de sclérose en plaques (107), montrant ainsi le rôle des facteurs environnementaux face aux prédispositions génétiques.

### Les facteurs génétiques HLA et non-HLA

Les 2 grandes classes de molécules HLA-I et -II revêtent une importance capitale dans la reconnaissance du soi et dans le déclenchement de l'auto-immunité. Les premières études ayant révélé cette importance sont les études sur l'hérédité de la sclérose en plaques, plus communément appelée histoire familiale de la maladie (108). Par la suite, les différences de risque relatif selon l'origine ethnique (109), selon les zones géographiques (110) et enfin les *genome wide association studies* (GWAS) (108) ont fini par dresser une cartographie assez complète des prédispositions génétiques pour la sclérose en plaques sur des gènes appartenant ou non au système CMH.

Les allèles du gène HLA-DR.B1 ont été les premiers incriminés. Leurs porteurs ont entre 1,2 et 4 fois plus de risque de déclencher une sclérose en plaques que ceux qui ne les portent pas (82, 111-118). Ce gène présente plusieurs variantes alléliques dans des populations d'origines ethniques différentes. Des estimations du risque de sclérose en plaques en présence de certaines variantes alléliques de ce gène sont résumées dans le tableau 2.1. Cependant, la génétique n'explique que 30% des cas de sclérose en plaques (17), et à lui seul, ce gène compte pour environ 10 % (82).

De nombreux autres allèles du CMH ont par ailleurs été associés à la survenue de sclérose en plaques (82, 111-118). Les porteurs de l'allèle HLA de classe II DQ.B1 (03:02) ont ainsi 1,4 fois plus de risque d'avoir la maladie par rapport aux témoins ne le portant pas, alors que d'autres variantes alléliques du même gène (04:01, 11:01 et 09:01) protègent leurs porteurs par rapport à des témoins ne les ayant pas (Tableau 2.1).

D'autres facteurs génétiques non localisés sur le bras court du chromosome 6, région codante des protéines du CMH, interviennent dans la susceptibilité à la sclérose en plaques. Ce sont à nouveau des allèles de gène (ex : gène NTA1 (119)) ou tout simplement des polymorphismes mononucléotidiques (SNP) qui altèrent une seule paire de base. Ces polymorphismes sont

nombreux et tous touchent la synthèse de protéines intervenant dans l'immunité, comme par exemple des molécules de co-stimulation (120) pour l'activation des lymphocytes, des chimiokines ou des cytokines. Ces exemples sont présentés dans la seconde partie du tableau 2.1.

**Tableau 2.1: Facteurs génétiques de risque ou de protection de la sclérose en plaques**

Facteurs	Estimation <sup>a</sup> de l'association avec la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>					
<b>Facteurs génétiques liés aux allèles* du CMH (liste non exhaustive d'allèles incluant différentes origines ethniques)</b>									
<b>HLA de classe II</b>	Allèles à risque des gènes DR.B1 ou DQ.B1 entre des patients sclérose en plaques par rapport à des témoins :	Le HLA-II permet au CPA la présentation de peptides étrangers aux lymphocytes. Le complexe peptides + HLA-II pourrait induire un mimétisme moléculaire à des peptides du soi (comme la myéline ou autres peptides du système nerveux central) ou au contraire être protecteur de ce mimétisme pour les allèles HLA-II protecteurs.	(82, 108, 111-118)	+ + + +					
<b>DR.B1</b>									
*15:01					~2 (1,5-2,6) – <b>3,9</b> (2,2-8,7)				
*15:03					<b>1,3</b> (1,1-1,5)				
*04:05					<b>2,2</b> (1,5–3,3) – <b>2,4</b> (1,3-4,4)				
*03:01					<b>1,2</b> (1,1-1,2) – <b>1,6</b> (1,3-1,9)				
*13:03					<b>2,6</b> (2,3-3,0)				
*08:01					<b>1,6</b> (1,4-1,7)				
<b>DQ.B1</b>									
*03:02					<b>1,4</b> (0,9–2,4)				
<b>HLA de classe I</b>					Autres allèles protecteurs des gènes de CMH de classe I ou II (DR.B1) entre des patients sclérose en plaques par rapport à des témoins :	Le HLA-I est porté sur l'ensemble des cellules nucléées, il présente des peptides étrangers aux T CD8+ (également des ligands pour les KIR sur les cellules <i>natural killers</i> ). Interactions plausibles avec des infections antérieures.			
<b>A*02:01</b>									<b>0,7</b> (0,6-0,7)
<b>B*44:02</b>									<b>0,8</b> (0,7-0,8)
<b>B*38:01</b>	<b>0,5</b> (0,4-0,8)								
<b>B*55:01</b>	<b>0,6</b> (0,5-0,7)								
<b>HLA de classe II</b>									
<b>DR.B1</b>									
*04:01	<b>0,3</b> (0,2–0,5)								
*11:01	<b>0,6</b> (0,5–0,8)								
*09:01	<b>0,8</b> (0,3–2,2)								
<b>Facteurs génétiques liés aux polymorphismes mononucléotidiques (SNP) sur les allèles* non-MHC</b>									
<b>gène NAT1</b> (Les variantes NAT1 modifient l'effet de l'exposition à la fumée de tabac)	Association tabagisme et sclérose en plaque : Chez les porteurs rs7388368A qui étaient hétérozygotes (A/C) : ~ <b>1,6</b> (1,1-2,5) <b>-1,8</b> (1,2-2,6)  Chez les individus homozygotes (A/A) : <b>5,2</b> (2,2-12,3)	Le gène NAT1 est un modificateur de l'exposition à la fumée du tabac chez les porteurs de susceptibilité génétique à la sclérose en plaques. Les loci du gène NAT1 codent plusieurs enzymes telle que la N-acétyl-transférase et d'autres enzymes de phase 2. Ces enzymes minimisent les dommages cellulaires causés par les agents toxiques comme les solvants organiques des cigarettes et	(119)	+ + +					

Facteurs	Estimation <sup>a</sup> de l'association avec la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>
		autres expositions, lors d'un processus de détoxification du foie.		
<p><b>Gènes GC</b> (rs12512631, rs4588, rs7041, rs222020, rs1155563, rs2282679 and rs3755967), <b>CYP2R1</b> (rs12794714, rs10741657, rs1562902, rs10766197; not shown, rs2060793), <b>CYP24A1</b> (rs2209314, rs2762939) <b>et VDR</b> (rs7968585)</p>	<p>La concentration de base dans le sérum de [25(OH)D] circulante est associé à la présence de ces trois gènes :</p> <p>Après 1 an en utilisant une régression linéaire multiple entre les patients recevant une supplémentation en vitamine D3 par rapport à ceux recevant un placebo (les deux groupes sont porteurs de ces SNP) :</p> <p>Ces quatre gènes modifient l'effet de la supplémentation en vitamine D<sub>3</sub> car la concentration de [25(OH)D] a augmenté en moyenne de 6,1 ± 8,9 ng/mL chez les patients supplémentés en D<sub>3</sub>, et diminué de 1,1 ± 8,4 ng/mL chez les patients non supplémentés (groupe placebo).</p>	<p>Les vitamine D<sub>2</sub> (Ergocalciférol) ou D<sub>3</sub> (cholécalférol) sont transformées en une forme circulante la 25-hydroxy vitamine D [25(OH)D] dans le foie par l'enzyme 25-hydroxylase (CYP27A1). La [25(OH)D] est elle-même transformée dans les reins en 1) 1<math>\alpha</math>,25(OH)D<sub>2</sub> = calcitriol (forme active de la vitamine D) par la 1-<math>\alpha</math> hydroxylase (CYP27B1 ou CYP2R1), et 2) 24,25 (OH)D<sub>2</sub> par la 24 hydroxylase (CYP24A1).</p> <p>La protéine de liaison à la vitamine D est une albumine synthétisée par le gène GC</p> <p>VDR est le récepteur nucléaire de la forme active soit le calcitriol.</p> <p>Le niveau de vitamine D influence les réponses immunitaires, le développement du cerveau, le stress oxydatif, la neuroplasticité, la neurotransmission et la neuroprotection</p>	(121, 122)	+++
<p><b>Gène du récepteur <math>\alpha</math> de IL-7 (CD127)</b> (Le marqueur rs6897932 comme allèle C et le polymorphisme IL7R T244I (rs6897932) comme allèle T)</p>	<p>La cytosine (C) en comparaison avec la thymine (T) CT par rapport à TT: <b>1,3</b> (1,1-1,6), CC par rapport à TT : <b>1,6</b> (1,3-2,0),</p> <p>T en comparaison avec C T par rapport à C, TT+CT par rapport à CC, TT par rapport à CT+CC <b>~0,8</b> (0,7-0,9)</p>	<p>L'IL-7 permet à la fois la thymopoïèse et l'homéostasie périphérique des LT CD4+ et LT CD8+ naïfs. Ce SNP est associé à la sclérose en plaques.</p>	(123, 124)	+++

Facteurs	Estimation <sup>a</sup> de l'association avec la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>
<b>Gène de IL-4 rs2243250 (C/T)</b> (pas le récepteur de IL4R)	C en comparaison avec T ~ <b>1,2</b> (1,02–1,5)	L'IL-4 produite par les lymphocytes Th2 permet leur propre développement et l'initiation de l'immunité humorale par les lymphocytes B, contre des agents pathogènes. Le SNP de l'IL4 rs2243250 (C/T) augmente la susceptibilité à la sclérose en plaques en régulant négativement cette réponse	(125)	+++
<b>Gènes de IL-2 et du récepteur α à IL-2 (CD25)</b> (parmi les patients atteints de sclérose en plaques et les témoins appariés selon l'ethnicité)	Pour le génotype mutant rs11594656 chez les patients atteints de sclérose en plaques (AA par rapport à (TA+TT)) : ~ <b>1,6</b> (1,1–2,5)	L'IL-2 est essentielle à la prolifération de tous les sous-ensembles de cellules T et à la survie périphérique de T Reg. Ce SNP augmente la susceptibilité à la sclérose en plaques	(126)	+++
<b>Gène du CD6</b> Polymorphismes (rs12360861G>A) et faibles niveaux d'expression d'ARNm pour CD6	Pour les individus AA au SNP rs12360861 par rapport aux individus GG : ~ <b>0,3</b> (0,1; 0,8)	L'interaction CD6 et ALCAM permet : l'activation, la prolifération et les synapses immunologiques CPA/cellules T. Ces interactions permettent aux leucocytes de traverser la barrière hémato-encéphalique. Les co-expressions de CD6 avec les chimiokines CCR1 et CCR5 augmentent également la transmigration des leucocytes sanguins vers le liquide céphalo-rachidien	(127)	++
<b>Gènes du CD28, CTLA-4, CD80 et CD86</b> (Distribution des polymorphismes chez les patients atteints de sclérose en plaques et les témoins)	<b>CD28</b> : pas d'association rs35593994G > A (G par rapport à A) rs3116496T > C (T par rapport à C)  <b>CTLA-4</b> : pas d'association rs5742909C > T (C par rapport à T) rs231775A > G (A par rapport à G) rs11571302G > T (G par rapport à T)  <b>CD80</b> : <b>rs1599795T &gt; A</b> ~ <b>0,7</b> (0,5-0,9) TA par rapport à TT <b>0,9</b> (0,4-2,1) AA par rapport à TT  <b>CD86</b> : <b>rs2715267T &gt; G</b> ~ <b>1.3</b> (0.9-1.8) TG par rapport à TT <b>1.6</b> (0.99-2.5) GG par rapport à TT	L'interaction entre CD80/CD86 sur les cellules dendritiques et CD28 sur les lymphocytes T permet la co-activation et la coalescence de la transduction dans les lymphocytes T. L'interaction CTLA-4–CD80/CD86 inhibe cette voie. L'équilibre entre CD28 et CTLA4 maintient la tolérance des lymphocytes T.	(120)	++

ALCAM, molécule d'adhésion cellulaire des leucocytes activés; C, la cytosine; Ref, références; SNP, polymorphisme mononucléotidique; T, la thymine

<sup>a</sup> Rapports de cotes, risque relatif ou hazard ratio.

<sup>b</sup> Niveau de preuve, selon le nombre et les méthodologies des études, évalué par P. Corsenac en adaptant la classification du *World Cancer Research Fund grading system 2007* (104), à partir des échelles d'évidence présentées par Burns 2011 dont celles du *Canadian Task Force on the Periodic Health Examination's Levels of Evidence* et *Levels of Evidence from Sackett* (105) :

+ FAIBLE Peu d'études, non répliquées, insuffisantes pour établir une association entre l'exposition et la maladie;

++ INTERMÉDIAIRE Études de cohorte, cas-témoins ou essais cliniques sans randomisation ou sans revue systématique de ces études; lacunes perçues (exemples : suivi insuffisant ou parcellaire; peu d'essais (ou d'études) disponibles ; échantillon non représentatif, faibles tailles; ajustement inadéquat ou incomplet).

+++ ÉLEVÉ Études de cohorte ou cas-témoins bien conçues et de grande envergure (d'une puissance adéquate), ou revue de la littérature ou méta-analyse mais en nombre limité; quelques lacunes et/ou résultats contraires subsistent; ajustement partiel. L'association devrait biologiquement être plausible.

++++ TRÈS ÉLEVÉ Associations cohérentes entre l'exposition et la maladie, avec peu ou pas de résultats contraires, basées sur des études épidémiologiques nombreuses et des revues systématiques et méta-analyses de ces nombreuses études (exemples : études observationnelles prospectives; Essais contrôlés randomisés avec une randomisation appropriée, une taille, une durée et une qualité suffisantes montrant des effets constants). L'association est biologiquement plausible.

\*Les allèles des antigènes de surface du complexe majeur d'histocompatibilité (CMH) sont également appelés *Human leukocyte antigen* (HLA) chez l'humain, car ces antigènes ont été découverts sur les leucocytes, deux classes de HLA sont particulièrement étudiées : HLA-Classe-I (sous-types: A, B, C) et HLA-Classe-II (sous-types: DR, DQ, DP). Le génotypage permet d'établir les allèles à risque et leurs fréquences chez des patients atteints par rapport aux allèles sains (ou protecteurs) et leur fréquence chez des sujets sains selon l'approche gène candidat. Les estimations de ces facteurs de risque ou de protection sont pour des présentations hétérozygotes. Ces associations sont les plus faibles rapportées. Certaines études d'association à l'échelle du génome (Génome Wide Association Studies, GWAS) peuvent rapporter des mesures d'association plus importantes.

### 2.2.2.2 Les facteurs infectieux de l'exposome

La notion d'exposome a été introduite par Wild en 2005 (95). Il s'agit de l'ensemble des expositions d'un sujet donné, de la naissance (et même *in-utero*) à la mort. Wild le subdivise en 3 composantes : interne, externe spécifique et externe générale (95). La composante interne comprend tous les processus métaboliques et physiologiques du sujet tel que l'homéostasie hormonale, l'avancée en âge, la sédentarité, la prise de poids et les processus inflammatoires et d'oxydation qui en résultent (95). La composante externe spécifique comprend entre autres les expositions à des agents polluants, chimiques, radioactifs ou infectieux, l'alimentation et la consommation d'alcool et de tabac (95). Enfin la composante externe générale recouvre les conditions sociales, économiques et psychologiques de l'individu (95).

Différentes primo-infections ont été successivement suspectées comme influençant la pathogénèse de la sclérose en plaques comme les virus Epstein-Barr (EBV), cytomégalovirus (CMV), et autres herpès virus dont le virus de la Varicelle et du Zona (*varicella-zoster virus*, VZV), et même les rétrovirus endogènes (HERVs) (128). Les risques ou la protection qu'induisent certaines de ces

primo-infections ainsi que les niveaux de preuves et la littérature en référence sont résumés au tableau 2.2.

A ce jour, seulement deux virus ont été fortement associés à la survenue de sclérose en plaques. L'infection symptomatique à l'EBV, la mononucléose, ou sa simple séropositivité augmentent le risque de sclérose en plaques par 2 et par 4, respectivement (129-131). Néanmoins, ce virus infecte très couramment la population puisque 90 à 100% des gens ayant ou non la sclérose en plaques, ont fait une mononucléose ou sa forme asymptomatique dans l'enfance. Au contraire, l'infection au cytomégalo virus réduirait le risque de sclérose en plaques de moitié environ (132-134).

D'autres virus humains de l'herpès, 6A, 6B et 7 (90, 95, 96), ainsi que le VZV (90, 97), le virus T lymphotrope humain (HTLV-1) (98-100), la rougeole et la rubéole (101) ont été étudiés. Néanmoins, leurs implications dans l'apparition de la sclérose en plaques sont encore débattues.

Encore aujourd'hui, l'origine infectieuse de la sclérose en plaques reste très controversée et non démontrée (21). Elle s'oppose à une origine auto-immune probable mais tout aussi débattue (21) qui impliquerait plus largement les facteurs environnementaux.

**Tableau 2.2 : Infections virales influençant l'apparition de la sclérose en plaques et leurs plausibles mécanismes immunologiques**

Facteurs	Estimation <sup>a</sup> de l'association avec la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>
<b>Mononucléose (Virus Epstein-Barr, EBV)</b>	Mononucléose infectieuse vs séronégativité IgG pour EBV : <b>2,2</b> (2,0–2,4)	<b>Cellules B</b> : EBV persiste dans les cellules B mémoires qui se réactivent à un nouveau contact, traversent la barrière hémato-encéphalique pour sécréter des IgG pathogènes contre les antigènes nucléaires latents ou lytiques de l'EBV mais deviennent aussi autoréactives contre les antigènes du système nerveux central comme la myéline en raison de mimétisme moléculaire.	(129-131, 135)	++++
<b>Séropositivité EBV (Non symptomatique)</b>	Séropositivité IgG pour EBV par rapport à séronégativité : <b>4,5</b> (3,3–6,1)	<b>Cellules T</b> : Grâce à la présentation du complexe CMH + antigène, les humains produisent des voies spécifiques EBV-Th1-CTL. Ces lymphocytes T CD8+ et CD4+ activés par l'EBV augmentent en périphérie, gagnent par le sang et ciblent le système nerveux central.  Les cellules B et T autoréactives facilitent l'amorçage des deuxièmes vagues de cellules immunitaires autoréactives à partir du cerveau.		

Facteurs	Estimation <sup>a</sup> de l'association avec la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>
<b>Cytomégalovirus (CMV)</b>	Séropositivité IgG pour le CMV par rapport à séronégativité IgG : <b>0,42</b> (0,2-0,9) <b>0,49</b> (0,3-0,8) <b>0,73</b> (0,6-0,9)	Réponses pro-inflammatoires initiales contre le CMV, puis réponses anti-inflammatoires (réduction des monocytes, expression anti-inflammatoire et réduction des antigènes de la surface cellulaire du CMH de classe I et II).	(132-134)	+++
<b>Virus HTLV-1 (Human T-cell lymphotropic virus)</b>	Sang de patients atteints de sclérose en plaques par rapport à témoins, avec des séquences de HTLV-1 obtenues par amplification du matériel génétique : <b>Pas d'association</b>	Mécanismes plausibles pour un effet néfaste ou protecteur.  Risque : Installation d'une myélopathie inflammatoire chronique similaire à la sclérose en plaques progressive primaire.  Protection : les cellules T auxiliaires infectées activées produisent de l'IL-2, IL-2R et IL-12, IL-12R, probablement intégrées dans le génome humain.	(136-138)	+
<b>Herpès virus humains (HHV 6A, 6B, 7)</b>	Sang de patients atteints de sclérose en plaques par rapport à témoins avec des séquences de HHV obtenues par amplification du matériel génétique : <b>Pas d'association</b>	Persistance de manière latente dans les monocytes-macrophages ; intégration chromosomique plausible; production de cytokines pro-inflammatoires; réduction du HLA de classe I à la surface des cellules infectées.	(130, 139, 140)	+
<b>Virus varicella-zoster (VZV)</b>	Survenue du zona chez les patients atteints de sclérose en plaques par rapport à témoins : <b>4</b> (2,2-7,1)	<b>Mécanismes putatifs à ce jour non expliqués:</b> Le virus est neurotrope et reste à l'état latent pendant de nombreuses années dans l'organisme. Des anticorps anti-VZV sont détectés dans le liquide céphalorachidien des patients avec une sclérose en plaques. Le mécanisme serait donc proche de celui de l'EBV avec une primo-infection et des réactivations. De plus, l'utilisation des médicaments anti-CD20 réactivent le VZV chez les patients sclérose en plaques et cela exacerbe l'attaque du VZV.	(130, 141)	+

EBV, virus Epstein-Barr; Réf, références

<sup>a</sup> Rapports de cotes, risque relatif ou hazard ratio.

<sup>b</sup> Niveau de preuve, selon le nombre et les méthodologies des études, évalué par P. Corsenac en adaptant la classification du *World Cancer Research Fund grading system 2007* (104), à partir des échelles d'évidence présentées par Burns 2011 dont celles du *Canadian Task Force on the Periodic Health Examination's Levels of Evidence* et *Levels of Evidence from Sackett* (105) :

+ FAIBLE Peu d'études, non répliquées, insuffisantes pour établir une association entre l'exposition et la maladie;

++ INTERMÉDIAIRE Études de cohorte, cas-témoins ou essais cliniques sans randomisation ou sans revue systématique de ces études; lacunes perçues (exemples : suivi insuffisant ou parcellaire; peu d'essais (ou d'études) disponibles ; échantillon non représentatif, faibles tailles; ajustement inadéquat ou incomplet).

+++ ÉLEVÉ Études de cohorte ou cas-témoins bien conçues et de grande envergure (d'une puissance adéquate), ou revue de la littérature ou méta-analyse mais en nombre limité; quelques lacunes et/ou résultats contraires subsistent; ajustement partiel. L'association devrait biologiquement être plausible.

++++ TRÈS ÉLEVÉ Associations cohérentes entre l'exposition et la maladie, avec peu ou pas de résultats contraires, basées sur des études épidémiologiques nombreuses et des revues systématiques et méta-analyses de ces nombreuses études (exemples : études observationnelles prospectives; Essais contrôlés randomisés avec une randomisation appropriée, une taille, une durée et une qualité suffisantes montrant des effets constants). L'association est biologiquement plausible.

### **2.2.2.3 Les facteurs individuels et environnementaux de l'exposome**

La description des facteurs individuels et environnementaux suspectés d'être impliqués dans l'étiologie de la sclérose en plaques sera présentée en deux sections : les facteurs de risque et les facteurs de protection.

#### Les facteurs de risque

Les facteurs de risque individuels et environnementaux, qui influencent la survenue de sclérose en plaques sont résumés au tableau 2.3 en fonction de leurs niveaux de preuves respectifs.

#### *Le sexe, les hormones sexuelles et l'âge*

L'incidence est 2 à 4 fois plus élevée chez les femmes que chez les hommes dans la sclérose en plaques récurrente-rémittente alors qu'elle est égale selon le sexe pour la sclérose en plaques progressive primaire (3, 70-72). Cette différence selon le sexe soulève la question du rôle des hormones sexuelles dans la survenue et l'évolution clinique de cette maladie (49-53). Les hormones comme la testostérone et la progestérone ont un effet anti-inflammatoire, tandis que l'œstradiol a un effet pro-inflammatoire à faible concentration et anti-inflammatoire à forte concentration (52). De plus, les niveaux de synthèse de ces hormones varient considérablement au cours de la vie d'une femme (puberté, grossesse, contraception, ménopause) (49). Il est probable que ces changements hormonaux, et la différenciation immunitaire qui en résulte, influencent l'apparition de la sclérose en plaques et la prépondérance des femmes par rapport aux hommes dans l'incidence de la forme récurrente-rémittente (49-53). Par exemple, à la puberté, des niveaux plus élevés d'androgènes stimulent les cellules T CD8+ et réduisent le rapport T CD4+/CD8+ chez les hommes, tandis que chez les femmes, les œstrogènes ont l'effet inverse, soutenant la survie des lymphocytes T CD4+ (49).

#### *Le surpoids et l'obésité*

Le rôle du surpoids et de l'obésité dans la sclérose en plaques reste encore controversé. Un indice de masse corporelle (IMC) > 27 à l'âge de 20 ans a été associé à un risque plus élevé de sclérose en plaques, mais cette même augmentation de l'IMC à un âge adulte plus avancé n'est pas associée à la survenue de la maladie (19, 89, 142, 143). De plus, une interaction significative a été observée entre l'allèle HLA-DR.B1\*15 et l'obésité au regard du risque de sclérose en plaques (143). L'association d'une obésité et de ce facteur génétique prédisposant chez des sujets augmente par 16 leur risque de la maladie par rapport à des sujets n'ayant ni d'obésité, ni l'allèle HLA-DR.B1\*15 (OR=16,2, IC 95% : 7,5–35,2) (143).

Les mécanismes biologiques via la synthèse des adipokines comme la leptine (142) agissant comme des cytokines pro-inflammatoires chez les personnes obèses, sont hautement pertinents (144). En effet, le tissu adipeux induit, chez l'humain et l'animal, une faible inflammation chronique et un déséquilibre en faveur de l'activation périphérique de lymphocytes Th1, Th17 et B, potentiellement auto-immuns, au détriment de lymphocytes T Reg, protecteurs (142, 144). Ce tissu adipeux participe aussi au dialogue entre le système nerveux central et la réponse immunitaire dans la sclérose en plaques, via la leptine pro-inflammatoire (89).

Les niveaux inférieurs de vitamine D observés chez les personnes obèses pourraient également constituer un mécanisme potentiel sous-jacent à la pathogenèse de la sclérose en plaques (144).

#### *La carence en vitamine D*

La vitamine D est un modulateur potentiel de l'immunité qui peut réguler les réponses inflammatoires des lymphocytes T (145) et B (146) autoréactifs par le récepteur de vitamine D (VDR). Apportée par l'alimentation ou l'exposition au soleil (vitamine D3, cholécalciférol), sa faible concentration sanguine chez l'humain (<50 nM/L) augmente le risque de sclérose en plaques d'environ 1,5 à 2 fois tandis que chaque augmentation de dose d'ultra-violet de 1 000 kJ / m<sup>2</sup> en diminue le risque de 1,5 fois (147-149). La vitamine D peut ainsi diminuer les réponses pro-inflammatoires, éventuellement par le biais d'une réduction de l'IL-2, IFN- $\gamma$  et d'IL-17 (78) et une augmentation de l'activité des réponses régulatrices (productions d'IL-4 et IL-10) (145).

La vitamine D agit également directement sur les cellules dendritiques, les cellules les plus habiles dans la présentation antigéniques (150). Cette vitamine les rendrait tolérogènes et abaisserait ainsi le risque de brèche de tolérance (150).

#### *Le tabagisme actif et passif*

Les fumeurs ont entre 1,5 et 1,8 fois plus de risque que les non-fumeurs de contracter la sclérose en plaques et ce risque grimpe même jusqu'à 3 pour les individus qui subissent le tabagisme passif de plusieurs individus (exemple : le tabagisme familial d'une mère, d'un père, d'une partie de la fratrie et des amis cumule les expositions) (151-153). La fumée du tabac contient de nombreux composés issus de la combustion comme les monoxydes d'azote (NO) et de carbone (CO) (154) ainsi que des solvants organiques potentiellement oxydants et cancérigènes (155). Les niveaux élevés de NO peuvent causer des dommages mitochondriaux, une nécrose des oligodendrocytes et une dégénérescence axonale (154). L'exposition répétée au CO affecte également l'oxygénation des tissus et peut notamment entraîner une démyélinisation, et une neurotoxicité (154). Cette fumée et les composants chimiques qu'elle contient peuvent également causer des lésions tissulaires répétitives et une inflammation chronique des voies respiratoires (154). Cela pourrait favoriser la production d'IL-17, via le récepteur *Aryl hydrocarbon* sur les cellules Th17, soutenant davantage cette inflammation chronique (152).

#### *Les catécholamines, mélatonine et interactions*

L'équilibre entre les réponses inflammatoires et anti-inflammatoires par la régulation des cytokines est au cœur de l'étiologie de la sclérose en plaques. Par conséquent, ce même équilibre peut également dépendre de l'homéostasie endocrinienne et des neurotransmetteurs tels que les catécholamines (156) ayant également des fonctions de cytokines. Il existe un haut niveau de preuve que la dérégulation du rythme circadien (hormone mélatonine) par le travail de nuit augmente le risque de sclérose en plaques (157, 158) alors que le stress (50, 159, 160), également générateur de catécholamines, ne semble pas être associé avec à ce même risque de sclérose en plaques.

#### *Les autres facteurs de risque*

La saison de naissance (161-163), la consommation de sel (19, 164) ainsi que la pollution de l'air (165-168), potentiels facteurs de risque pour la survenue de sclérose en plaques ont tous été incriminés sans niveau de preuve suffisant à ce jour.

### **Tableau 2.3: Facteurs de risque ou de protection individuels et environnementaux soupçonnés d'influencer la survenue de la sclérose en plaques**

Facteurs	Estimation <sup>a</sup> de risque ou de protection de la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>
<b>Sexe et hormones sexuelles</b> (rapport femmes/hommes dans l'incidence de la sclérose en plaques)	Femmes par rapport à hommes: ~ <b>2-3</b> (-)	Meilleure signalisation TLR des dangers et présentation antigénique par CMH-II ; La rééquilibration des <b>œstrogènes</b> entre la voie pro-inflammatoire NF-kB via ER- $\alpha$ (faible concentration d'oestradiol) et la voie anti-inflammatoire via l'acoalescence ER- $\alpha$ /œstrogènes (forte concentration d'oestradiol) permet les signalisations GPCR30 et améliore la voie PD-1 (FOXP3, régulatrice) ; Les œstrogènes modifient la perméabilité de la barrière hémato-encéphalique des cellules dendritiques ; une meilleure hypermutation somatique et recombinaison de classe pour les cellules B; <b>Testostérone</b> : maintient l'activité T Reg et la non-méthylation de FOXP3, inhibe Th1.	(51-53)	++++
<b>Obésité des jeunes adultes</b>	IMC > 27 à l'âge de 20 ans par rapport aux sujets de poids normal avec IMC entre 18,5 et 21 kg/m <sup>2</sup> : ~ <b>1,8</b> (1,2-2,6) – <b>2,2</b> (1,7-3,0)	Synthèse de cytokines pro-inflammatoires via la leptine; Déséquilibre Th1-Th17/T Reg ; Diminution de la biodisponibilité de la vitamine D ; induction du facteur ROR $\gamma$ t (et donc des Th17).	(19, 142, 143)	+++
<b>Obésité de l'adulte</b>	IMC $\geq$ 30 par rapport aux sujets de poids normal : <b>Pas d'association</b>		(19, 89, 142)	++
<b>Faible niveau de vitamine D ou faible exposition aux UV</b>	Vitamine D < 50nM/l par rapport à $\geq$ 50nM/l: ~ <b>1,4</b> (1,2-1,7) – <b>2,2</b> (1,5-3,3)  Par incrément de dose UV de 1 000 kJ/m <sup>2</sup> : <b>0,7</b> (0,5-0,9)  Utilisation auto-déclarée de suppléments de vitamine D entre 13 et 18 ans par rapport à aucun : <b>0,7</b> (0,5-0,9)	La vitamine D inhibe les cytokines pro-inflammatoires via le récepteur de la vitamine D ; Rééquilibre la balance Th1-Th17/T Reg.	(147-149)	++++
<b>Tabac</b>	Fumeurs par rapport à non-fumeurs: ~ <b>1,5</b> (1,4-1,6) – <b>1,8</b> (1,2-2,6)	Dommages créés à l'ADN ; interaction avec le gène NAT1 ; synthèse de cytokines pro-inflammatoires induite par solvants; Déséquilibre Th1-Th17/T Reg via le récepteur <i>Aryl hydrocarbon</i> .	(151-153)	++++
<b>Tabagisme passif</b>	Exposés par rapport à non-exposés au tabagisme passif : ~ <b>1,3</b> (1,1-1,6) – <b>2,9</b> (1,3-6,3)		(151)	++++
<b>Exposition aux solvants organiques (SO)</b>	Exposés par rapport à non-exposés au SO: ~ <b>1,5</b> (1,03-2,3) - <b>9</b> (2,5-33,8)	Synthèse de cytokines pro-inflammatoires induite par solvants; Déséquilibre Th1-Th17/T Reg; stress oxydatif; autoanticorps; affecte l'expression des gènes.	(155)	+++

Facteurs	Estimation <sup>a</sup> de risque ou de protection de la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>
<b>Travail de nuit (perturbation du rythme circadien)</b>	Quarts de travail à 15-19 ans, exposés par rapport à jamais exposés : ~ <b>1,1</b> (0,9-1,5) - <b>1,2</b> (0,9-1,4)  Travail par quarts pendant 3 ans ou plus avant l'âge de 20 ans par rapport à jamais de travail par quarts : ~ <b>2,1</b> (1,3-3,4)	Les glucocorticoïdes et le manque de mélatonine augmentent les leucocytes, les cytokines pro-inflammatoires, les molécules d'adhésion; Déséquilibre Th17/T Reg.	(157, 158)	+++
<b>Stress psychologique</b>	Stress auto-rapporté à la maison ou au travail, stress sévère par rapport à stress minimal : <b>Aucune association</b>	Réduction des œstrogènes et de la testostérone ; Déséquilibre Th1-Th17/T Reg.	(50, 159, 160)	+
<b>Saison de naissance</b>	Naissance au mois d'avril par rapport au mois de novembre : ~ <b>1,21</b> (1,08-1,4) - <b>1,24</b> (1,1-1,4)	Faibles concentrations maternelles de vitamine D ; le dérèglement des rythmes circadiens via la mélatonine augmente les cytokines Th1 et perturbe l'action d'ICAM-1 dans les vaisseaux sanguins.	(161-163)	+
<b>Consommation élevée de sel</b>	Consommation excessive par rapport à consommation normale établie par questionnaire sans mesure précise de la quantité ingérée : <b>Pas d'association</b>	Induction de la <i>serum glucocorticoid kinase 1</i> (SGK1), promeut l'expression de Th17.	(19, 164)	+
<b>Pollution de l'air</b>	Les quintiles supérieurs par rapport à inférieurs de particules : <b>Aucune association</b> Pour des particules de taille $\leq 10 \mu\text{m}$ , $\leq 2,5 \mu\text{m}$ , et entre 2,5 et $10 \mu\text{m}$ - Exposé par rapport à non exposé aux gaz d'échappement (NO <sub>2</sub> , O <sub>3</sub> ) <b>Aucune association</b> - Zone urbaine abritant une usine de fer et d'acier par rapport à zone rurale : <b>Variation de la prévalence brute de la sclérose en plaques</b> 96,1 par rapport à 45,5/100 000 habitants (urbaine/rurale)	Perméabilité de la barrière hémato-encéphalique, neuroinflammation et neurodégénérescence.	(165-168)	+

(-), Non disponible

a Rapports de cotes, risque relatif ou hazard ratio.

<sup>b</sup> Niveau de preuve, selon le nombre et les méthodologies des études, évalué par P. Corsenac en adaptant la classification du *World Cancer Research Fund grading system 2007* (104), à partir des échelles d'évidence présentées par Burns 2011 dont celles du *Canadian Task Force on the Periodic Health Examination's Levels of Evidence* et *Levels of Evidence from Sackett* (105) :

+ FAIBLE Peu d'études, non répliquées, insuffisantes pour établir une association entre l'exposition et la maladie;

++ INTERMÉDIAIRE Études de cohorte, cas-témoins ou essais cliniques sans randomisation ou sans revue systématique de ces études; lacunes perçues (exemples : suivi insuffisant ou parcellaire; peu d'essais (ou d'études) disponibles ; échantillon non représentatif, faibles tailles; ajustement inadéquat ou incomplet).

+++ ÉLEVÉ Études de cohorte ou cas-témoins bien conçues et de grande envergure (d'une puissance adéquate), ou revue de la littérature ou méta-analyse mais en nombre limité; quelques lacunes et/ou résultats contraires subsistent; ajustement partiel. L'association devrait biologiquement être plausible.

++++ TRÈS ÉLEVÉ Associations cohérentes entre l'exposition et la maladie, avec peu ou pas de résultats contraires, basées sur des études épidémiologiques nombreuses et des revues systématiques et méta-analyses de ces nombreuses études (exemples : études observationnelles prospectives; Essais contrôlés randomisés avec une randomisation appropriée, une taille, une durée et une qualité suffisantes montrant des effets constants). L'association est biologiquement plausible.

### Les facteurs protecteurs

De nombreux facteurs potentiellement protecteurs vis-à-vis de la survenue de sclérose en plaques, comme les consommations de tabac à priser, i.e., broyé en fine poudre et consommé par insufflation (151, 169), café (170-172), alcool (173), vitamine A (174, 175) ainsi qu'une alimentation saine (176) et un microbiote en santé (177-179) sont en cours d'études. Ils sont tous résumés au tableau 2.4. Hormis la nicotine (151, 169) pouvant être ingérée dans le tabac à chiquer et non à fumer (puisque le tabagisme reste un facteur de risque), aucun facteur n'atteint le niveau de preuve suffisant, à ce jour, pour être discuté dans ce paragraphe.

**Tableau 2.4 : Facteurs de protection individuels et environnementaux, soupçonnés, qui influenceraient la survenue de la sclérose en plaques**

Facteurs	Estimation <sup>a</sup> de risque ou de protection de la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>
<b>Tabac à priser /nicotine</b>	Consommateurs par rapport à non-consommateurs : ~ <b>0,3</b> (0,1-0,8) – <b>0,8</b> (0,7-0,9)	La nicotine inhiberait la libération de cytokines pro-inflammatoires.	(151, 169)	++

Facteurs	Estimation <sup>a</sup> de risque ou de protection de la sclérose en plaques et intervalle de confiance (IC) à 95%	Mécanismes immunologiques	Réf	Niveau d'évidence <sup>b</sup>
<b>Café /Caféine</b>	Consommation quotidienne par rapport à non : ~0,6 (0,4-0,8) – 0,7 (0,5-0,99) <b>Néanmoins ces associations n'ont pas été reproduites.</b>	Inhibe la voie pro-inflammatoire NF-kB et diminue ainsi les niveaux de TNF- $\alpha$ , IL-2 et IFN- $\gamma$ ; Suppression de l'activation Th1; Favorise l'activité T Reg en augmentant IL-10.	(170-172)	++
<b>Alcool</b>	Consommation modérée ou faible par rapport à forte : <b>aucune association</b>	Anti-inflammatoire grâce à l'augmentation de l'IL-10 et diminution des réponses inflammatoires des monocytes.	(173)	+
<b>Vitamine A</b>	<i>Retinol Binding Protein</i> - un marqueur de substitution pour les niveaux de vitamine A dans le deuxième quintile (niveau normal de la vitamine) par rapport au premier (le plus faible) : ~0,38 (0,2-0,7)	La vitamine A inhibe la transcription des facteurs de la voie Th17 et régule positivement les facteurs de la voie T Reg. Cela réduit l'activité du sous-ensemble Th17 et améliore l'activité T Reg.	(174, 175)	+
<b>Microbiote intestinal et acides gras à chaîne courte (SCFAs = acétate, propionate, et butyrate)</b>	Abondance et type de bactéries commensales du microbiote des patients atteints de sclérose en plaques par rapport à donneurs sains: <b>N.A</b>	Inhibent les voies pro-inflammatoires NF-kB, MAPK/AP1 (SCFA/GPCR); promeuvent T Reg exprimant CD39; réduisent les réponses des neutrophiles, macrophages et cellules dendritiques qui initialisent l'inflammation dans les voies respiratoires, la moelle osseuse (propionate), l'épithélium intestinal et peut-être d'autres tissus ; Réduisent le stress oxydatif; Le butyrate serait un inhibiteur des histones désacétylases.	(177-179)	++
<b>Alimentation saine</b>	Consommation plus élevée de fruits ou de légumes par rapport à la médiane de consommation : ~ <b>0,2</b> (0,09-0,6) – <b>0,4</b> (0,2-0,99)	Le tryptophane alimentaire peut être métabolisé par le microbiote intestinal en agoniste du récepteur <i>Aryl hydrocarbon</i> qui réduirait l'inflammation du système nerveux central; les graisses polyinsaturées et les acides gras oméga-3, le succinate et l'acide kynurénique présents dans les aliments sains agissent comme des <i>SCFAs</i> .	(176)	+

SCFAs, Short Chain Fatty Acids ou acides gras à chaîne courtes; N.A, Non Applicable

a Rapports de cotes, risque relatif ou hazard ratio.

<sup>b</sup> Niveau de preuve, selon le nombre et les méthodologies des études, évalué par P. Corsenac en adaptant la classification du *World Cancer Research Fund grading system 2007* (104), à partir des échelles d'évidence présentées par Burns 2011 dont celles du *Canadian Task Force on the Periodic Health Examination's Levels of Evidence* et *Levels of Evidence from Sackett* (105) :

+ FAIBLE Peu d'études, non répliquées, insuffisantes pour établir une association entre l'exposition et la maladie;

++ INTERMÉDIAIRE Études de cohorte, cas-témoins ou essais cliniques sans randomisation ou sans revue systématique de ces études; lacunes perçues (exemples : suivi insuffisant ou parcellaire; peu d'essais (ou d'études) disponibles ; échantillon non représentatif, faibles tailles; ajustement inadéquat ou incomplet).

+++ ÉLEVÉ Études de cohorte ou cas-témoins bien conçues et de grande envergure (d'une puissance adéquate), ou revue de la littérature ou méta-analyse mais en nombre limité; quelques lacunes et/ou résultats contraires subsistent; ajustement partiel. L'association devrait biologiquement être plausible.

++++ TRÈS ÉLEVÉ Associations cohérentes entre l'exposition et la maladie, avec peu ou pas de résultats contraires, basées sur des études épidémiologiques nombreuses et des revues systématiques et méta-analyses de ces nombreuses études (exemples : études observationnelles prospectives; Essais contrôlés randomisés avec une randomisation appropriée, une taille, une durée et une qualité suffisantes montrant des effets constants). L'association est biologiquement plausible.

### **2.2.3 Physiopathologie de la sclérose en plaques**

La sclérose en plaques est une maladie inflammatoire et neurodégénérative qui se caractérise par la démyélinisation progressive puis la dégénérescence axonale des neurones, induisant un handicap progressif (4, 5). Son origine auto-immune ou virale reste encore discutée mais tous s'accordent sur les processus inflammatoires et la neurodégénérescence axonale secondaire (21).

#### **2.2.3.1 Les phénotypes cliniques**

La sclérose en plaques se compose de deux phénotypes principaux et prédominants : les formes récurrente-rémittente aussi appelée cyclique (85-90%) et progressive primaire (10-15%) (1-5), par lesquelles débute la maladie. Ces deux phénotypes sont dans la grande majorité des cas précédés d'un syndrome clinique isolé, dont les symptômes de démyélinisation sont l'atteinte du nerf optique (se traduisant par une névrite optique), du tronc cérébral, du cervelet, ou d'une myélite (5). Le syndrome clinique isolé est un épisode inflammatoire de démyélinisation observé pour la première fois chez un patient (4). Pour devenir une sclérose en plaques, il faut donc au moins deux poussées inflammatoires de démyélinisation, en l'absence de fièvre ou d'infection, espacées de 30 jours (4).

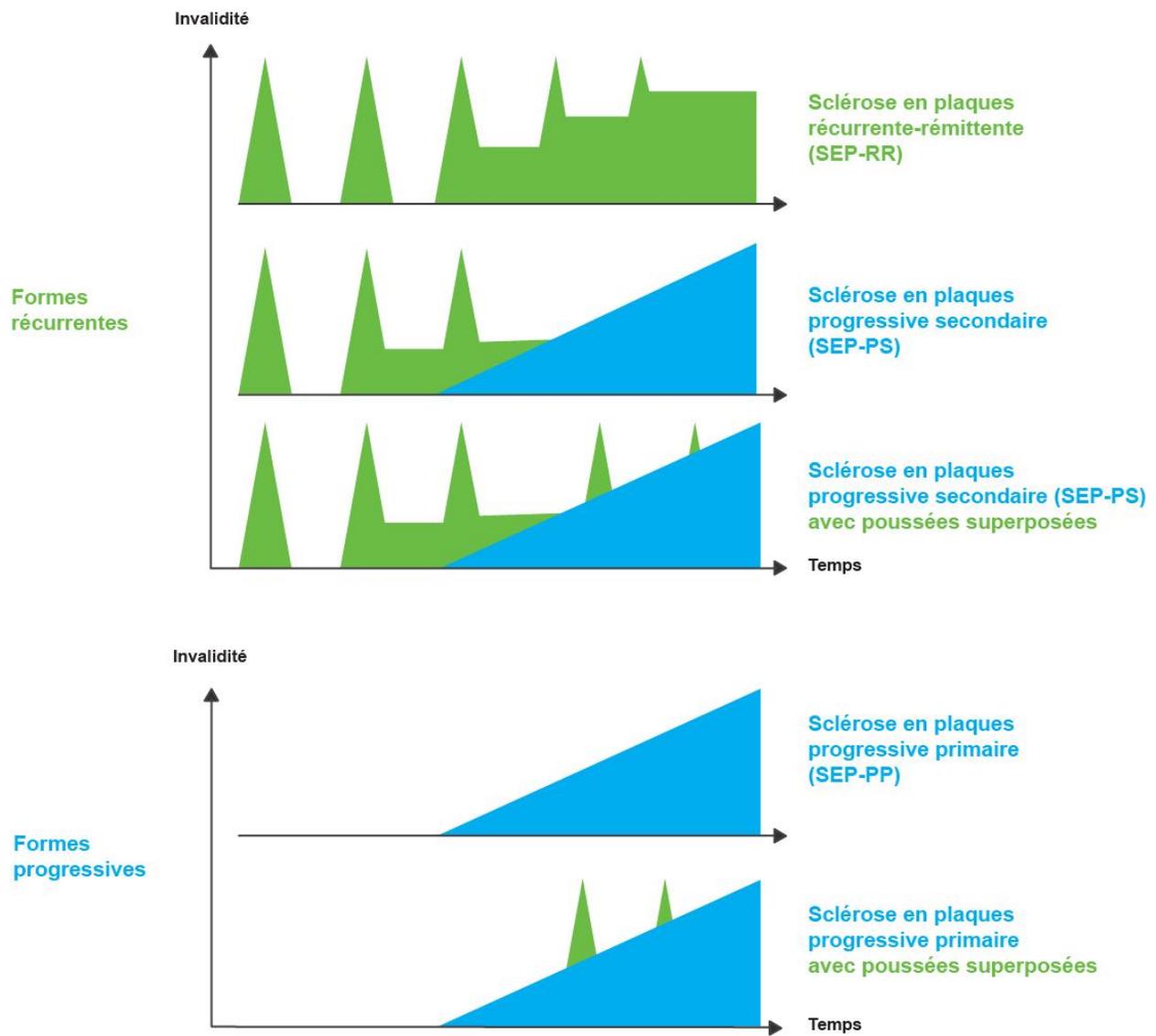
La forme récurrente-rémittente se caractérise par des poussées symptomatiques entraînant un handicap neurologique suivies de rémissions (4, 5, 180). Ce handicap peut totalement disparaître entre deux poussées, ou revenir à son niveau antérieur, en raison des capacités régénératrices des neurones (180). Il peut également progresser, augmentant ainsi successivement au cours des poussées (180). L'âge médian de survenue de cette forme clinique est de 30 ans (2).

Alors que 50% des patients atteints initialement de sclérose en plaques récurrente-rémittente restent sous cette forme, l'autre moitié des patients évolue vers une forme dite progressive

secondaire (sclérose en plaques-PS) (1, 181). Cette forme progressive secondaire survient après 10 ou 20 ans et se caractérise par l'apparition d'un handicap croissant avec ou sans interruption des poussées-rémissions (180). Néanmoins, le handicap ne décroît jamais (180).

La forme progressive primaire (10-15%) ne présente aucune rechute ni rémission, contrairement à la forme récurrente-rémittente, mais un handicap croissant qui ne régresse jamais (180, 181). L'âge médian de survenue de cette forme clinique est de 40 ans (2).

Certains cliniciens distinguent également deux autres phénotypes : la sclérose en plaques progressive secondaire avec rechutes superposées et la sclérose en plaques progressive récurrente, mêlant toutes deux les caractéristiques des deux principaux phénotypes de sclérose en plaques récurrente-rémittente et progressive primaire (180). La Figure 1 illustre les différents phénotypes de sclérose en plaques, selon le type initial (progressif ou récurrent), les poussées et l'accumulation des incapacités.



**Figure 2.1 : Les différents phénotypes cliniques de la sclérose en plaques**

Inspiré de Kamm et al. 2014 (180)

Légende : Les poussées sont illustrées en vert, et l'invalidité croissante en bleu.

### 2.2.3.2 L'origine auto-immune ou virale de la sclérose en plaques

La sclérose en plaques se caractérise anatomiquement par l'apparition de plaques, majoritairement localisées dans la substance blanche du système nerveux central (cerveau, tronc cérébral, cervelet) (182). Ces plaques représentent une concentration de cellules immunitaires innées (monocytes,

macrophages, cellules dendritiques) et adaptatives (lymphocytes T CD8+, T CD4+, B) activées contre les peptides de la gaine de myéline (183). Secondairement, une neurodégénérescence axonale s'installe sur un terrain inflammatoire préexistant (21, 182, 184, 185).

La théorie la plus largement acceptée est donc que la sclérose en plaques est probablement une maladie auto-immune (21) qui commence par une infiltration de LT CD4+ autoréactifs, puis de LT CD8+ et LB autoréactifs qui installent l'inflammation et une première salve de destruction autoréactive provenant de la périphérie (184). Plus tard, la maladie est ensuite dominée par une activation de la microglie, les cellules immunitaires résidentes du système nerveux central, et une neurodégénérescence chronique qui s'opère à partir de celui-ci (186).

De nombreuses preuves étayent l'origine auto-immune. Différents sous-types histopathologiques de lésions démyélinisantes décrivent en effet la présence de ces cellules innées et adaptatives inflammatoires dans les biopsies de patients (182). Des cellules T (CD4+ Th1/Th17 et CD8+) activées se trouvent dans les plaques mais aussi dans le liquide céphalorachidien et la circulation périphérique des patients atteints de sclérose en plaques (84). Ces cellules effectrices inflammatoires périphériques, résistantes aux mécanismes régulateurs, ont la capacité de passer la barrière hémato-encéphalique, d'être réactivées par les CPA locales, de recruter de nouvelles cellules T et macrophages résidents, et d'initier les premières lésions inflammatoires des neurones et oligodendrocytes (84). Dans le modèle animal de sclérose en plaques, l'EAE (*Experimental Autoimmune Encephalomyelitis*), l'activation immunitaire des Th17, médiée en partie par l'expression de l'interleukine 23, est associée à des lésions actives apparentées à la sclérose en plaques (187). Enfin des drogues immuno-modulatrices ont démontré leur action contre l'auto-immunité de la sclérose en plaques : les interférons  $\beta$  1-a et 1-b réduiraient la réponse inflammatoire des cellules innées et des Th1 (188); l'acétate de glatiramère augmente la réponse T régulatrice, rétablissant le déséquilibre Th1, Th17 versus T Reg (188, 189); le natalizumab bloque la migration des lymphocytes T à travers la barrière hémato-encéphalique (188); et l'ocrelizumab neutralise les cellules B autoréactives contre la gaine de myéline (188).

Cependant, s'il existe un ou des auto-antigènes qui activent ces lymphocytes T et B autoréactifs, qui échappent à la sélection (périphérique ou centrale), où sont-ils (21) ? À ce jour, aucun n'a pu être identifié (21) et transféré à des animaux pour induire une véritable sclérose en plaques auto-immune chez l'animal et non un modèle imparfait comme celui de l'EAE (190). En effet, le modèle

EAE n'est pas induit par un auto-antigène caractérisé (21). Il induit une auto-immunité contre le système nerveux central de la souris certes comparable à la sclérose en plaques mais non équivalent puisqu'il utilise des artifices comme la composante de Freund et la toxine de la coqueluche pour forcer l'activation de lymphocytes autoréactifs (Forcer les signaux 1, 2 et 3) contre des peptides de myéline qui leurs sont massivement présentés sous la forme d'une injection. De plus, si la perte de myéline et d'oligodendrocytes (les cellules qui la produisent) pourrait tout à fait s'expliquer par des processus auto-immuns dirigés vers plusieurs de leurs peptides (183), la sclérose en plaques évolue secondairement vers une neurodégénérescence axonale et une atrophie cérébrale qui n'est pas entièrement attribuable à l'inflammation et aux processus auto-immuns (185). Nous parlerons donc plus dans le cas présent de lymphocytes T et B activés et effecteurs que de réels lymphocytes autoréactifs, en absence de preuve (21).

Dès lors, une origine infectieuse pourrait être plausible, même si aucune preuve ne vient à ce jour corroborer ce fait (21). Un ou plusieurs antigènes viraux pourraient activer indirectement l'immunité (mais pas de façon autoréactive) par mimétisme moléculaire. Néanmoins, à ce jour, aucun agent pathogène n'a clairement été identifié comme responsable de la sclérose en plaques (21).

### **2.2.3.3 Description des processus immunitaires médiés par les lymphocytes menant à la destruction de la gaine de myéline**

#### L'implication des différentes familles de lymphocytes

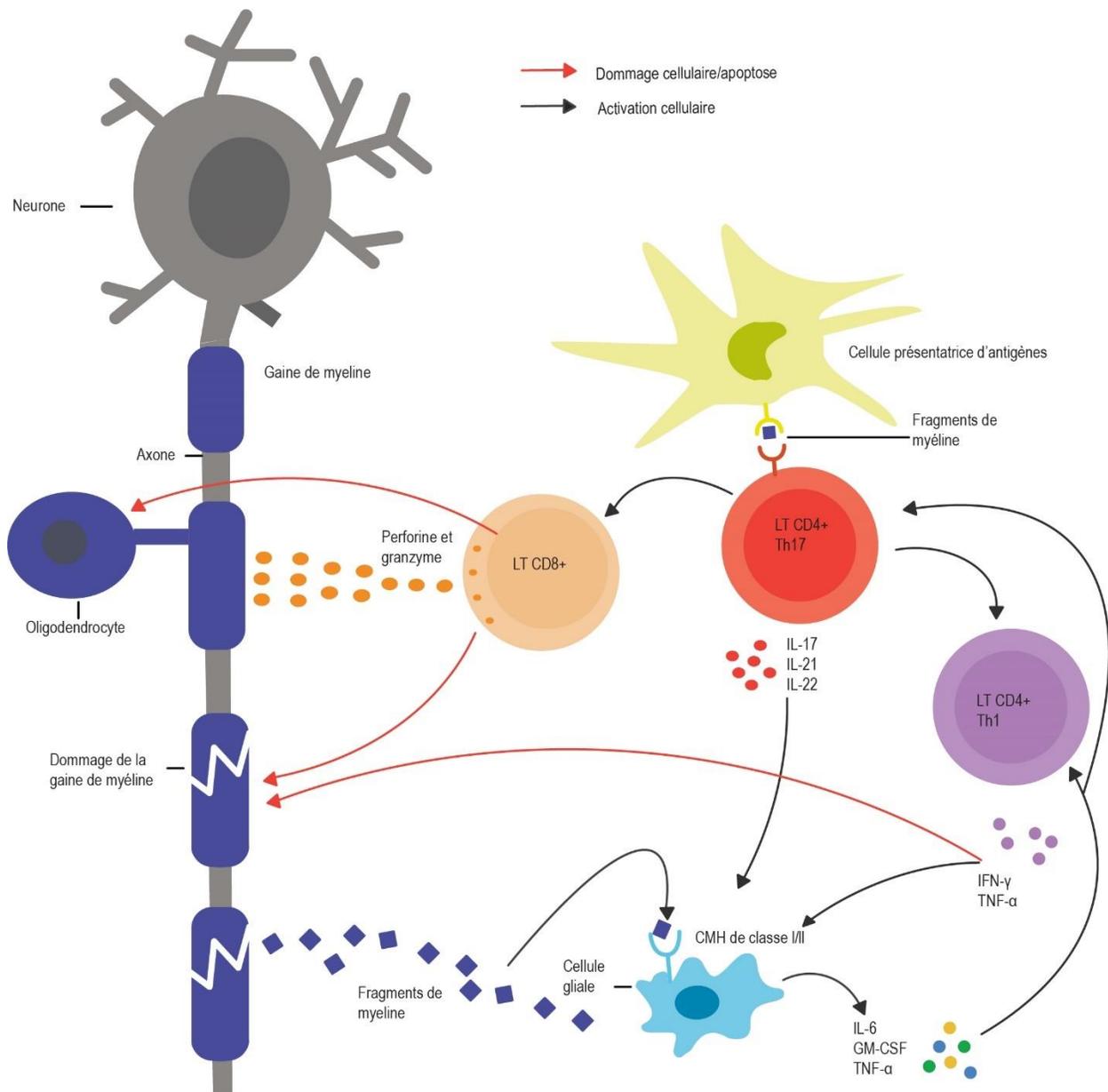
Si l'absence d'identification d'auto-antigènes amorçant les lymphocytes en périphérie ne permet pas tout à fait d'utiliser le terme de maladie auto-immune (21), les différentes familles de lymphocytes, T CD4+, T CD8+ et B participent toutes au premier stade de la pathogénèse de la maladie (184), soit le ciblage et la destruction immunitaire des peptides de la gaine de myéline (21, 183).

Nous avons brièvement abordé dans l'introduction générale, trois orientations différentes dans la réponse immunitaire adaptative : 1) une réponse cellulaire médiée par les LT CD4+, Th1, Th17 et les LT CD8+; 2) une réponse humorale médiée par les T folliculaires (TFH), les Th2 et les cellules B; et 3) une réponse régulatrice médiée par des LT (pouvant être des CD4+ ou CD8+ régulateurs) et LB régulateurs. Par leur plasticité phénotypique, les LT CD4+, appelés également T *helpers*,

orientent comme des chefs d'orchestre dans les trois directions de l'immunité adaptative en fonction de l'attaque (pro-inflammatoire) ou de la régulation (anti-inflammatoire) à mener.

Dans la cadre de la pathogénèse de la sclérose en plaques, les premiers lymphocytes incriminés sont des LT CD4+. Les Th1 et Th17 sont les premiers à infiltrer le système nerveux central (184) en passant la barrière hémato-encéphalique afin d'installer un contexte pro-inflammatoire. Cette barrière est normalement infranchissable par le système immunitaire périphérique. Les Th1 produisent de l'IFN- $\gamma$ , les Th17 de l'IL-17 et les CPA de l'IL-23 qui installent l'inflammation propice à la destruction de la gaine de myéline (84, 184, 187). Leurs nombres augmenteraient au détriment de réponses régulatrices (T Reg), anti-inflammatoires, submergées (184). Cette initiation est supportée par des observations faites chez les patients atteints de sclérose en plaques (84) mais la pathogénèse actuelle reste encore majoritairement expliquées par le modèle animal imparfait de l'EAE (184).

Survient ensuite les lymphocytes T CD8+, faisant partie de la réponse cellulaire. Ils sont normalement spécialisés dans la destruction cytotytique des cellules infectées par un agent pathogène intracellulaire. Néanmoins, activés par un auto-antigène en périphérie (21) et dirigés par les chefs d'orchestres Th1 et Th17, les lymphocytes T CD8+ activés détruisent les peptides de la gaine de myéline et secondairement les oligodendrocytes eux-mêmes. Néanmoins certains LT CD8+ pourraient également produire une réponse régulatrice (191). Ils ne participeraient pas à la destruction mais à la désactivation de lymphocytes T CD4+ qui orchestrent la destruction auto-immune (191).



**Figure 2.2 : Lymphocytes T CD4+ et CD8+ activés contre les peptides de la gaine de myéline dans le système nerveux central.**

Inspiré de Kaskow et al. 2018 (84),

Légende Figure 2 : Les lymphocytes T CD4+ et CD8+ activés établissent puis maintiennent un environnement inflammatoire dans le système nerveux central contribuant à la mort des oligodendrocytes, à la démyélinisation et à la perte neuronale.

Les lymphocytes B font partie de la réponse humorale. La cellule B a la particularité de se différencier en plasmocyte (184), après activation, pour produire des anticorps (ou immunoglobulines (Ig)) qui ciblent les protéines contenant le peptide l'ayant activée et détruisent

ainsi les cellules ou les agents pathogènes incriminés par une cytotoxicité cellulaire dépendante des anticorps. Or des IgM (précoces) et IgG (tardifs) ciblant des auto-protéines sont retrouvés dans le liquide céphalo-rachidien chez respectivement 90% et 40% des patients atteints de sclérose en plaques, en fonction de l'activité de la maladie (184). Ces anticorps incriminent directement les cellules B dans la première salve d'attaque. À l'intérieur du système nerveux central, ils ciblent et détruisent la gaine de myéline et les cellules associées, les oligodendrocytes (184, 192). Les cellules B sont aussi de super CPA professionnelles, capables de présenter des antigènes aux lymphocytes T et d'induire leur différenciation phénotypique, par la synthèse de cytokines. Certaines cellules B, résidentes du système nerveux central, peuvent donc : 1) présenter des débris cellulaires de la gaine de myéline et des oligodendrocytes et chroniciser ainsi l'attaque de Th1, Th17 et LT CD8+, de novo; ou 2) orchestrer une réponse régulatrice, anti-inflammatoire, protectrice (184, 192). Néanmoins, ce second point reste encore à démontrer chez l'humain (184, 192).

Les derniers lymphocytes incriminés dans cette pathogénèse font partie de l'immunité adaptative humorale (193). Ce sont des lymphocytes T CD4+, les T auxiliaires folliculaires (193). Ces derniers doivent nécessairement participer à l'activation « autoréactive » des cellules B, soit au niveau des ganglions lymphatiques, soit en passant eux-mêmes la barrière hémato-encéphalique (193), en même temps que les premiers Th1 et Th17, car les processus de maturation des cellules B et d'affinement des anticorps nécessitent le soutien de ces dernières (193).

#### La migration et les désordres de la barrière hémato-encéphalique

La barrière hémato-encéphalique se compose de l'endothélium vasculaire, de sa membrane basale, de cellules péricytes (capables de mobilité) et des astrocytes (194). L'endothélium vasculaire au niveau du système nerveux central a la particularité d'être composé de cellules très serrées par un ensemble de protéines appelées « *tight junction* » qui empêche les éléments figurés du sang, érythrocytes ou leucocytes, ainsi que les grosses protéines comme l'albumine ou les immunoglobulines de diffuser (194).

Dans la sclérose en plaques, lors de l'activation (*priming*) des différentes familles de lymphocytes en périphérie, les concentrations de cytokines pro-inflammatoires augmentent dans la circulation générale. Ces cytokines provoquent la perte d'intégrité de la barrière hémato-encéphalique en relâchant totalement ou modulant les *tight junctions* (194). Ces conditions inflammatoires vont

également permettre d'exprimer différentes molécules d'adhésion au niveau de la lumière endothéliale, par exemple VCAM-1 (195), qui vont faciliter la diapédèse des lymphocytes activés. Parallèlement, les lymphocytes activés vont développer des molécules d'adressage ou *homing proteins* comme les intégrines  $\alpha 4$  (par exemple, VLA4) à leur surface (188, 195). En regard des molécules d'adhésion exprimées par un endothélium inflammatoire, celles-ci permettront la migration de ces lymphocytes activés au travers d'une barrière hémato-encéphalique devenue également perméable par la modulation des *tight junctions* (194).

#### **2.2.3.4 Description de la neurodégénérescence axonale**

La première salve d'inflammation et d'attaque « auto-immune » du système nerveux central provient de la périphérie induisant les premiers dommages sur la gaine de myéline et les axones (185). Les débris cellulaires de cette première attaque permettent l'amorçage d'une seconde salve d'attaque provenant de la microglie, les cellules immunitaires résidente du système nerveux central (185). La neurodégénérescence axonale, seconde phase de la sclérose en plaques, est partiellement distincte de ces deux types d'attaque mais elle nécessite les désordres initiaux de ces attaques pour se mettre en place.

Les oligodendrocytes fabriquent la gaine de myéline qui protège les axones et permet la propagation de l'influx nerveux, le potentiel d'action, via les nœuds de Ranvier (185). Leur destruction, par la destruction de la gaine de myéline, ne permet donc plus des échanges ioniques harmonieux de  $Na^+$ ,  $K^+$  et  $Ca^{2+}$  le long des axones et empêche donc la propagation du signal (185). Les neurones demandent alors plus d'énergie à leurs mitochondries pour maintenir ces répartitions ioniques de part et d'autre de la membrane axonale (185). Or, ces mitochondries sont sensibles à des neurotoxiques, les espèces réactives de l'oxygène et de l'azote, produites par la première et la seconde salves d'inflammation et d'attaques autoréactives (185). Les dysfonctions mitochondriales et l'activation de la microglie pérennisent une neuroinflammation et une neurodégénérescence localisées via les cytokines pro-inflammatoires et l'activation de LT  $CD4^+$  résidents qui accentuent le phénomène (185).

Par la suite les astrocytes jouent un rôle central (185): 1) ils changent la perméabilité membranaire de la barrière hémato-encéphalique et permettent le passage de nouveaux lymphocytes T et B

activés contre les peptides de la gaine de myéline (accentuation des salves périphériques du premier type) ; 2) ils produisent des chimiokines qui attirent ces cellules périphériques y compris des macrophages de l'immunité innée; 3) ils augmentent l'expression du CMH de classe II et donc la présentation antigénique des débris cellulaires de myéline et d'oligodendrocytes (salves de second type); et 4) ils augmentent l'activité de phagocytose de la microglie.

## **2.2.4 Les thérapeutiques médicamenteuses**

La plupart des médicaments décrits dans cette section sont encore aujourd'hui exclusivement prescrits à des patients atteints de la forme récurrente-rémittente de la sclérose en plaques. L'avènement des anti-CD20 en 2017, changea la donne pour la première fois, avec l'introduction de médicaments qui agissent également sur la forme progressive primaire.

### **2.2.4.1 Médicaments immunomodulateurs de première génération et de première intention**

Acétate de glatiramère : Il agit à la fois sur l'immunité innée et adaptative (189). Il permet aux CPA une différenciation tolérogène et augmente les réponses T régulatrices et Th2 (modulant l'activation de cellules B), inhibe les lymphocytes T activés contre les peptides de la gaine de myéline et diminue la sécrétion de l'interféron- $\gamma$  (pro-inflammatoire) (189). Ainsi, il rétablit le déséquilibre Th1, Th17 versus T Reg (189).

Interférons  $\beta$  1a et 1b : Les modes d'action sont mal connus (188). Néanmoins, tous deux auraient une action anti-inflammatoire et neuroprotectrice (188). Sur les cellules de l'immunité innée, ils réduiraient la production de cytokines pro-inflammatoires et permettraient, par exemple une activation de la voie Th2 anti-inflammatoire ( $\beta$  1-a) au niveau de l'immunité adaptative (188).

### **2.2.4.2 Traitements de seconde génération et de deuxième intention qui modifient l'évolution de la maladie**

Fumarate de diméthyle: Ce médicament a un mécanisme d'action inconnu (188). Néanmoins, il activerait supposément le facteur nucléaire 2 similaire au dérivé d'érythroïde (Nrf2), qui réduirait le stress oxydatif (188). Il serait, toujours supposément, anti-inflammatoire par la réduction des cytokines IFN- $\gamma$  et IL-17 et permettrait de rendre tolérogène les monocytes en permettant leur différenciation vers un phénotype M2, anti-inflammatoire et permettant la régénérescence tissulaire (188).

Modulateur des récepteurs de la sphingosine-1-phosphate : Le fingolimod séquestre les lymphocytes activés dans les ganglions lymphatiques, empêchant leur libération dans la circulation générale (188). Il réduit donc le nombre de tous les lymphocytes activés en périphérie (B, Th1,

Th17) et réduit les niveaux de cytokines pro-inflammatoires de ces derniers (TNF $\alpha$ , IFN- $\gamma$  et IL-17) (188). Il augmenterait également l'activité Th2 anti-inflammatoire (188).

Anticorps monoclonal anti-CD52 : L'alemtuzumab vise les récepteurs CD52 sur les lymphocytes activés afin de les détruire (188).

Anticorps monoclonal anti-intégrine  $\alpha 4$  : Le natalizumab bloque les intégrines  $\alpha 4$ , empêchant les lymphocytes T activés de s'arrimer à l'endothélium vasculaire pour commencer leur diapédèse à travers la barrière hémato-encéphalique (188).

Les anticorps monoclonaux anti-CD20 : Les ocrelizumab et rituximab neutralisent les cellules B activés contre la gaine de myéline en ciblant la molécule CD20 (188). Ils réduiraient également les cytokines pro-inflammatoires IFN- $\gamma$  et IL-17 produites par les Th1 (188). Ce sont actuellement les seuls médicaments de cette liste à être utilisés pour la forme progressive primaire en plus d'être utilisés chez les patient atteints de la forme récurrente-rémittente de la sclérose en plaques.

Anticorps monoclonal anti-CD25 : Le daclizumab cible la sous-unité alpha du récepteur à IL-2 mais il est actuellement retiré du marché canadien car il n'a pas fait preuve de sa totale innocuité chez l'humain. En inhibant l'activité de l'IL-2, il bloquerait ainsi l'expansion clonale des lymphocytes activés (188).

Inhibiteur de la synthèse de pyrimidine de novo : Le tériflunomide est un inhibiteur de l'enzyme dihydro-orotate déshydrogénase qui réduit l'activité des lymphocytes T et B (188). Il réduirait les réponses de Th1 et Th17 activés et augmenterait les réponses T Reg (188). Il réduirait de plus l'activité pro-inflammatoire des monocytes et des astrocytes (188).

Immunosuppresseur : Le mitoxantrone réduit la fonction et la mobilité des cellules immunitaires innées et adaptatives en perturbant la synthèse de cytokines pro-inflammatoires et des chimio-attractants (188).

Reconstituant immunitaire : La cladribine permet en autres une déplétion des cellules B activées contre la gaine de myéline (196).

## 2.3 Le diabète

Le diabète est une maladie polymorphique complexe. Les phénotypes cliniques sont nombreux et la seule caractéristique qui permet de tous les regrouper sous le terme de diabète est un taux élevé de glucose dans le sang, l'hyperglycémie.

Ce chapitre débutera donc volontairement par la présentation des différents phénotypes cliniques afin de les clarifier. Il se limitera à trois formes d'intérêt majoritaires, les diabètes de type 1, de type 2 et LADA. Leur épidémiologie, étiologie et physiopathologie sera ensuite abordée, ainsi que leurs thérapies médicamenteuses.

### 2.3.1 Les phénotypes cliniques

Le diabète de type 1 se caractérise par une destruction auto-immune des cellules  $\beta$  des îlots de Langerhans du pancréas conduisant à une carence sévère ou totale en insuline (7). La destruction nécessite une supplémentation extérieure de l'hormone hypoglycémiante afin de tarir un syndrome polyuro-polydipsique, un amaigrissement, voire une acidocétose pouvant conduire au coma et à la mort (7). Ce diabète est même appelé 1A, car il requière la présence d'auto-anticorps lors du diagnostic (7).

Les diabètes de type 1B correspondent à une destruction ou une dysfonction des cellules  $\beta$ , sans être d'origine auto-immune. Citons entre-autres, les diabètes MODY (*Maturity Onset Diabetes of the Young*), qui surviennent chez l'enfant par une mutation monogénique touchant les cellules  $\beta$  (7).

Le diabète de type 2 comprend une insulino-résistance et une insuffisance en insuline (7, 197). Il peut néanmoins à terme, devenir insulino-requérant (8 ans en moyenne après le diagnostic) (198).

Le diabète de type 1,5 ou LADA est dit intermédiaire entre le type 1 et 2, car il présente une destruction auto-immune et une insulino-résistance caractéristique du diabète de type 2 mais moins prononcée (7, 199). Il s'agit d'un diabète auto-immun de type 1A, à progression lente, qui survient à l'âge adulte (7). Il requière une supplémentation en insuline plus tardive, entre 3 à 6 mois et 1 an après le début des premiers symptômes (6, 200) (pouvant aller jusqu'à 6 ans (201)). Les marqueurs de l'auto-immunité sont majoritairement limités à deux autoanticorps (7), suggérant une différence physiopathologique entre le diabète de type 1A conventionnel et le diabète de type 1,5.

Le diabète gestationnel est soit transitoire (la femme ne présentera aucun diabète au cours de la vie, seulement lors de ses grossesses), soit inaugural d'un diabète de type 1 ou de type 2 qui se déclenche/décompense lors de la grossesse puis perdure par la suite. En fonction de son trimestre de survenue lors de la grossesse, ce diabète gestationnel peut donc être un premier épisode de diabète de type 1 (1<sup>er</sup> trimestre) ou de type 2 (2<sup>ème</sup> et 3<sup>ème</sup> trimestre) dit inaugural qui s'installera ultérieurement de façon pérenne (202). Le plus souvent, ce diabète gestationnel est un diabète de type 2 transitoire ou un premier épisode inaugural d'un diabète de type 2. Les femmes ayant un diabète gestationnel de type 2 ont en effet entre 7 et 12 fois plus de risque de développer un diabète de type 2 par ailleurs (203).

### **2.3.2 Épidémiologie**

L'organisation mondiale de la santé estimait la prévalence globale de diabète à plus de 400 millions d'individus en 2016 (25). Néanmoins, ce chiffre serait sous-estimé puisque Khan et al. l'estimaient en 2020 à 462 millions d'individus pour le seul diabète de type 2 (204). La très grande majorité des formes de diabète, près de 90% des cas prévalent, sont des diabètes de type 2, incluant le diabète gestationnel, (204) et 5 à 10 % sont des diabètes auto-immuns (7, 13).

La prévalence globale du diabète de type 1 serait d'environ 10% et son incidence annuelle standardisée sur l'âge serait de 5 à 15 cas pour 100 000 habitants dans le monde (13, 205). Au Canada, cette incidence grimperait même jusqu'à 20 à 24 cas pour 100 000 habitants entre 2006 et 2017 (13, 25). Le diabète de type 1 survient préférentiellement dans l'enfance, l'adolescence et chez le jeune adulte avant 19 ans (69). Le pic d'incidence se situe entre 10 et 14 ans (69) mais la survenue chez l'adulte, même au-delà de 40 ans, reste tout à fait possible.

La prévalence du diabète de type 2 était estimée à 6% (204) et son incidence globale standardisée sur l'âge était de 279 cas pour 100 000 habitants dans le monde en 2017 (205). En Amérique du nord, toujours en 2017, cette incidence dépassait même les 317 cas pour 100 000 habitants. Le diabète de type 2 survient habituellement chez l'adulte avec un pic d'incidence qui se situe à 55 ans (204). Néanmoins, les cas incident de diabète de type 2 chez l'enfant sont en constante augmentation, notamment au Canada (14).

L'incidence du diabète de type 2 semble similaire chez les hommes et les femmes, bien qu'une légère prédominance masculine (206) ait été rapportée (207). Les hommes (notamment adultes) auraient 1,5-1,8 fois plus de risque de développer un diabète de type 1 (69, 208). Cependant, cette prédominance est largement controversée (207).

En comparaison avec celle des caucasiens, l'incidence du diabète de type 2 serait plus élevée dans les populations insulaires du Pacifique (OR 3,1, IC à 95% 1,4–6,8), suivie des populations d'origine asiatique (RR 1,94, IC à 95% : 1,46–2,58), africaine (RR 1.36, IC à 95% : 1,14–1,63 et OR 2,3, IC à 95% : 2,1–2.6), des Amérindiens (OR 2,2, IC à 95%: 1,6–2,9), des hispaniques (RR 1,70, IC à 95%: 1,28–2,26) et même de toutes ethnies confondues (OR 1,8, IC à 95%: 1,5–2,9) (209). Néanmoins, ces résultats sont intrinsèquement corrélés à l'appartenance socioéconomique de ces différents groupes (209). Pour le diabète de type 1, les caucasiens seraient plus à risque que les autres ethnies (210).

Coincé entre les diabètes de type 1 et 2, ainsi qu'entre deux âges, la forme LADA débute chez l'adulte entre 20 et 40 ans, avec une prédominance après 30 ans (6, 200). Selon les populations, sa prévalence est estimée entre 2 et 12 % (6, 7, 200). Son incidence est difficile à estimer car 1) sa présentation clinique initiale comprend les phénotypes de diabète de type 1 et 2 et rend son diagnostic difficile; et 2) les études de cohorte, qui établissent son diagnostic et permettent une estimation de l'incidence, sont rares (6, 200). À ce jour, aucune disparité selon le sexe et l'ethnie n'est établie hormis les différences liées au système HLA souvent corrélées à l'ethnicité.

La prévalence et l'incidence du diabète gestationnel varient dans le monde en raison des caractéristiques des mères, comme le surpoids, l'obésité, l'âge à la procréation et l'ethnicité, ainsi que des critères diagnostiques différents (211). La prévalence standardisée sur l'âge du diabète gestationnel variait de 9,5% en Afrique à 26,6% en Asie du Sud-Est en 2017 (202). Au Canada, la prévalence était estimée entre 7 et 16% en 2016 (211). En Ontario, dans une vaste étude de cohorte de Feig et al., l'incidence annuelle ajustée sur l'âge du diabète gestationnel était entre 2,7 % et 5,6 % parmi les femmes ontariennes ayant donné naissance entre 1996 et 2010 (211, 212).

### 2.3.3 Physiopathologie des différents diabètes

Le point commun à tous les diabètes est une hyperglycémie chronique due à une carence en insuline légère, sévère ou absolue qu'il est possible de suivre par la concentration en peptide C (7, 197). L'origine auto-immune (type 1 et LADA) ou métabolique (type 2) du diabète permet une dichotomie qui s'estompe face à un besoin en insuline et des processus inflammatoires immunitaires qui s'exercent dans les diabètes de type 1, type 2 et LADA (7). Le contexte d'obésité favorise la synthèse de cytokines pro-inflammatoires (144) qui initialisent, renforcent et soutiennent les processus auto-immuns mais aussi métaboliques, ci-dessous décrits, conduisant à la mort des cellules  $\beta$ .

#### L'homéostasie glycémique, la carence en insuline et le peptide C

La glycémie (glucose dans le sang) est rétrocontrôlée par son propre taux. Deux hormones antagonistes, l'insuline et le glucagon, ayant respectivement une action hypo- et hyper-glycémiant, participent à cette homéostasie.

Les cellules  $\beta$  des îlots de Langerhans sont responsables de la production d'insuline, qui est synthétisée sous forme de pré-proinsuline puis transformée en proinsuline dans le réticulum endoplasmique (197). Cette proinsuline, transférée à l'appareil de Golgi, est ensuite clivée en peptide C et en insuline active en fonction du taux de glycémie (à la hausse) et tous deux sont excrétés de façon endocrine (197).

La carence en insuline, quel que soit le diabète, peut donc être tracée par la concentration plasmatique du peptide C (7, 197).

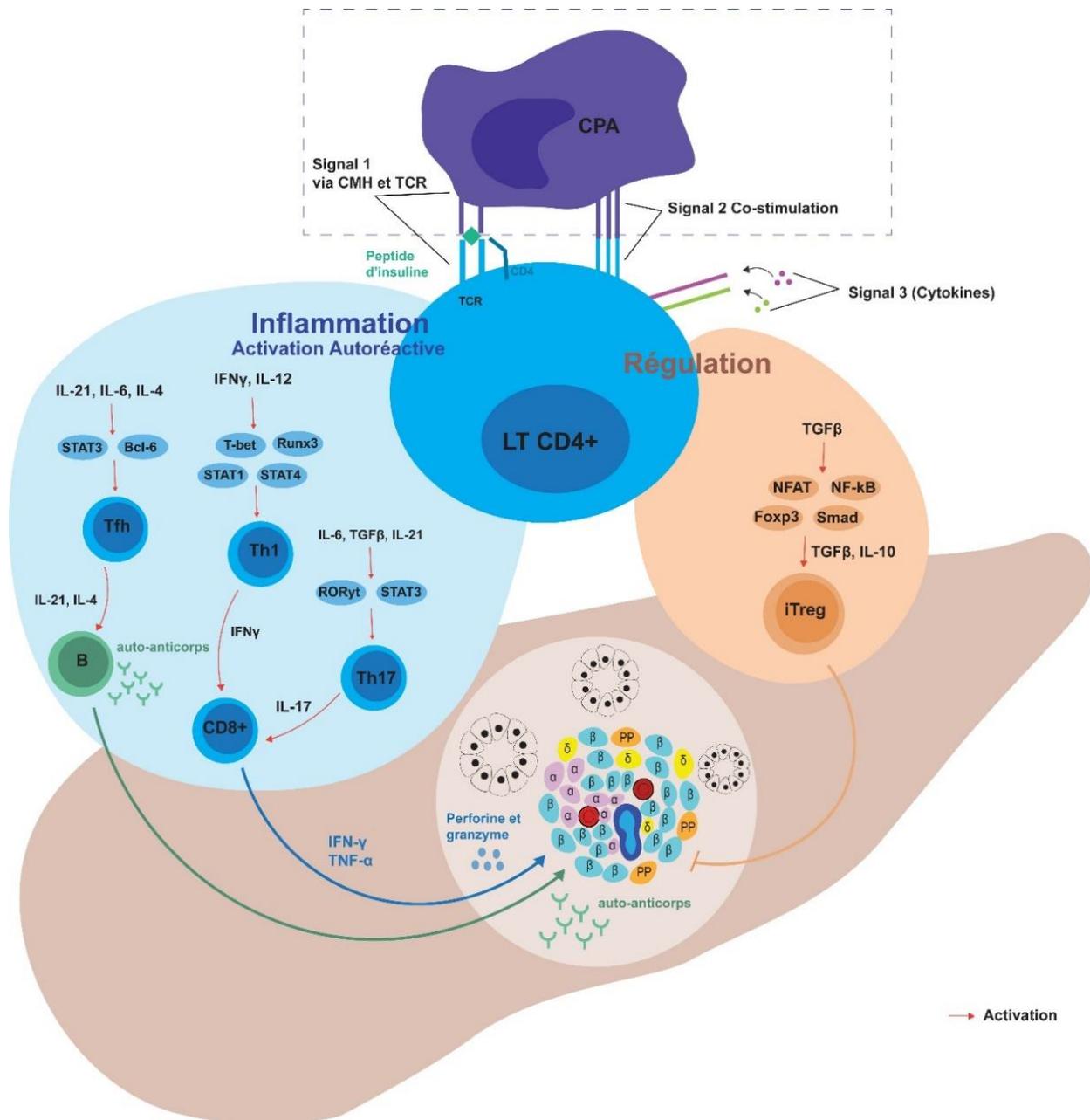
#### Les processus auto-immuns du diabète de type 1

La physiopathologie exacte du diabète de type 1 n'est pas totalement comprise (7). Néanmoins, le modèle animal de souris NOD apporte un scénario parmi l'hétérogénéité des processus pathologiques intervenant chez l'humain (85). Par extension de ce modèle, une destruction auto-immune ciblant les cellules  $\beta$  des îlots de Langerhans interviendrait chez des sujets présentant une susceptibilité génétique et est probablement accentuée par des facteurs environnementaux qui ne sont à ce jour pas tous établis (26).

La présence d'au moins un autoanticorps, traduisant l'implication de lymphocytes B autoréactifs dirigés contre des peptides des cellules  $\beta$  est une condition préalable à son diagnostic. Ce sont les

autoanticorps de l'acide glutamique décarboxylase (GADA) tels que GAD65, les autoanticorps des cellules des îlots de Langerhans (*Islet cells Antibodies*, ICA) tel que ICA512, les autoanticorps contre les tyrosines phosphatases, IA-2 et IA-2 $\alpha$ , les autoanticorps de l'insuline (IAA) et les autoanticorps dirigés contre l'isoforme 8 du transporteur de zinc (ZnT8) spécifique des îlots Langerhans (7). Quatre-vingt-dix pourcent des patients diabétiques de type 1 présentent l'ensemble de ces anticorps ou une majorité d'entre eux (7).

La présence d'immunoglobulines spécifiques dirigées contre des autoantigènes des îlots ne suffit pas pour expliquer le processus d'auto-immunité de ce diabète. L'immunité cellulaire participe majoritairement à la pathogenèse du diabète de type 1. En premier lieu, les lymphocytes T CD4+ folliculaires permettent une maturation des cellules B et l'affinement des autoanticorps (85, 213). Néanmoins, les T CD4+ et CD8+ sont majoritairement responsables de cette destruction ciblée (85).



**Figure 2.3 : Dérèglement de la tolérance périphérique et rôle de la signalisation du récepteur des cellules T (TCR) dans la pathogenèse du diabète de type 1.**

Inspiré de Clark et al. 2020 (85).

Les premiers autoantigènes présentés par les cellules présentatrices d'antigènes seraient des peptides d'insuline et de pro-insuline, ayant une très forte affinité pour les TCR (85). Les lymphocytes, chefs d'orchestre, les Th1 et Th17 autoréactifs activés, migrent et induisent dès lors un contexte pro-inflammatoire par leurs cytokines respectives (exemples: IFN- $\gamma$ , Il-17) dans les

îlots de Langerhans. Les Th1 induisent ainsi une première apoptose des cellules  $\beta$  par l'IFN- $\gamma$  (85). Les lymphocytes T CD8+, coactivés par les Th1, exerceraient leur action cytolytique (perforine et granzyme B). Enfin, la force ou la persistance du signal TCR induirait une augmentation des Th1 et Th17 autoréactifs au détriment de réponses T régulatrices (iFoxp3+Treg, favorisé par une faible signalisation TCR), anti-inflammatoire, dès lors submergées (85). L'attaque devient chronique par 1) les hauts niveaux de synthèse de l'IL-21 qui empêche l'épuisement des T autoréactifs effecteurs, 2) une stimulation chronique de cette tolérance périphérique dérégulée, et 3) la formation de cellules T mémoires qui, maintenues dans le microenvironnement des îlots, sont sans cesse réactivées et maintiennent ainsi l'inflammation chronique (85).

### La résistance et l'insuffisance à l'insuline du diabète de type 2

Le diabète de type 2 est une maladie métabolique multifactorielle complexe, impliquant prédispositions génétiques et de nombreux facteurs environnementaux (197). Il se caractérise par la juxtaposition de deux anomalies : la résistance puis l'insuffisance à l'insuline (7, 197).

Une résistance à l'insuline s'installe tout d'abord en raison de facteurs environnementaux, comme une avancée en âge, la sédentarité, la prise de poids, de mauvaises habitudes alimentaires et secondairement selon des facteurs génétiques (197). Cette résistance s'opère au niveau des cellules musculaires, du tissu adipeux et du foie (197), mais les nombreux mécanismes de cette résistance (214) ne seront pas ici détaillés.

Face à cette résistance, les cellules  $\beta$  augmentent leur sécrétion d'insuline (7) jusqu'à épuisement, dysfonctionnement et même mort cellulaire (199).

L'insuffisance en insuline s'explique par différents phénomènes qui se juxtaposent dont le premier est un stress du réticulum endoplasmique. Dans la cellule  $\beta$ , la mobilisation du  $\text{Ca}^{2+}$  à partir du réticulum endoplasmique puis de l'appareil de Golgi permet une excrétion vésiculaire de l'insuline active et du peptide C (197). Néanmoins, l'hyperglycémie et l'hyperlipidémie provoquées par les facteurs environnementaux favorisent sur le réticulum endoplasmique : 1) un stress oxydatif direct, 2) l'activation de la voie de signalisation UPR (*unfolded protein response*) et de son second messenger cAMP (*cyclic adenosine monophosphate*) pour la transduction du signal intracellulaire et 3) l'accumulation de proinsuline et de polypeptides amyloïdes précurseurs (dont la conformation spatiale n'est pas finalisée) (197). Ces trois voies conduisent à la génération d'espèces réactives de l'oxygène (ROS) par le réticulum endoplasmique qui inhibent les signaux naturels de

mobilisation du Ca<sup>2+</sup> et permettent ainsi l'excrétion de l'insuline active et du peptide C (197). De plus, ces voies activent la destruction du stock d'ARN messager de la cellule  $\beta$  (pour une réponse rapide de synthèse en insuline), et provoquent également des signaux pro-apoptotiques conduisant à la mort et à l'induction de cytokines pro-inflammatoires dans le microenvironnement des îlots de Langerhans (197).

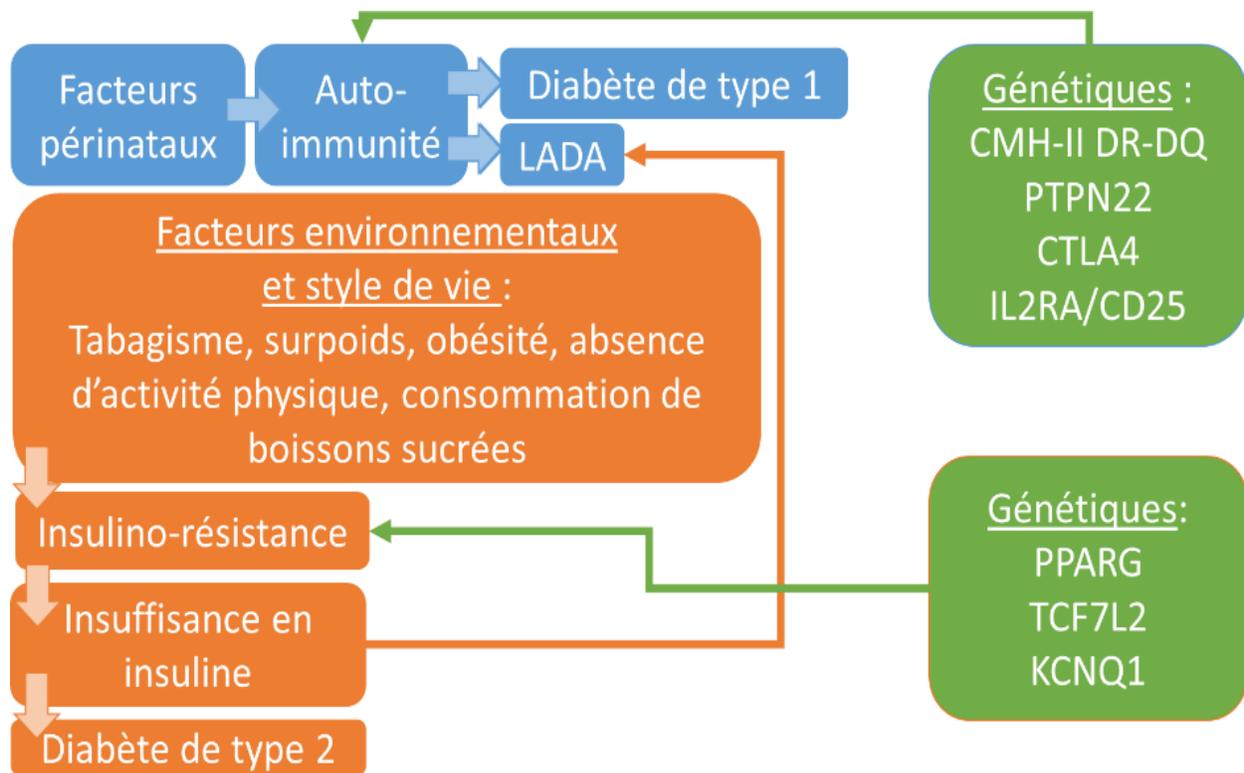
Ce premier mécanisme oxydatif et inflammatoire entraîne un stress au niveau des mitochondries qui l'amplifie et s'ajoute à une inflammation systémique causée par les mauvaises habitudes alimentaires, la sédentarité et le déséquilibre de la flore intestinale (197). Même en régulant le taux de glycémie à la normale, une mémoire métabolique se met en place et permet une chronicisation de ces phénomènes inflammatoires et de stress oxydatif (197) qui conduisent à l'épuisement et à la mort de la cellule  $\beta$ .

#### Le diabète LADA, un diabète intermédiaire entre les types 1 et 2

Devant la diversité des processus auto-immuns caractérisant le diabète de type 1 dit aigu (85), communément appelé diabète de type 1, le diabète LADA est également un diabète de type 1A, où se produisent des processus auto-immuns hétérogènes (parfois d'un individu à l'autre) comme ceux précédemment décrits pour le diabète de type 1 (199). Néanmoins, l'attaque est lente et progressive comparativement au diabète de type 1, dont l'attaque est rapide et soudaine (199). Tous deux conduisent à terme à une carence en insuline totale, mais le diabète de type 1 y conduit immédiatement alors que le diabète LADA y parvient en 6 à 12 mois (199). Ce diabète présente de plus des caractéristiques auto-immunes qui lui sont propres et un caractère hybride entre le diabète de type 1 et 2 (199). C'est pourquoi, le diabète LADA est aussi qualifié de diabète de type 1,5.

Au niveau de l'auto-immunité, seuls les autoanticorps GADA suivis parfois des ICA sont retrouvés chez les patients atteints de diabète LADA (7, 199). Les autres autoanticorps, caractéristiques du diabète de type 1, sont parfois précocement retrouvés chez des patients avec un diabète LADA mais cela reste rare (7, 199).

Ce diabète est dit intermédiaire entre le type 1 et 2, car il présente une destruction auto-immune et une insulino-résistance caractéristique du diabète de type 2, cependant moins prononcée (199). Cette caractéristique physiopathologique s'explique par le partage de nombreux facteurs de risque génétiques et environnementaux entre ce diabète et le diabète de type 2 (199).



**Figure 2.4 : Étiologie du diabète LADA, et des diabètes de type 1 et 2.**

Inspiré de Carlsson et al. 2019 (199)

Légende : CMH-II, complexe majeur d'histocompatibilité de classe 2; CTLA4, *T-lymphocyte-associated antigen* ; IL2RA, Gène de l'*Interleukin-2 Receptor alpha*; KCNQ1, Gène de *Potassium voltage-gated channel, KQT-like subfamily, member 1*; PPARG, Gène de la *Peroxisome Proliferator Activated Receptor Gamma*; PTPN22, Gène de la *Protein tyrosine phosphatase non-receptor type 22*; TCF7L2, Gène de la *Transcription Factor 7 Like 2*; Polymorphismes mononucléotidiques d'origine génétique.

### 2.3.4 Étiologie

Dans cette section portant sur l'étiologie, les connaissances sont résumées dans une série de tableaux de synthèse, précisant le niveau d'évidence épidémiologique et/ou immunologique pour chacun des facteurs ou groupes de facteurs mentionnés. Ce niveau d'évidence a été élaboré à partir de classifications existantes, celle du *World Cancer Research Fund grading system 2007* (104), adaptée à partir d'autres échelles de niveau d'évidence présentées par Burns 2011 dont celles du *Canadian Task Force on the Periodic Health Examination's Levels of Evidence* et *Levels of Evidence from Sackett* (105). Cette échelle est croissante à quatre niveaux comme celle du *World Cancer Research Fund grading system 2007* (104) mais elle reprend l'idée des autres échelles présentées par Burns (105) afin d'y inclure tous les types de devis d'étude (essais randomisés et non randomisés, cas-témoins, cohortes, études transversales, séries temporelles, conférence de consensus) bien menées ou avec des lacunes perçues. Parfois les études immunologiques ne pouvaient répondre aux critères épidémiologiques de ces échelles. Cette classification prend donc en considération le nombre d'études, la méthodologie utilisée, la convergence ou divergence des résultats entre les études, l'existence de revues, revues systématiques ou méta-analyses de la littérature ainsi que la cohérence globale des résultats et la plausibilité biologique.

#### 2.3.4.1 Les facteurs familiaux, génétiques

##### Histoire familiale de diabète

L'histoire familiale du diabète s'explique par l'intrication de facteurs génétiques et environnementaux (209, 215). Les facteurs génétiques constituent une part importante dans la survenue du diabète de type 1 alors que les facteurs environnementaux semblent constituer une part plus importante dans la survenue du diabète de type 2 (209, 215).

Pour le diabète de type 1, les personnes ayant des antécédents familiaux, maternels et/ou paternels de ce diabète (au 1<sup>er</sup> degré) ont jusqu'à 9 fois plus de risque d'en développer un eux-mêmes par rapport aux personnes n'ayant aucun antécédent familial pour ce diabète (215, 216).

Pour le diabète de type 2, les antécédents familiaux de ce diabète prédestinent un enfant à avoir entre 2 à 3 fois plus de risque de le développer lui-même par rapport à ses camarades n'ayant aucun antécédent familial de ce diabète (209, 215). Néanmoins, pour le diabète de type 2, le génotype

n'explique pas tout, puisque des variations de ce risque (209) ont été observées en fonction du poids (217) et des modes de vie des parents (régime alimentaire et activité physique), suggérant une part épigénétique et comportementale dans cette transmission (218).

Les personnes atteintes de diabète LADA partagent les caractéristiques des diabètes de type 1 et 2. L'héritabilité serait néanmoins d'autant plus prononcée pour des antécédents familiaux de diabète de type 1 (OR: 5,8; IC à 95%: 3,2–10,3) que des antécédents de diabète de type 2 (OR: 1,9; IC à 95%: 1,5–2,5) (215).

#### Les facteurs génétiques HLA et non-HLA

*Le diabète de type 1* : Les allèles du CMH de classe II, impliqués dans la présentation antigénique des CPA professionnelles aux LT CD4+, sont les plus largement incriminés dans le risque de survenue de diabète de type 1. Il s'agit des haplotypes HLA-DR3 (OR=3,5, IC à 95% : 3,3-3,8) et HLA-DR4 (OR=6,8, IC à 95% : 6,5-7,2), qui présentent les plus fortes associations (Tableau 2.5) (219, 220). Quatre-vingt-dix pour cent des patients atteints de diabète de type 1 sont même porteurs des deux haplotypes DR4/DQ8 ou DR3/DQ2 dont la combinaison avec DQ8 et DQ2 augmente le risque de diabète de type 1 (OR=13,2, IC à 95% : 10,1-17,2) (25, 221, 222). Les allèles du CMH de classe I participeraient également à l'augmentation de ce risque mais dans une moindre mesure : HLA-B8 (OR=3,2, IC à 95% : 3,0-3,4) et HLA-B15 (OR=3,7, IC à 95% : 3,6-3,8) (219, 220). Enfin, il existe des allèles protecteurs, tels que HLA-DR2, -DR5, et -DR7 (219, 220). Néanmoins, les variantes alléliques sont nombreuses selon les origines ethniques et ces facteurs de risque ou de protection ont largement été établis dans des populations majoritairement caucasiennes (25, 219).

En dehors des gènes du CMH, de nombreux polymorphismes mononucléotidiques peuvent altérer la réponse immunitaire (altération des molécules de costimulations ou des cytokines) ou des protéines utilisées par la cellule  $\beta$  et ainsi entraîner un risque accru de diabète de type 1 (Tableau 2.5) (219, 223). Parmi ces polymorphismes, citons les gènes de la protéine PTPN22 (qui inhibe la transduction de la signalisation des lymphocytes T et B), de l'immunoglobuline CTLA4 (inhibition des lymphocytes T), ou du récepteur de l'IL-2 (IL2RA/CD25) (Tableau 2.5) (219, 224).

*Le diabète de type 2* : Une récente méta-analyse suggère que les patients atteints de diabète de type 2 partageraient certains des allèles à risque du CMH présents chez les patients diabétiques de type 1 (225). Néanmoins, cette analyse nécessite d'être reproduite.

La susceptibilité génétique au diabète de type 2 résulterait donc de facteurs polygéniques complexes (plus d'une centaine identifiée à ce jour) (226), en dehors des gènes du CMH, et en interaction avec des facteurs environnementaux prédisposants (209). Le tableau 2.5 énumère quelques gènes (SNP) qui sont majoritairement liés à des fonctions des cellules  $\beta$  (226, 227).

*Le diabète LADA* : Du point de vue des risques associés aux allèles du CMH, le diabète LADA est beaucoup plus similaire au diabète de type 1 (6). Néanmoins, ils partagent également certains risques liés à des polymorphismes retrouvés pour le diabète de type 2 (199, 200). La revue de Nishimura et al. résume très bien dans un tableau l'ensemble des gènes HLA et les différents polymorphismes incriminés dans la survenue du diabète LADA (6). Par exemple (Tableau 2.5), les haplotypes HLA-DR3 (OR=3.4, IC 96% : 2.92–3.90) et HLA-DR4 (OR=3.3, IC 95% : 2.94–3.81) présentent une association moindre avec la survenue de diabète LADA (6) qu'avec le diabète de type 1. Le polymorphisme du gène du facteur de transcription TCF7L2 est associé aux diabètes LADA et de type 2 mais pas au diabète de type 1 (200). Les polymorphismes des gènes de la protéine PTPN22 et l'immunoglobuline CTLA4 ou du récepteur de l'IL-2 (IL2RA/CD25) sont associés à un risque plus élevé de survenue de diabète LADA et de diabète de type 1 mais pas de diabète de type 2 (200).

**Tableau 2.5: Facteurs familiaux et génétiques de risque ou de protection des diabètes de type 1, 2 et LADA**

Facteurs	Estimation <sup>a</sup> de l'association avec le diabète et intervalle de confiance (IC) à 95% ou valeur p	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
<b>Antécédents familiaux</b> (patients diabétiques par rapport à des témoins)				
<b>Diabète de type 1</b>	<b>9</b> (P<0.0001)	Cumul des facteurs génétiques liés au diabète de type 1	(215, 216)	++++
<b>Diabète de type 2</b>	<b>2,7</b> (2,2–3,3)	Cumul des facteurs environnementaux liés au diabète de type 2	(209, 215)	++++
<b>Diabète LADA</b>	Antécédents de type 1 : <b>5,8</b> (3,2–10,3) Antécédents de type 2 : <b>1,9</b> (1,5–2,5)	À la fois les facteurs génétiques liés au diabète de type 1 et les facteurs environnementaux liés au diabète de type 2	(215)	++++
<b>Facteurs génétiques liés aux allèles* du CMH</b> (liste non exhaustive d'allèles)				
<b>Diabète de type 1</b>				

Facteurs	Estimation <sup>a</sup> de l'association avec le diabète et intervalle de confiance (IC) à 95% ou valeur p	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
<b>HLA de classe II</b>	Allèles à risque (>1) ou protecteur (<1) des gènes HLA entre des patients diabétiques de type 1 par rapport à des témoins :	Le HLA-II permet au CPA la présentation de peptides étrangers aux lymphocytes. Le complexe peptides + HLA-II pourrait induire un mimétisme moléculaire à des peptides du soi.  Le HLA-I est porté sur l'ensemble des cellules nucléées, il présente des peptides étrangers aux T CD8+.	(219, 220, 222)	++++
Combinaison DR4/DQ8 ou DR3/DQ2	<b>13,2</b> (10,1-17,2)			
DR3	<b>3,5</b> (3,3-3,8)			
DR4	<b>6,8</b> (6,5-7,2)			
DR2	<b>0,2</b> (0,1-0,3)			
DR5	<b>0,3</b> (0,2-0,4)			
DR7	<b>0,2</b> (0,1-0,3)			
<b>HLA de classe I</b>				
B8	<b>3,2</b> (3,0-3,4)			
B15	<b>3,7</b> (3,6-3,8)			
<b>Diabète de type 2</b>				
<b>HLA de classe II</b>	Allèles à risque (>1) ou protecteur (<1) des gènes HLA entre des patients diabétiques de type 2 par rapport à des témoins :	Hypothèse des auteurs : une potentielle origine auto-immune des processus étiologiques du diabète de type 2	(225)	+
DRB5	<b>0,83</b> (0,74-0,94) (présence) <b>1,19</b> (1,07-1,32) (absence)			
DQB*06:02	<b>0,84</b> (0,74-0,95)			
DQA*01:02	<b>0,85</b> (0,75-0,95)			
DRB1*15:01~DQA1*01:02~DQB1*06:02	<b>0,84</b> (0,74-0,96)			
DRB1*07:01~DQA1*02:01~DQB1*03:03	<b>1,37</b> (1,12-1,67)			
<b>Diabète LADA</b>				
<b>HLA de classe II</b>	Allèles à risque (>1) ou protecteur (<1) des gènes HLA entre des patients LADA par rapport à des témoins :	Mêmes mécanismes que le diabète de type 1	(6)	++++
DR2 (DQB1*0601-DRB1*1502)	<b>0,1</b> (IC 95% non communiqué)			
DR3	<b>3,4</b> (2,92-3,90)			
DR4	<b>3,3</b> (2,94-3,81)			
DRB1	<b>0,27</b> (0,13-0,56)			
<b>Facteurs génétiques liés aux polymorphismes mononucléotidiques (SNP) sur les allèles* non-MHC</b> (liste non exhaustive de SNP)				
<b>Diabète de type 1</b>				
<i>T-lymphocyte-associated antigen</i> (CTLA4)	Les mutations SNP G par rapport à A, ou A par rapport à G :  +6230G>A (rs3087243) <b>0,84</b> (IC 95% non communiqué)  .49A>G (rs231775) <b>~1,4</b> (IC 95% non communiqué)	CTLA4 est exprimé sur les cellules T et régule négativement leur activation. La mutation affecte l'expression de CTLA4, augmente l'activité des T autoréactifs et donc la susceptibilité au T1D	(219, 224)	++++

Facteurs	Estimation <sup>a</sup> de l'association avec le diabète et intervalle de confiance (IC) à 95% ou valeur p	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
<i>Protein tyrosine phosphatase non-receptor type 22</i> (PTPN22)	Mutation G>A(rs2476601) 1,9 (IC 95% non communiqué)	Cette protéine inhibe la transduction de la signalisation des lymphocytes T et B. Sa mutation entraîne une activation plus facile des T et notamment des T autoréactifs, augmentant la susceptibilité au T1D	(219, 224)	++++
<i>Interleukin-2 Receptor alpha</i> (IL2Ra)	Mutation de la molécule CD25 (rs61839660) C>T, C par rapport à T <b>0,62</b> (IC 95% non communiqué)  (rs10795791) A>G, A par rapport à G <b>1,16</b> (IC 95% non communiqué)  (rs41295121) C>T, C par rapport à T <b>0,49</b> (IC 95% non communiqué)	Permet d'éteindre les cellules T autoréactives et régule les fonctions des Natural killers, des B et des T reg. Sa mutation augmente l'activité auto-immune et réduit la régulation. Par conséquent cette mutation augmente la susceptibilité au T1D	(219, 224)	++++
<i>Tumor-necrosis factor-α</i> (TNFα)	Mutation -308 (rs1800629) G>A, A par rapport à G ~ <b>1,1</b> (IC 95% non communiqué)	Augmente la production de cette cytokine pro-inflammatoire et promeut le T1D	(219, 224)	++++
<b>Diabète de type 2</b>				
Gène de la <i>Peroxisome Proliferator Activated Receptor Gamma</i> (PPARG)	379 non diabétiques par rapport à 333 patients avec diabète, tolérance au glucose altérée ou glycémie à jeun altérée par rapport à : <b>0,78</b> (0,59–1,05) L'alanine moins fréquente allèle (16 %) est associée avec une diminution du risque de diabète par rapport à C et T	PPAR code pour PPARγ, un récepteur d'hormone nucléaire qui régule l'adipogenèse et est une cible pour les thiazolidinediones, des médicaments utilisés pour traiter le diabète	(226-230)	++++
Gène de la <i>Transcription Factor 7 Like 2</i> (TCF7L2)	Génotypage de 228 marqueurs microsatellites chez des individus atteints de diabète de type 2 et des témoins tout au long d'un intervalle de 10,5 Mb sur le chromosome 10q : ~ <b>1,4</b> (IC 95% non communiqué)	Protéine impliquée dans la régulation glycémique		++++
<i>Potassium voltage-gated channel, KQT-like subfamily, member 1</i> (KCNQ1)	GWAS parmi 207 097 marqueurs SNP chez des individus japonais atteints de diabète de type 2 et de témoins : <b>1,26</b> (1,18-1,34)	Transport du K <sup>+</sup> dans la cellule β		++++
<b>Diabète LADA</b>				
<i>T-lymphocyte-associated antigen</i> (CTLA4)	SNP G>A <b>0,75</b> (0,68–0,83)	Mêmes mécanismes que le diabète de type 1	(6)	++++
<i>Protein tyrosine phosphatase non-receptor type 22</i> (PTPN22)	SNP A>C <b>1,5</b> (1,4–1,5)	Mêmes mécanismes que le diabète de type 1		

Facteurs	Estimation <sup>a</sup> de l'association avec le diabète et intervalle de confiance (IC) à 95% ou valeur p	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
Gène de la <i>Transcription Factor 7 Like 2</i> (TCF7L2)	SNP CT par rapport à TT <b>1,5</b> (1,2–1,9)	Mêmes mécanismes que le diabète de type 2		
<i>Interleukin-2 Receptor</i> (IL2Ra)	SNP A>G <b>0,70</b> (0,58–0,83)	Mêmes mécanismes que le diabète de type 1		

Ref, références; CMH, complexe majeur d'histocompatibilité; SNP, polymorphisme nucléotidique.

<sup>a</sup> Rapports de cotes, risque relatif ou hazard ratio.

<sup>b</sup> Niveau de preuve, selon le nombre et les méthodologies des études, évalué par P. Corsenac en adaptant la classification du *World Cancer Research Fund grading system 2007* (104), à partir des échelles d'évidence présentées par Burns 2011 dont celles du *Canadian Task Force on the Periodic Health Examination's Levels of Evidence* et *Levels of Evidence from Sackett* (105) :

+ FAIBLE Peu d'études, non répliquées, insuffisantes pour établir une association entre l'exposition et la maladie;

++ INTERMÉDIAIRE Études de cohorte, cas-témoins ou essais cliniques sans randomisation ou sans revue systématique de ces études; lacunes perçues (exemples : suivi insuffisant ou parcellaire; peu d'essais (ou d'études) disponibles ; échantillon non représentatif, faibles tailles; ajustement inadéquat ou incomplet).

+++ ÉLEVÉ Études de cohorte ou cas-témoins bien conçues et de grande envergure (d'une puissance adéquate), ou revue de la littérature ou méta-analyse mais en nombre limité; quelques lacunes et/ou résultats contraires subsistent; ajustement partiel. L'association devrait biologiquement être plausible.

++++ TRÈS ÉLEVÉ Associations cohérentes entre l'exposition et la maladie, avec peu ou pas de résultats contraires, basées sur des études épidémiologiques nombreuses et des revues systématiques et méta-analyses de ces nombreuses études (exemples : études observationnelles prospectives; Essais contrôlés randomisés avec une randomisation appropriée, une taille, une durée et une qualité suffisantes montrant des effets constants). L'association est biologiquement plausible.

\*Les allèles des antigènes de surface du complexe majeur d'histocompatibilité (CMH) sont également appelés *Human leukocyte antigen* (HLA) chez l'humain, car ces antigènes ont été découverts sur les leucocytes, deux classes de HLA sont particulièrement étudiées : HLA-Classe-I (sous-types: A, B, C) et HLA-Classe-II (sous-types: DR, DQ, DP). Le génotypage permet d'établir les allèles à risque et leurs fréquences chez des patients atteints versus les allèles sains (ou protecteurs) et leur fréquence chez des sujets sains selon l'approche gène candidat.

### 2.3.4.2 Les facteurs infectieux de l'exposome

Le diabète de type 1 est bien une maladie auto-immune. Néanmoins, la possibilité d'une auto-immunité induite par un virus (*bystander activation*) ou par mimétisme moléculaire est soutenue par certains auteurs (231) et réfutée par d'autres (26). Cette idée s'est largement répandue par des modèles animaux démontrant cette possibilité (231).

Les entérovirus (coxsackie) ont ainsi été les premiers et les plus largement soupçonnés par cette littérature établissant que ces infections étaient 10 fois plus courantes chez des patients diabétiques de type 1 que chez des personnes saines (26, 231). Néanmoins, rien à ce jour n'est formellement démontré; ces virus ne peuvent par leur seule infection causer directement le diabète de type 1 (26,

232, 233). Comme les susceptibilités génétiques, ils peuvent probablement être un préalable (*in utero*, ou dans l'enfance) ou un facilitant (232, 233).

D'autres virus, comme les herpès virus (Epstein Barr, Cytomégalo virus), ou des virus à ARN de la rubéole (Rubivirus) et des oreillons (Myxovirus paratodis) ont été incriminés (26, 231) mais aucune évidence à ce jour ne permet de les lier directement à la survenue de diabète de type 1.

Pour les diabètes de type 2 et LADA d'origine métabolique ou partiellement métabolique, les infections, par les processus inflammatoires concomitants, peuvent accentuer la résistance et l'insuffisance à l'insuline (234). Néanmoins, à ce jour, les résultats scientifiques ne soutiennent pas l'existence de facteurs de risque infectieux pouvant être associés à la survenue de ces types de diabète.

#### **2.3.4.3 Les facteurs individuels et environnementaux de l'exposome**

##### Facteurs de risque individuel, périnataux et de mode de vie pour les diabètes de type 1, 2 et LADA

###### *Le sexe*

Pour le diabète de type 1, un sur-risque masculin a été établi (69, 208), ce qui est extrêmement rare et controversé puisque les maladies auto-immunes sont le plus souvent à prédominance féminine (Tableau 2.6) (207). Hommes et femmes ont des tissus adipeux et musculaires, en répartition et en masse, différents (207). Ils n'ont donc pas la même résistance à l'insuline (207). Les diabètes de type 2 et LADA devraient donc être naturellement impactés par cette différence physiologique. D'un point de vue épidémiologique, il existerait une légère prédominance masculine de l'incidence du diabète de type 2 chez des sujets canadiens  $\geq 50$  ans (pic d'incidence de ce phénotype) mais sans ajustement sur l'indice de masse corporelle (IMC) et l'âge au diagnostic (206, 207). Pour le diabète LADA, aucune étude ne permet de répondre à une éventuelle disparité selon le sexe.

###### *Les facteurs périnataux et les facteurs environnementaux*

Le diabète de type 1 présente de nombreux facteurs de risque périnataux comme l'excès de poids à la naissance, la prématurité (ou la durée de gestation), l'avancée en âge de la mère et du père à la naissance, le nombre de frères ou sœurs plus âgés (ou l'ordre de naissance) et même la zone de résidence (urbaine ou rurale) (27, 235-241). Pour les facteurs environnementaux, beaucoup ont été

incriminés sans clairement être associés à sa survenue (26); citons les facteurs infectieux, précédemment abordés, les facteurs alimentaires, la pollution de l'air, les vaccins, l'environnement familial et le stress, mais aucun n'a d'association avérée avec la survenue du diabète de type 1 (26).

Les facteurs environnementaux et ceux liés au mode de vie, semblent donc avoir un impact peu important (s'il y en a un) sur la survenue du diabète de type 1 alors que les facteurs périnataux prennent une place prépondérante. Ceci est en totale opposition avec les diabètes de type 2 et LADA.

Hormis le faible poids de naissance (27, 199) dû à une malnutrition *in utero* favorisant une insulino-résistance et/ou un déficit de développement des cellules  $\beta$ , les diabètes de type 2 et LADA sont pratiquement dépourvus de facteurs de risque périnataux. Les facteurs environnementaux et ceux liés au mode de vie, prennent par contre une place prépondérante pour les diabètes de type 2 et LADA (199) (Figure 2.4). Ces facteurs, de risque ou de protection, qu'ils partagent en grand nombre, sont en effet communs à ces deux types de diabète, sans doute par leur aspect métabolique commun.

### Les facteurs de risque liés au mode de vie

#### *Le tabagisme*

La nicotine induit une résistance à l'insuline et possède une action anti-inflammatoire et immunomodulatrice (242). Elle serait donc censée réduire le risque de diabète de type 1 par son action anti-inflammatoire et favoriser les diabètes de type 2 et LADA en augmentant la résistance à l'insuline. Pour le diabète type 1 qui survient en jeune âge, l'exposition *in utero* au tabagisme de la mère protège de sa survenue (RR= 0,79, IC 95%, 0,71-0,87) (243). Néanmoins, la nicotine est le plus souvent fumée, le tabagisme expose donc à bon nombre de substances autres que cette dernière. Pour le diabète de type 2, le tabagisme semble donner lieu à un risque accru (209, 242). Pour le diabète LADA, les résultats épidémiologiques sont faibles ou contradictoires et ne permettent pas d'établir de preuves formelles (27, 243).

#### *Le surpoids et l'obésité*

Le surpoids et l'obésité favorisent l'insulino-résistance (214), l'inflammation chronique et l'auto-immunité (144). Le risque du diabète de type 1 serait doublé voire triplé (90) mais pour les seuls enfants obèses comparativement aux enfants non obèses. Les personnes obèses ont ainsi un risque

entre 5 et 40 fois plus élevé d'avoir un diabète de type 2 (en fonction de leur IMC, Tableau 2.6) et 5 fois plus élevé d'avoir un diabète LADA que les personnes non obèses (27, 209).

#### *La vitamine D, l'immunomodulateur*

La vitamine D module la production de cytokines pro-inflammatoires des lymphocytes T et B grâce aux récepteurs de vitamine D exprimés à leur surface (145, 146). Elle permet également de rendre tolérogène les cellules dendritiques (150). Elle devrait en toute vraisemblance protéger de l'auto-immunité et de l'inflammation chronique et donc des diabètes de type 1, 2 et LADA. Malgré une plausibilité physiologique et clinique (supplémentation en vitamine D) (244, 245), l'évidence épidémiologique d'une association causale entre la déficience de cette vitamine et la survenue de ces trois diabètes n'est pas au rendez-vous (244, 245).

#### *Consommation de boissons sucrées*

À notre connaissance, aucune étude épidémiologique n'a étudié l'association entre cette consommation et la survenue du diabète de type 1. La consommation quotidienne de boissons sucrées multiplie par 2 le risque de survenue des diabètes de type 2 et LADA. Pour une consommation de 2 boissons sucrées (de 200 ML) par jour, le risque de diabète LADA est estimé à : OR = 1,99, IC à 95% : 1,11-3,56 (246) par rapport aux personnes n'en consommant pas ou en consommant juste 1 fois par mois. Pour une consommation de 1 boisson sucrée (de 250 ML) par jour le risque de diabète de type 2 est estimé à : RR=1,19, IC à 95% : 1,13-1,25 (247), mais chaque consommation supplémentaire (de 250 ML) par jour l'augmenterait de 19% par rapport aux personnes n'en consommant pas ou en consommant juste 1 fois par mois (247). Pour le diabète de type 1 déjà diagnostiqué, la consommation de sucre en ferait vraisemblablement progresser l'avancée par une augmentation de la demande en insuline, une apoptose des cellules  $\beta$  et/ou une augmentation de l'auto-immunité cellulaire et humorale (248).

#### *Le cortisol et le stress*

La survenue du diabète de type 1 est clairement associée au stress dans la revue de Sharif et al. (249). Par exemple, Karavanaki et al. ont établi que le risque de développer un diabète de type 1 était plus élevé chez les enfants ayant eu une période de stress deux années en amont de leur diagnostic: RR : 2,49 (IC à 95% : 1,14-5,40) (250). Les associations entre le stress et le diabète de type 2 sont très bien relatées dans un tableau récapitulatif de la recension des écrits de Hackett (251). L'ensemble des résultats ne seront pas ici décrits en détail mais ils suggèrent également

clairement des associations à risque. Beaucoup de ces résultats diffèrent entre les hommes et les femmes (Tableau 6, exemple de Kato et al. 2009 (252)) chez les personnes ayant subi un stress en amont de leur diagnostic. Pour le diabète LADA, peu de données probantes existent mais elles tendent à corroborer cette même association entre le stress et la survenue de ce diabète (27, 199).

### *Le café*

Le café augmenterait la sensibilité à l'insuline mais aussi l'auto-immunité (170). Il serait donc protecteur vis-à-vis du diabète de type 2 (253) et un facteur de risque pour les diabètes de type 1 (170) et LADA (27). Pour le diabète de type 1, les associations sont contradictoires (170). Les mécanismes physiologiques et la relation causale entre cette exposition et la survenue de diabète de type 2 n'est à ce jour pas établie (254). Pour le diabète LADA, la littérature ne permet pas actuellement de se positionner (27, 255).

### Les facteurs protecteurs liés au mode de vie

#### *L'activité physique*

L'activité physique augmente la consommation de glucose et réduit l'hyperglycémie et l'insulino-résistance. Pour le diabète de type 1, aucune association n'est établie (256). Le manque d'activité physique est donc associé à entre 3 et 5 fois plus de risque d'avoir un diabète LADA (Tableau 2.6) (27). Dans une méta-analyse de 28 cohortes, Smith et al. ont corroboré ce résultat pour le diabète de type 2 (257). Ils ont montré une dose-réponse entre le temps d'activité de loisir d'intensité modérée et la survenue de ce diabète; passant d'un risque relatif de 0,93 (IC à 95% : 0,92 -0,95) à 0,47 (IC à 95% : 0,34-0,65) avec respectivement 2,25 et 60 heures en équivalents métaboliques par semaine d'activité physique modérée par rapport à des personnes inactives (257).

#### *Consommation modérée d'alcool*

Pour le diabète de type 1, la consommation et même l'initiation à l'alcool n'est pas conseillée pour les enfants et les adolescents par les recommandations de santé publique. Pour les formes plus tardives, son effet bénéfique serait attribué à une augmentation de la sensibilité à l'insuline et des effets anti-inflammatoires qui agiraient positivement sur les diabètes de type 2 et LADA (27, 258). Néanmoins, pour le diabète de type 2 les méta-analyses existantes sont contradictoires (Tableau 2.6) (27, 199, 258, 259) et pour le diabète LADA, les études existantes manquent de reproductibilité (27, 199, 260).

## Alimentation

Les bienfaits d'une bonne hygiène alimentaire ne sont clairement plus à démontrer. Elle est prescrite comme un « médicament » de première intention pour les diabétiques de type 2 (261). Pour le diabète de type 1, la diète méditerranéenne riche aurait une action anti-inflammatoire, en relation avec un bon équilibre de la flore intestinale (262). Néanmoins une méta-analyse récente démontre qu'il n'existe aucune association entre des suppléments de l'enfant en acides gras polyinsaturés Omega-3 et Omega-6 et la survenue clinique d'un diabète de type 1 (Tableau 2.6) (263). Pour le diabète de type 2, les bienfaits proviendraient d'une alimentation riche en acide gras essentiels et en fibres composée de céréales entières, de poisson gras, de fruits et de produits laitiers, au détriment de la viande rouge, viandes transformées et boisson sucrées (Tableau 2.6) (264). Pour le diabète LADA, la consommation de poisson gras diminuerait également le risque de survenue de ce diabète ( $\geq 1$  par rapport à  $< 1$  portion de poisson gras par semaine, OR= 0,51 (IC à 95% : 0,30-0,87) (27, 265) et celle de viande rouge transformée l'augmenterait (OR=1,27, IC à 95% : 1.07–1.52) pour une portion par jour par rapport à aucune (266). Néanmoins, ces deux analyses provenant d'études cas-témoins nécessitent d'être répliquées.

**Tableau 2.6: Facteurs de risque ou de protection individuels, périnataux et de mode de vie qui influencent les survenues de diabètes de type 1, 2 et LADA**

Facteurs	Estimation <sup>a</sup> de risque ou de protection des diabètes et intervalle de confiance (IC) à 95%	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
<b>Facteur individuel</b>				
<b>Sexe</b> (rapport hommes / femmes dans l'incidence)	<b>Diabète de type 1</b> <b>1,47</b> (1,33-1,60)  Incidence Hommes/femmes âgés de 15-34 ans entre 1983-2002= 16,4/8,9 pour 100 000 soit ratio= <b>1,8</b>	Meilleure signalisation TLR des dangers et présentation antigénique par CMH-II ; la rééquilibration des <b>œstrogènes</b> entre la voie pro-inflammatoire NF-kB via ER- $\alpha$ (faible concentration d'oestradiol) et la voie anti-inflammatoire via la coalescence ER- $\alpha$ /œstrogènes (forte concentration d'oestradiol) permet les signalisations GPCR30 et améliore la voie PD-1 (FOXP3, régulatrice) ; une meilleure hypermutation somatique et recombinaison de classe pour les cellules B; <b>Testostérone</b> : maintient l'activité T Reg et la non-	(69, 207, 208)	+++ Mais controversé

Facteurs	Estimation <sup>a</sup> de risque ou de protection des diabètes et intervalle de confiance (IC) à 95%	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
		méthylation de FOXP3, inhibe Th1.		
	<b>Diabète de type 2</b> Incidence Hommes/femmes ≥ 50 ans 15,9/1000 par rapport à 12,7/1000; p < 0.001 Ratio=1,3	Masse grasse et masse musculaire différentes chez hommes et femmes n'induit pas la même insulino-résistance	(206, 207)	+++ Mais controversé
	<b>Diabète LADA</b> -		-	Aucune évidence
<b>Facteurs périnataux</b>				
<b>Diabète de type 1</b>				
<b>Excès de poids</b> à la naissance	>4Kg versus 3 à 3,5Kg à la naissance : <b>1,10</b> (1,04-1,19) <b>1,43</b> (1,11-1,85)	Inconnu	(235, 236, 239)	++++
<b>Prématurité</b> (ou durée de gestation)	Naissance prématurée par rapport à Non : ~ <b>1,20</b> (1,11-1,29)	Inconnu, néanmoins la prématurité du système adaptatif pourrait entraîner un défaut de la régulation T et B Reg facilitant l'auto-immunité du diabète de type 1	(237-239)	++++
<b>Nombre de frères ou sœurs plus âgés</b> (ou l'ordre de naissance)	≥4 par rapport à 0 : <b>0,65</b> (0,55-0,76)	Inconnu	(240, 267)	++++
<b>L'âge de la mère et du père</b> à la naissance	Mères <25 ans par rapport à 25-29 ans: <b>1,94</b> (1,34-2,81),  Pères >30 ans par rapport à 25-29 ans : <b>1,56</b> (1,16-2,10)	Inconnu	(241, 268)	++++
<b>Zone de résidence</b> (rurale/urbaine)	Zones rurales périphériques ou très périphériques par rapport à zone urbaine : <b>1,2</b> (1,04-1,45)	Inconnu	(269, 270)	++++
<b>Diabètes de type 2 et LADA</b>				
<b>Faible poids</b> à la naissance	<b>Diabètes de type 2 et LADA</b> <3000g par rapport à ≥4000: ~ <b>5</b> (IC 95% non communiqué)  <b>LADA</b> <3000g par rapport à ≥4000: <b>2,38</b> (1,23-4,60)	Inconnu	(27, 199)	++++
<b>Facteurs de risque liés au mode de vie</b>				
<b>Tabagisme</b>	<b>Diabète de type 1</b> Mères fumeuses par rapport à Non-fumeuses pendant la grossesse <b>0,79</b> (0,71-0,87)  <b>Diabète LADA :</b> Fumeurs par rapport à Non-fumeurs	La nicotine a des effets immuno-modulateurs et anti-inflammatoires contrebalancés par la fumée de cigarette, cette dernière étant pro-inflammatoire	(199, 243)	+++ Mais controversé

Facteurs	Estimation <sup>a</sup> de risque ou de protection des diabètes et intervalle de confiance (IC) à 95%	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
	<b>1,00</b> (non-significatif, IC 95% non communiqué)			
	<b>Diabète de type 2 :</b> Fumeurs>25cigarettes/j par rapport à Non-fumeurs ~2 (1,25-3,03) (Un exemple sur une multitude décrit par Ismail et al. 2021)	La nicotine augmente l'insulino-résistance	(199, 209, 242)	++++
Surpoids et obésité	<b>Diabète de type 1</b> Obésité infantile selon IMC Obèses par rapport à normal : <b>2,03</b> (1,46– 2,80)  Une relation dose-réponse a été soutenue par une association continue entre l'IMC de l'enfant et diabète de type 1 (rapport de cotes combiné=1,25 (1,04–1,51) pour une élévation d'1 déviation standard d'IMC.	Le surpoids et l'obésité favorisent l'insulino-résistance (214), l'inflammation chronique et l'auto-immunité (144).  Synthèse de cytokines pro-inflammatoires via la leptine; Déséquilibre Th1-Th17/Th Reg ; Diminution de la biodisponibilité de la vitamine D ; induction du facteur ROR $\gamma$ t (et donc des Th17).	(90, 271)	++++
	<b>Diabète de type 2</b> IMC (kg/m <sup>2</sup> ) [18,5;29] par rapport à $\leq$ 18,5 : <b>5,16</b> (1,92–13,80) IMC (kg/m <sup>2</sup> ) [30;35] par rapport à $\leq$ 18,5 : <b>5,25</b> (1,96–14,04)  IMC (kg/m <sup>2</sup> ) $\geq$ 30 par rapport à <23 <b>11</b> (IC 95% non communiqué)  IMC (kg/m <sup>2</sup> ) $\geq$ 35 par rapport à <23 <b>38,8</b> (31,9–47,2)		(199, 209)	++++
	<b>Diabète LADA</b> IMC (kg/m <sup>2</sup> ) $\geq$ 30 par rapport à <23 <b>5</b> (IC 95% non communiqué)		(199)	++++
Vitamine D	<b>Diabètes de type 1, 2 et LADA N.A</b> (N.A) Toujours basé sur des suppléments cliniques ou alimentaires en vitamine D.	La vitamine D inhibe les cytokines pro-inflammatoires via le récepteur de la vitamine D ; Rééquilibre la balance Th1-Th17/Th Reg.	(221, 222)	+++ Données non concluantes
Consommation de boissons sucrées	<b>Diabète de type 1</b> N.A (N.A)	La consommation de sucre ferait progresser le diabète de type 1 par une augmentation de la demande en insuline, une apoptose des cellules $\beta$ et/ou une augmentation de l'auto-immunité cellulaire et humorale	-	+
	<b>Diabète de type 2</b> 1 boisson sucrée de 250 ML/j par rapport à des personnes n'en consommant pas ou en consommant juste 1 fois par mois :	Augmente la demande en insuline et la résistance à l'insuline	(247)	+++

Facteurs	Estimation <sup>a</sup> de risque ou de protection des diabètes et intervalle de confiance (IC) à 95%	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
	<p><b>1,19</b> (1,13-1,25)</p> <p>Augmente de <b>19%</b> pour chaque consommation de 250ML/j supplémentaire</p>			
	<p><b>Diabète LADA</b> 2 boissons sucrées de 200 ML/j par rapport à des personnes n'en consommant pas ou en consommant juste 1 fois par mois :</p> <p><b>1,99</b> (1,11-3,56)</p>		(246)	++
Cortisol et stress	<p><b>Diabète de type 1</b> Évènement stressant dans la période de 2 ans précédant le diagnostic par rapport à aucun stress :</p> <p><b>2,49</b> (1,14-5,40)</p>	Stimulation neuroendocrine de l'axe hypothalamo-hypophysaire favoriserait une pro-inflammation et une stimulation de l'auto-immunité.	(249, 250)	++++
	<p><b>Diabète de type 2</b> Stress perçu basé sur la question : « Combien de stress ressentez-vous dans la vie quotidienne ? » (Échelle à trois points : faible, moyen, élevé; exemple de Kato et al. 2009) Élevé par rapport à faible :</p> <p>Hommes: <b>1,39</b> (1,15-1,65) Femmes: <b>1,25</b> (1,01-1,56)</p>	En réponse au stress, l'allostase adrénaline et cortisol permet une production accrue d'insuline pour anticiper un besoin accru en glucose (qui n'est pas réel) des cellules. La permanence ou la répétition de cette boucle de rétrocontrôle augmenterait progressivement la résistance à l'insuline	(251, 252)	++++
	<p><b>Diabète LADA</b> N.A Augmentation de 80% du risque de survenue de diabète LADA en présence d'un bien-être faible ou de troubles du sommeil (pas stress direct)</p>	Les deux mécanismes précédents.	(27, 199)	+++
Café	<p><b>Diabète de type 1</b> Enfants consommant 1 tasse/j par rapport à pas de consommation :</p> <p><b>1,69</b> (1,21-2,37) Exemple d'association de Virtanen et al. (non probante)</p>	Augmenterait la sensibilité à l'insuline et l'auto-immunité mais mécanismes non prouvés et relation causale non établie, car d'autres auteurs comme Sharif et al. parlent d'immunomodulation potentielle avec une baisse de synthèse de TNF $\alpha$ , IFN- $\gamma$ et une augmentation de l'IL-10. Elle améliorerait donc la régulation.	(170, 272)	+
	<p><b>Diabète de type 2</b> Médiane d'environ 5 tasses par jour par rapport à l'absence de consommation de café :</p> <p><b>0,7</b> (IC 95% non-communiqué, méta-analyse)</p>		(254)	+++ Mais controversé
	<p><b>Diabète LADA</b> Quantité de tasse de café/j :</p> <p><b>1,11</b> (1,00-1,23) Lorsque catégorisée, <math>\geq 6</math> tasses/j par rapport à <math>&lt;2</math> tasses/j</p> <p><b>2,02</b> (0,99-4,12) Exemple d'association de Lofvenborg et al. (non probante)</p>		(27, 199, 255)	+

Facteurs	Estimation <sup>a</sup> de risque ou de protection des diabètes et intervalle de confiance (IC) à 95%	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
<b>Facteurs de protection liés au mode de vie</b>				
Activité physique	<b>Diabète de type 1</b> Activité physique modérée tous les jours par rapport à moins d'une fois par semaine : 0,87 (0,24–3,10)	Hypothèse : l'insulino-résistance participe à l'initiation de processus auto-immuns.	(27, 199, 256)	++
	<b>Diabète de type 2</b> Effet dose-réponse avec respectivement 2,25 et 60 heures en équivalent métabolique par semaine d'activité physique modérée par rapport à des personnes inactives : <b>0,93</b> (0,92 -0,95) à <b>0,47</b> (0,34-0,65)	L'activité physique augmente la sensibilité à l'insuline et l'utilisation du glucose.	(257)	++++
	<b>Diabète LADA</b> Activité physique modérée tous les jours par rapport à moins d'une fois par semaine : <b>3,04</b> (1,23-7,50) <b>3,16</b> (1,18-8,46) <b>5</b> (IC 95% non communiqué)		(27, 199, 256)	+++
Consommation modérée d'alcool	<b>Diabète de type 1</b> La consommation d'alcool est déconseillée aux enfants et adolescents	Effet bénéfique serait attribué à une augmentation de la sensibilité à l'insuline (pour les diabètes de type 2 et LADA) et des effets anti-inflammatoires (pour le diabète de type 1)	-	+
	<b>Diabète de type 2</b> Par rapport à la catégorie minimale de consommation d'alcool : une consommation d'alcool légère <b>0,83</b> (0,73, 0,95) et modérée <b>0,74</b> (0,67-0,82)		(258, 259)	++++ Méta-analyses contradictoires
	<b>Diabète LADA</b> Pour chaque tranche de 5 g d'alcool/j par rapport à aucune consommation : Faibles niveaux de GADA <b>0,85</b> (0,76-0,94) Niveaux élevés de GADA 1,00 (0,94-1,06).		(27, 199, 260)	+
Alimentation	<b>Diabète de type 1</b> Supplémentation en acides gras polyinsaturés Omega-3 par rapport à aucune 0,87 (0,71-1,08)  Supplémentation en acides gras polyinsaturés Omega-6 par rapport à aucune 1,05 (0,92-1,20)	Les fibres contenues dans les céréales entières et les acides gras essentiels contenus dans les poissons gras favorisent une bonne santé de la flore intestinale en équilibre avec le système immunitaire inné et adaptatif. Une bonne alimentation a donc action anti-inflammatoire et tolérogène.	(263)	
	<b>Diabète de type 2</b> Consommation élevée par rapport à faible		(264)	++++

Facteurs	Estimation <sup>a</sup> de risque ou de protection des diabètes et intervalle de confiance (IC) à 95%	Mécanismes immunologiques ou métaboliques	Réf	Niveau d'évidence <sup>b</sup>
	Céréales entières (ingestion globale entre 0–302 g/j) : <b>0,77</b> (0,71-0,84) Poisson (ingestion globale entre 0–225 g/j) : 1,04 (0,95-1,13) Fruits (ingestion globale entre 10–618 g/j) : 0,96 (0,93-1,00) Produits laitiers (ingestion globale entre 0–2000 g/j) : <b>0,91</b> (0,85-0,97) Viande rouge (ingestion globale entre 0–207 g/j) : <b>1,21</b> (1,13-1,30) Viandes rouge transformées (ingestion globale entre 0–142 g/j) : <b>1,27</b> (1,20-1,35) Boisson sucrées (ingestion globale entre 0–748 ml/j) : <b>1,30</b> (1,20-1,40)	permet de limiter l'hyperglycémie, la synthèse d'insuline et l'insulino-résistance		
	<b>Diabète LADA</b> ≥1 par rapport à <1 portion de poisson gras par semaine : <b>0,51</b> (0,30-0,87) Viandes rouge transformées (pour une portion/j) : <b>1,27</b> (1,07-1,52)		(265, 266)	++

IMC, Indice de masse corporel; N.A, Non applicable

<sup>a</sup> Rapports de cotes, risque relatif ou hazard ratio.

<sup>b</sup> Niveau de preuve, selon le nombre et les méthodologies des études, évalué par P. Corsenac en adaptant la classification du *World Cancer Research Fund grading system 2007* (104), à partir des échelles d'évidence présentées par Burns 2011 dont celles du *Canadian Task Force on the Periodic Health Examination's Levels of Evidence* et *Levels of Evidence from Sackett* (105) :

+ FAIBLE Peu d'études, non répliquées, insuffisantes pour établir une association entre l'exposition et la maladie;

++ INTERMÉDIAIRE Études de cohorte, cas-témoins ou essais cliniques sans randomisation ou sans revue systématique de ces études; lacunes perçues (exemples : suivi insuffisant ou parcellaire; peu d'essais (ou d'études) disponibles ; échantillon non représentatif, faibles tailles; ajustement inadéquat ou incomplet).

+++ ÉLEVÉ Études de cohorte ou cas-témoins bien conçues et de grande envergure (d'une puissance adéquate), ou revue de la littérature ou méta-analyse mais en nombre limité; quelques lacunes et/ou résultats contraires subsistent; ajustement partiel. L'association devrait biologiquement être plausible.

++++ TRÈS ÉLEVÉ Associations cohérentes entre l'exposition et la maladie, avec peu ou pas de résultats contraires, basées sur des études épidémiologiques nombreuses et des revues systématiques et méta-analyses de ces nombreuses études (exemples :

études observationnelles prospectives; Essais contrôlés randomisés avec une randomisation appropriée, une taille, une durée et une qualité suffisantes montrant des effets constants). L'association est biologiquement plausible.

### **2.3.5 Les thérapeutiques médicamenteuses**

Les patients diabétiques disposent de deux thérapeutiques médicamenteuses en dehors des règles hygiéno-diététiques : l'insulinothérapie et les antidiabétiques oraux.

#### **2.3.5.1 L'insulinothérapie**

L'insuline est le traitement des diabètes de type 1 et LADA dans le cadre d'une carence sévère ou absolue d'insuline qui n'est plus synthétisée par les cellules  $\beta$  du patient. Pour le diabète LADA, l'insuline est nécessaire après une période d'environ 3 à 6 mois et 1 an, avant laquelle elle n'est pas requise. L'insuline est également l'un des traitements du diabète de type 2 lorsque l'insuffisance insulinaire est devenue trop prononcée (7).

#### **2.3.5.2 Les antidiabétiques oraux (ADO)**

Ce sont les traitements du diabète de type 2. Ils sont utilisés seuls ou en association avec l'insulinothérapie en fonction des besoins (7). Les différentes classes seront ici énoncées, ainsi que leur mode d'action. Ils sont parfois utilisés dans le diabète de type LADA durant la période qui précède la nécessaire insulinothérapie, car ces derniers sont souvent initialement confondus avec un diabète de type 2 (201). Des médicaments de première intention, les biguanides et les sécrétagogues de l'insuline, les sulfonylurées, ont également été utilisés spécifiquement pour traiter le diabète LADA et le sont encore aujourd'hui même en connaissance du diagnostic, même s'ils ne sont plus à ce jour préconisés dans le traitement de ce phénotype de diabète (201). De même, les autres classes présentées ci-dessous, peuvent néanmoins continuer à être prescrites par les cliniciens pour leur action modérée sur le diabète de type LADA (201).

Les biguanides (metformine) : Ils diminuent la production hépatique de glucose et augmentent l'insulino-sensibilité et l'utilisation périphérique du glucose (261).

Les sécrétagogues de l'insuline, les sulfonylurées: Également appelés sulfamides hypoglycémiant, stimulent la production d'insuline par les cellules  $\beta$  (261).

Les sécrétagogues de l'insuline, non-sulfonylurées ou méglitinides : Sont également appelés glinides, ils stimulent la production d'insuline par les cellules  $\beta$  (261).

Les agonistes des récepteurs du peptide-1 apparenté au glucagon (ou GLP-1) : Ce sont des agonistes des incrétines (hormones intestinales) dont le taux est réduit ou absent chez les patients diabétiques (261). Ils permettent une augmentation de la sécrétion d'insuline, une diminution de la synthèse du glucagon, un retard de la vidange gastrique, et une augmentation de la satiété (261).

Les inhibiteurs du cotransporteur sodium-glucose de type 2 (ou iSGLT-2) : Ils favorisent l'élimination du glucose dans les urines en bloquant sa réabsorption (261).

Les inhibiteurs de la dipeptidyl peptidase-4 (ou iDPP-4) : Ils augmentent les effets des incrétines en inhibant la dégradation des GLP-1 (261).

Les thiazolidinediones (ou TZD) : Ils augmentent l'action de l'insuline en augmentant l'insulinosensibilité (261).

Les inhibiteurs des  $\alpha$ -glucosidases : Ils ralentissent la digestion des glucides dans l'intestin (273).

## 2.4 Le bacille Calmette-Guérin (BCG)

### 2.4.1 Historique

#### 2.4.1.1 Son développement

Le bacille Calmette et Guérin (du nom de leurs inventeurs, médecin et vétérinaire de formation) est un vaccin vivant atténué qui confère une protection croisée à la souche humaine, responsable de la tuberculose (274). Il se compose d'une mycobactérie de source et de souche animale, *Mycobacterium tuberculosis bovis* (275).

La souche humaine de la tuberculose nommée *Mycobacterium tuberculosis* a été isolée pour la première fois par Robert Koch en 1882 (274, 275). La tuberculose, dont la forme la plus courante est une infection pulmonaire, évolue vers une forme latente; elle peut également avoir une forme méningée et miliaire (275).

Benjamin Weill-Hallé and Raymond Turpin utilisèrent le vaccin au BCG pour la première fois en 1921 afin de protéger un enfant de 3 jours dont la mère était morte après l'accouchement, de la forme pulmonaire de la tuberculose (274, 275). L'enfant gardé par sa grand-mère, elle-même infectée, a également reçu une dose au 5<sup>ème</sup> et 7<sup>ème</sup> jour et il ne développa pas la maladie (275). Les docteurs Calmette et Guérin ayant déjà testé à cette époque le vaccin sur l'animal, confièrent à ces deux médecins le soin de commencer les tests sur l'homme (276). De 1922 à 1924, ces derniers vaccinèrent ainsi leurs 200 puis 5000 premiers nourrissons, suivis jusqu'en 1928 de 100 000 autres, ce qui permettra aux docteurs Calmette et Guérin de démontrer et publier les données épidémiologiques de son efficacité et de son innocuité en 1928 (276).

Depuis les premiers travaux des docteurs Calmette et Guérin en 1908 à l'Institut Pasteur de Lille, la souche initiale du vaccin au BCG a circulé dans différents laboratoires, et a fortement muté à cause de différentes atténuations (274). La perte de certaines sections génomiques (la délétion) explique la subdivision en une dizaine de souches différentes du vaccin comme la perte de la région RD1 (*region of difference*) pour la première souche développée par Calmette et Guérin (274).

En 1925, le Dr. Joseph-Albert Baudoin, professeur d'hygiène de la faculté de médecine de l'Université de Montréal propose au comité consultatif sur la recherche en tuberculose de réitérer les expérimentations des Drs Calmette et Guérin sur des nouveau-nés québécois (277). Le Dr Petit, de l'Institut Pasteur de Paris, ramène alors la souche de France et le Dr. Baudoin est alors chargé

d'en vérifier l'innocuité et l'efficacité (277). Ses travaux effectués entre 1925 et 1940 seront salués dans les Rapports et documents de l'Institut Pasteur en 1932 et par différentes publications scientifiques (277).

En octobre 1932, le Dr Armand Frappier alors jeune diplômé en médecine, se rend en France à l'Institut Pasteur de Paris où il apprend en 3 mois, directement auprès de ces maîtres Calmette et Guérin, les secrets de l'atténuation de ce vaccin vivant (278). En 1933, il ramène sa première souche vaccinale sur laquelle il commence à travailler à l'Institut de microbiologie et d'hygiène de l'Université de Montréal jusqu'à la fin des années 40 (278, 279). Les atténuations successives et la perte des régions RD1, RD2 et RD15 caractérisent la souche Frappier (274).

#### **2.4.1.2 Programme de vaccination BCG au Québec**

À partir de 1949 débute un programme de vaccination gratuit, organisé sur la base du volontariat dans la province de Québec sous l'initiative du Dr Armand Frappier (279). Ce programme de vaccination a progressivement été éliminé à partir de 1975 et a officiellement pris fin en 1976, selon les autorités de santé publique (280). Il fut donc en vigueur de 1949 à 1976. L'hétérogénéité des études montrant l'efficacité du vaccin au BCG sur la tuberculose humaine et la circulation endémique de la bactérie dans certaines régions du monde comme l'Europe ou l'Amérique du nord, ont poussé les autorités de santé publique à réviser la vaccination systématique et à privilégier les populations à risque (281). Malgré l'abandon de cette vaccination systématique chez le nouveau-né, il n'y a pas eu de recrudescence épidémique de la tuberculose dans ces régions (281).

Ce programme de vaccination au BCG concernait tous les nouveau-nés, les enfants d'âge scolaire jusqu'à leur 11<sup>ème</sup> année (2<sup>ème</sup> cycle, 5<sup>ème</sup> secondaire) et les populations à risque de tuberculose (282). Un test à la tuberculine par scarification était préalablement réalisé sur les sujets, sauf les nouveau-nés, afin de s'assurer de leur négativité (282). La vaccination des enfants d'âge scolaire était menée par des équipes volantes envoyées dans les écoles de la province.

Entre 1956 et 1975, la moyenne des proportions d'enfants vaccinés annuellement avant l'âge d'un an était de 43,8% (IC 95% : 42.8-44.9) alors que cette proportion moyenne était de 6,9% (IC 95% : 6,5-7.4) chez les enfants âgés de 5-9 ans (282). Les autres âges de vaccination étaient minoritaires.

Entre 1970 et 1974, cette moyenne de vaccination annuel augmente même jusqu'à 46% pour les enfants vaccinés avant l'âge de 1 an, alors que ce même moyenne reste stable pour les enfants vaccinées entre 5 et 9 ans (282).

#### **2.4.1.3 Registre de vaccination au BCG**

Le registre de vaccination au BCG (1926-1992) est né sous l'égide du Professeur Baudoin en 1926 et a été repris par le Dr Frappier en 1933 au retour de ce dernier de l'Institut Pasteur de Lille (279). Il rassemble plus de 4 millions d'événements de vaccination au BCG. Ces informations ont été colligées à partir des certificats originaux de vaccination et conservées initialement sous la forme de microfilms (1926-1955) et de fiches papiers (1956-1992) jusqu'en 1992 (Figure 5) (279).


**GOUVERNEMENT DU QUÉBEC**  
 MINISTÈRE DES  
 AFFAIRES SOCIALES

**CERTIFICAT DE VACCINATION ET IMMUNISATION**

BCG

Nom de famille										Prénom														
Date de naissance					Sexe					Prénom du père					Nom de fille de la mère									
An.		Mo.		Jr.		F <input type="checkbox"/>		M <input type="checkbox"/>			sinon													
Adresse du vacciné										Code municipal														
Vaccinations antérieures					Épreuve pré-vaccinale					Date					réaction									
1. dernière		2. avant-dern			Tine (T) <input type="checkbox"/>		P.P.D. (Mantoux) (S) <input type="checkbox"/>			An.		Mo.		Jr.		mm		68		69				
61 An.		64			65		Heaf (H) <input type="checkbox"/>			CutI BCG (C) <input type="checkbox"/>			66		67		19							
Vaccination actuelle										La personne vaccinée est:										Contacts tuberculeux				
No emulsion										étudiant (e) en médecine <input type="checkbox"/>										73				
Type										étudiant (e) infirmier (e) <input type="checkbox"/>										74				
scarification (S) <input type="checkbox"/>										personnel hospitalier <input type="checkbox"/>										oui <input type="checkbox"/>				
intradermique (I) <input type="checkbox"/>										Nom du service de santé (CLSC, DSC, etc.)										non <input type="checkbox"/>				
19										75										77				
19										Nom de l'opérateur														

RT - 26



**Figure 2.5 : Photos d'un certificat de vaccination au BCG vierge (conservé sous forme de microfilm 1926-1955 et sous forme papier 1956-1992) et des listes de vaccination en papier (1956-1992).**

## 2.4.2 Protection contre les tuberculoses

Les réponses immunitaires humaines au BCG ne sont à ce jour pas complètement comprises (37, 39). Les premières hypothèses immunologiques expliquaient la protection croisée contre la tuberculose par les seules voies, cellulaire Th1/CTL (37) puis plus tardivement par la voie humorale B/Immunoglobulines (283). A ce jour, les réponses du vaccin ont été précisées. Elles seraient à la fois de type cellulaire avec des lymphocytes Th1 (IFN- $\gamma$ ) (37), Th17 (IL-17) et CD8+ (37) et humorale (37, 283), toutes deux spécifiques d'antigènes communs à *Mycobacterium tuberculosis bovis* et *Mycobacterium tuberculosis*. L'inoculation du vaccin permettrait ainsi de répondre rapidement à l'infection de *Mycobacterium tuberculosis* en ciblant des antigènes communs avec les deux bras armés de l'immunité adaptative (Th1/th17-CTL et cellules B-plasmocytes-Immunoglobulines spécifiques). Comme toutes infections ou tous autres vaccins vivants atténués, l'inoculation du BCG induit également des voies régulatrices permettant de neutraliser les voies humorale et cellulaire de l'immunité adaptative (37).

L'efficacité du BCG sur la tuberculose est variable selon la forme (pulmonaire, miliaire ou méningé), l'âge d'inoculation du vaccin, l'existence d'un programme de vaccination systématique et d'un contrôle préalable par test cutané à la tuberculine avant inoculation, et la distance à l'équateur du pays dont l'étude est issue (37, 284). Succinctement, plus l'enfant est vacciné précocement et plus la protection semble prononcée. Lorsque les nouveau-nés sont vaccinés à la naissance, cette protection serait supérieure à 50% sur la forme pulmonaire (RR= 0,41; IC à 95% : 0,29-0,58); avec un test tuberculinique négatif en amont, écartant toutes contaminations antérieures par des mycobactéries environnementales, cette protection avoisinerait même les 75% (RR = 0,26; IC à 95% = 0,18 ; 0,37) (284). Lorsque la vaccination a lieu plus tard dans l'enfance, en ayant toujours un test tuberculinique négatif en amont, cette protection serait de seulement 12% (RR= 0,88; IC à 95%, 0,59-1,31) (284). Enfin la protection contre les formes miliaire et méningé serait quasiment égale à 100% (RR= 0,08; IC à 95% : 0,03-0,25) lorsque vacciné plus tard dans l'enfance, en ayant toujours un test tuberculinique négatif en amont (284). L'efficacité est indépendante de la souche vaccinale (37, 284). Cette protection se perpétue à long terme et avait initialement été évaluée à 10-15 ans avec une seule injection (285). Néanmoins, de plus récentes évaluations rapportent une très bonne efficacité de l'ordre de 40 à 70% chez des personnes âgées de plus de 45 à 50 ans (37, 286), toujours avec une seule injection.

### **2.4.3 Les protections hétérologues**

La vaccination au BCG ne protège pas seulement de la tuberculose. Dans les pays où cette maladie a encore une circulation épidémique ou endémique, le BCG (systématisée par des programmes de vaccination) s'est également révélé efficace pour réduire la mortalité infantile toutes causes confondues notamment par son action sur d'autres infections. Cette protection est appelée hétérologue car elle agit sur des agents pathogènes différents du *Mycobacterium tuberculosis bovis* et de *Mycobacterium tuberculosis*, mais aussi sur d'autres maladies non-vectorielles comme le cancer, l'auto-immunité et les diabètes.

Afin de prouver ces actions hétérologues variées, les scientifiques sortent parfois du cadre pur de la vaccination néonatale avec une seule dose intradermique pour vie entière. Ils réalisent le plus souvent des essais cliniques thérapeutiques ou expérimentations animales qui démontrent les mécanismes immunologiques de ce vaccin, sans toutefois être totalement extrapolables au cadre vaccinal comme le sont des essais vaccinaux.

Dans cette recherche sur les mécanismes hétérologues du vaccin, un tournant a été opéré par Netea en 2014, lorsqu'il proposa la « trained immunity » (31). Il s'agit d'une immunité innée entraînée, boostée par l'injection du vaccin au BCG qui lui apprend et imprime une mémoire immunitaire innée au cœur des cellules myéloïdes (38). Le vaccin agirait donc sur les deux immunités, innée et adaptative, en créant deux apprentissages et deux mémoires distinctes.

#### **2.4.3.1 Diminution de la mortalité infantile et protection contre les infections**

Le premier exemple pour introduire la protection hétérologue du vaccin au BCG contre d'autres infections est celui de la lèpre, causée par *Mycobacterium leprae*. Étant de la famille des mycobactéries, il est aisé de comprendre qu'une réactivité croisée à des antigènes communs réamorcent la mémoire lymphocytaire induite par le vaccin (cellulaire : Th1/Th17-CTL et humorale : cellules B-IgM puis IgG). Son efficacité semble comparable à son action antituberculeuse, de l'ordre de 57%, selon une méta-analyse récente (287). Cette efficacité varie également selon les mêmes variables précédemment citées pour son action antituberculeuse (287).

De nombreux et récents essais vaccinaux randomisés témoignent également d'une diminution de la mortalité infantile toutes causes confondues (75) et notamment sur des agents infectieux

d'origine virale. Citons, par exemple, Biering-Sørensen et al. qui montrent une réduction de 40% de la mortalité infantile pour des causes infectieuses (MRR, 0,62; IC à 95%, 0,46–0,83) (288); l'essai contrôlé randomisé de Arts et al. montrant une protection contre le virus de la fièvre jaune (289) et celui de Podder et al. sur le papillomavirus humain (290). Comment une mémoire immunitaire adaptative spécifique pourrait-elle seule contrecarrer le virus de la fièvre jaune (289), du papillomavirus humain (290), mais aussi du virus de l'hépatite B (75), ou encore Influenza A (75) (tous deux démontrés chez l'animal) ? L'immunité innée pourrait agir de façon non spécifique via les *Toll-like* récepteurs contre tous ces agents pathogènes et ainsi expliquer ces protections (86).

Les cellules innées, entre-autres monocytes et macrophages résidents, les neutrophiles circulants (leurs précurseurs) et les cellules dendritiques, présentent en effet de très nombreux récepteurs de reconnaissance de motifs moléculaires d'agents pathogènes (exemples : les Toll-like récepteurs, les NOD-like récepteurs, ou les récepteurs de mannose) qui permettent de signaler le danger et de répondre rapidement par une réponse pro-inflammatoire conséquente (exemples : IFN- $\gamma$ , TNF- $\alpha$ ) (86). Récemment, il a été démontré que plusieurs récepteurs de reconnaissance de motifs moléculaires s'expriment également sur des cellules du système immunitaire adaptatif (cellules B et T), qui semblent également capables de reconnaître les signaux de danger. Néanmoins au lieu d'induire l'activation d'une cascade de signalisation intracellulaire conduisant à la synthèse de cytokines pro-inflammatoires, la stimulation de ces récepteurs permettrait une maturation des lymphocytes T et B (291-293).

Le BCG induit une très forte inflammation de type IFN- $\gamma$  et TNF- $\alpha$  à son point d'injection intradermique, sans commune mesure avec d'autres vaccins (37). Elle se répercute dans les ganglions locaux puis à l'ensemble de l'organisme, y compris la moelle osseuse qui héberge les cellules progénitrices de ces cellules myéloïdes (37). Cette inflammation éduque les cellules innées qu'elles soient résidentes ou circulantes. Cependant, leur durée de vie est faible et ne peut expliquer la persistance à long terme d'une mémoire immunitaire innée. Ce sont les cellules progénitrices de ces cellules myéloïdes qui subissent par cette inflammation conséquente, des changements épigénétiques et métaboliques (du glucose (37, 43)). Ces changements épigénétiques s'imprimeraient durablement, même 50 ans après l'injection (dans un cadre vaccinal) (37). Ils induisent par la suite des niveaux de synthèse conséquents de cytokines pro-inflammatoires (37-39, 43, 289), chez l'homme et l'animal, qui permettent lors d'infections ultérieures, contre tous

agents étrangers, de répondre par une inflammation précoce, massive et soutenue. Secondairement, ce type d'inflammation massive permettrait la différenciation et le recrutement de cellules lymphoïdes innées (exemple : ILC-2) et de cellules Natural killer qui aideraient dans la clairance de ces infections (37, 39).

#### **2.4.3.2 Protection contre le cancer**

Hormis l'immunité entraînée, le vaccin au BCG induit de fait plusieurs réponses adaptatives polyfonctionnelles qui permettent une protection avérée contre le cancer.

L'injection thérapeutique intravésicale du vaccin au BCG, en induction puis en traitement de maintenance, démontre depuis plus de 40 ans, une meilleure survie des patients atteints du cancer de la vessie (HR=0,52; IC à 95% : 0,43-0,63 sur la récurrence et HR=0,74; IC à 95% : 0,57-0,96 sur la progression de la tumeur en traitement de maintenance par rapport à une seule injection d'induction) (294). Le BCG induit donc une immunostimulation qui permet d'activer des lymphocytes T contre le BCG et contre la tumeur cancéreuse (294). En plus des Th1 et LT CD8+, première voie cellulaire décrite, le BCG induit également la différenciation de Th17 (et la synthèse d'IL-17) qui vont potentialiser l'action antitumorale des Th1 et des LT CD8+ (294). Néanmoins, ceci n'est qu'un exemple d'action sur l'immunité adaptative induite par le BCG dans le cadre de la thérapie de ce cancer, dont les mécanismes exacts ne sont toujours pas élucidés (294).

En parallèle de son action antitumorale, l'injection thérapeutique intravésicale du vaccin au BCG induit également la production locale de T Reg conventionnel (FoxP3+) et de LT CD4+ circulants, dont la plasticité leur permet une différenciation en T Reg de novo exprimant PD-L1 (45). Cette expression phénotypique a la capacité de lier le point de contrôle immunitaire PD-1 (ou B7) et d'inactiver des cellules T activées (45).

Ces voies sont ici démontrées dans un cadre thérapeutique et non vaccinal. Pourtant, ces réponses régulatrices ne sont pas spécifiques de cette injection thérapeutique. Cette capacité d'induire des réponses régulatrices, en parallèle de voies adaptatives effectrices spécifiques (LT CD4+, LT CD8+ et LB) aux peptides du BCG, se rencontre également dans la protection croisée vis-à-vis de la tuberculose (37).

Ces voies régulatrices pourraient-elles neutraliser l'amorçage de l'auto-immunité ? et seraient-elles suffisamment pérennes pour le faire ? surtout dans un cadre vaccinal où une seule injection doit neutraliser des processus auto-immuns survenant des années voire des décennies plus tard ?

### **2.4.3.3 Protection contre l'auto-immunité**

Dans le cas de la sclérose en plaques, la voie PD-L1/PD1, précédemment évoquée, pourrait potentiellement expliquer l'extinction de cellules T et B autoréactives ainsi que la modulation de leurs actions par des T Reg conventionnels ou induits selon la récente revue de processus expérimentaux de Mi et al. (45, 295). A ce jour, rien dans notre recension de la littérature ne permet d'extrapoler de tels résultats thérapeutiques et expérimentaux à de potentiels effets induits par une vaccination unique menée pendant l'enfance.

Dans le cadre du diabète de type 1 et de la sclérose en plaques, de nombreux essais cliniques ont également été menés sur l'Homme, avec des injections parfois répétées et un recul limité (40, 296, 297). Ces essais concluent à une protection efficace du vaccin au BCG contre l'évolution de l'auto-immunité. Là encore, ces résultats thérapeutiques sont difficilement extrapolables à un cadre vaccinal.

Rappelons de plus, qu'à ce jour, aucune étude épidémiologique d'association entre la vaccination au BCG et les survenues de ces deux maladies auto-immunes n'a clairement établi de lien (sclérose en plaques (54-59), regroupées dans la méta-analyse de Farez et al. (60), et le seul diabète de type 1 (61-67)).

Pourtant, dans le cadre de leurs essais cliniques sur l'auto-immunité du diabète de type 1 et de la sclérose en plaques (40), Ristori et Faustman tentent d'apporter des réponses d'un point de vue mécanistique. Ils proposent ainsi un potentiel et hypothétique mécanisme régulateur induit par une cytokine pro-inflammatoire, le TNF- $\alpha$  (40). D'après eux, cette cytokine dispose en effet de deux récepteurs TNF-R1, plus susceptible d'être porté par les LT autoréactifs, et TNF-R2, plus susceptible d'être porté par les T Reg (40). Ils supposent donc qu'en fonction de sa concentration, l'un conduit à la transduction d'un signal intracellulaire via TNF-R1 qui conduit à la mort des LT autoréactifs; et le second, via TNF-R2, induit une survie des T Reg (40, 296, 298). Cette action régulatrice à la particularité d'être mis en œuvre par une cytokine pro-inflammatoire.

Dès lors, les cytokines pro-inflammatoires pourraient-elles être protectrices vis-à-vis de l'auto-immunité ? Une inflammation accrue d'origine myéloïde causée par l'immunité entraînée pourrait-elle être le mécanisme qui explique la protection observée lors des essais cliniques ? et qui pourrait potentiellement être observé dans un cadre vaccinal et d'étude populationnelle ?

Arts est un fervent opposant à cette idée (299). Pour cet auteur, l'immunité entraînée n'est que le terreau des maladies auto-immunes. De plus, outre l'immunité entraînée, le vaccin au BCG induit des réponses Th1, Th17 et LT CD8+ tout aussi pro-inflammatoires. Cela semble donc contre-intuitif comme mécanisme protecteur vis-à-vis des processus auto-immuns. Rappelons en effet que la sclérose en plaques et les diabètes auto-immuns sont dus à un débalancement de l'inflammation (exemples : Th1, Th17 et LT CD8+ activés de façon autoréactive) versus la régulation (exemples : T Reg et B Reg).

Pourtant, toute inflammation a besoin d'être régulée sinon elle pourrait finir par détruire l'environnement dans lequel elle a été créée (300, 301). Dans cette idée, si l'immunité entraînée implique la synthèse massive de cytokines pro-inflammatoires, ces seuils de synthèse élevés ne pourraient-ils pas eux-mêmes induire par rétrocontrôle des voies régulatrices exacerbées ? C'est là un mécanisme hypothétique qui pourrait à la fois expliquer les résultats thérapeutiques des essais cliniques, et qui serait suffisamment pérenne pour être observé dans un cadre vaccinal, des décennies après une injection, puisque enregistré dans la mémoire immunologique innée.

#### **2.4.3.4 Changement de la phosphorylation oxydative vers la glycolyse aérobie**

Les changements épigénétiques de l'immunité entraînée et l'activation des voies adaptatives précédemment décrites s'accompagnent également d'un changement de métabolisme cellulaire des cellules innées et adaptatives (37, 39, 41-43). Citons par exemple pour les voies régulatrices adaptatives la modulation de iFOXP3 des iTReg, ou expression de CTLA-4, point de contrôle immunitaire permettant de modérer l'action de lymphocytes T activés de façon autoréactive (37, 39, 41-43). Les cellules immunitaires innées et adaptatives utilisent en effet préférentiellement la phosphorylation oxydative du cycle de Krebs (dans les mitochondries) pour phosphoryler l'ADP en ATP et ainsi fournir de l'énergie (42). Cette première voie utilise peu de glucose. Or le vaccin au BCG, utilisé en thérapie contre le diabète de type 1 ou la sclérose en plaques (40) ou de manière

expérimentale chez des animaux (41-43), induit toujours un changement de métabolisme cellulaire vers une glycolyse aérobie qui est soutenue et consommatrice de glucose.

Kühnreiter et al. ont même prouvé chez la souris NOD que cette glycolyse aérobie permettait bien la consommation accrue de glucose et donc la réduction de l'HbA1c, même en l'absence d'auto-immunité (HbA1c : l'hémoglobine glyquée est un dosage sanguin permettant de traduire l'homéostasie glycémique sur une période de trois mois) (41, 42). Néanmoins le recul maximum observé de cette action hétérologue expérimentale n'est que de 8 années, qui plus est pour des injections répétées (41).

Une seule injection en bas âge pourrait-elle garantir ces mêmes effets expérimentaux ? Actuellement rien ne le démontre. Cependant, une telle vaccination au BCG avec ces mêmes effets observés en population, pourrait dès lors devenir très intéressante pour toutes les personnes obèses et/ou atteintes d'hyperglycémie chronique comme les diabètes de type 1, de type 2 et LADA.

En conclusion, les protections hétérologues du vaccin au BCG n'ont pas encore livré tous leurs secrets. Cependant, les littératures tant épidémiologiques qu'immunologiques témoignent de l'intérêt grandissant pour ce vaccin ayant un siècle d'existence (37, 39). Elles montrent en effet une très bonne aptitude de ce dernier à protéger le nourrisson contre de multiples infections sur le long terme mais qu'en est-il pour la protection de l'auto-immunité massivement montrée à court terme de façon clinique ou expérimentale ? Les changements épigénétiques qui accompagnent les apprentissages de l'immunité entraînée et l'activation des voies adaptatives permettraient de plus de favoriser la glycolyse aérobie et ainsi d'augmenter l'utilisation du glucose. Néanmoins ce seul mécanisme métabolique, très courant lors de l'activation des cellules immunitaires, peut-il à lui seul lutter contre l'hyperglycémie chronique de tous les diabètes, qui plus est à long terme et dans un cadre vaccinal ?

### 3 Problématique

---

Les études épidémiologiques menées à ce jour mesuraient l'association entre la vaccination au BCG et les survenues de sclérose en plaques et du seul diabète de type 1, dans l'enfance et l'adolescence. Toutes présentaient des lacunes méthodologiques importantes comme une taille d'échantillon ou une période de suivi toutes deux limitées, un ajustement partiel pour les facteurs de confusion et/ou l'utilisation de données auto-rapportées pour mesurer l'exposition principale. De plus, l'exposition principale ne prenait pas toujours en compte l'âge à la vaccination qui pourrait induire des apprentissages immunologiques différents et donc se traduire épidémiologiquement par des associations différentes selon ces âges d'exposition au BCG. La modification potentielle des effets selon le sexe n'a pas non plus été systématiquement étudiée. Les hormones sexuelles jouent en effet un rôle prépondérant dans les réponses immunitaires. Enfin, aucune de ces études ne différenciait les phénotypes cliniques de la sclérose en plaques ou du diabète, ni leur âge de survenue, pourtant très aidants dans la différenciation phénotypique de ces maladies. Ces lacunes laissent planer de nombreux doutes car ces études rapportaient toutes une absence d'association avec la vaccination BCG, en dépit des effets hétérologues immunomodulateurs et métaboliques décrits pour ce vaccin dans les expérimentations animales et essais cliniques.

Afin de pallier ce manque de connaissances, il est nécessaire de combler les lacunes méthodologiques des précédentes études par des études d'envergure ayant une puissance statistique et un suivi de cohorte adéquats ainsi qu'un ajustement le plus complet possible pour atténuer la confusion. L'exposition principale doit être étudiée selon l'âge à la vaccination et non seulement selon le statut vaccinal des sujets. L'utilisation d'un registre de vaccination fiable doit être privilégiée afin de minimiser les erreurs de classification, contrairement à certaines études ayant utilisé de l'information auto-rapportée. L'effet du sexe sur les associations doit aussi être systématiquement évalué car les réponses/apprentissages immunitaires du vaccin au BCG pourraient être différentes. Enfin, les différenciations phénotypiques des maladies d'intérêt doivent être formalisées lorsque cela est possible à partir des données pharmaceutiques. Dans le cas de la sclérose en plaques et des diabètes, les médicaments sont hautement spécifiques de la forme clinique pour laquelle ils sont prescrits. Or, les formes cliniques n'ont pas été précédemment

étudiées (exemples : diabète de type 1 et 2 à l'âge adulte, et le diabète LADA) et doivent faire l'objet d'études.

#### **4 Objectifs de recherche et hypothèses**

---

La vaccination au BCG en jeune âge permettrait un apprentissage non spécifique des immunités innée et adaptative (28-39) ainsi qu'une utilisation accrue du glucose dans les cellules immunitaires (40-43). Ces mécanismes hétérologues pourraient limiter le déclenchement et la chronicisation de l'auto-immunité et de l'hyperglycémie. Ce projet doctoral s'intéressera donc à sa potentielle protection vis à vis de ces deux mécanismes sur les survenues de sclérose en plaques et de diabète en caractérisant les phénotypes de ces deux maladies lorsque les informations disponibles le permettent. Les limites méthodologiques principales des précédentes études seront palliées par l'utilisation d'une vaste cohorte de naissance québécoise à laquelle ont été intégrées les informations du registre provincial de vaccination au BCG. Le projet se décline en trois objectifs de recherche principaux :

**Objectif 1:** Estimer les associations entre la vaccination au BCG (ainsi que l'âge à la vaccination) et l'incidence de la sclérose en plaques, globalement puis selon le sexe.

- a) Dans un premier temps, étudier ensemble toutes les formes cliniques de sclérose en plaques;
- b) Dans un deuxième temps, effectuer les analyses séparément pour la forme récurrente-rémittente et les autres formes (e.g : progressive primaire, progressive secondaire).

**Objectif 2 :** Mesurer les associations entre la vaccination au BCG (ainsi que l'âge à la vaccination) et l'incidence du diabète de type 1 à l'adolescence, globalement puis selon le sexe.

**Objectif 3 :** Estimer les associations entre la vaccination au BCG (ainsi que l'âge à la vaccination) et les diabètes de type 1, type 2, et LADA à un âge adulte, globalement puis selon le sexe.

#### **Hypothèses de recherche :**

Sur la base d'études immunologiques, nous postulons qu'il existe des associations inverses entre la vaccination au BCG (particulièrement en très jeune âge) et le développement de l'auto-immunité et/ou de l'hyperglycémie chez l'Homme. Ces associations devraient être observables dans une cohorte populationnelle bien menée et d'envergure, avec des expositions sans biais d'information. Plus particulièrement, ces associations inverses seraient observables, à des âges pédiatriques et

adultes, avec les différents phénotypes de sclérose en plaques, notamment la forme récurrente-rémittente (la plus fréquente) et les diabètes de type 1, 2 et LADA. Ces associations inverses pourraient différer de plus selon le sexe (46) et l'âge à la vaccination (47, 48). De nombreux essais cliniques récents ont en effet révélé que l'action non spécifique du BCG différerait chez les filles et les garçons (46). Les effets bénéfiques chez les garçons apparaissent rapidement au cours de la première semaine après la vaccination puis s'atténuent rapidement avec le temps. Chez les filles, l'effet bénéfique commence plus tardivement après la vaccination que chez les garçons mais se prolonge dans le temps plus tardivement se traduisant par une baisse de l'incidence des infections respiratoires (46). Administré très tôt dans l'enfance, ce vaccin permettrait également une meilleure stimulation et une éducation hétérologue des cellules innées (47, 48) que lors d'une vaccination plus tardive. Néanmoins, ces résultats ont été montrés dans un cadre clinique et/ou expérimental mais jamais dans un cadre vaccinal conventionnel (une injection vaccinale en bas âge) et une cohorte populationnelle.

## 5 Méthodes épidémiologiques

---

### 5.1 Schéma, population d'étude et couplage probabiliste

Les trois articles originaux de ce doctorat sont des études longitudinales menées sur une cohorte rétrospective. Pour ce faire, nous utilisons une nouvelle agrégation de la Cohorte de naissance québécoise sur l'immunité et la santé (*Quebec Birth Cohort on Immunity and Health*, QBCIH) créée par la Pr. Rousseau. Cette cohorte regroupe l'ensemble des personnes nées au Québec entre 1970 et 1974, après au moins 32 semaines de grossesse. Les très grands prématurés (avant 32 semaines) ont été exclus des 5 années de naissances considérées en raison de mortalité et morbidité importantes de ces enfants (302). Une première édition de cette cohorte a été initialement menée sur la seule année de naissance 1974 (N=81,496) (303), dans le cadre d'études de moindre taille portant sur l'asthme et le diabète de type 1 (63). La cohorte utilisée pour le présent projet comprend 400,611 individus, représentant 90,4% des sujets éligibles de ces 5 années de naissance. Un couplage probabiliste réalisé sur différentes sources de données médico-administratives et pharmaceutiques québécoises a permis d'assembler cette cohorte.

Des liens probabilistes ont été réalisés par l'Institut de la statistique du Québec, à l'aide de logiciels développés par Statistique Canada en commençant par le registre des naissances et le fichier des personnes assurées (FIPA) de la Régie de l'assurance maladie du Québec (RAMQ). Les deux logiciels créés par Statistique Canada sont : les logiciels *Generalized Record Linkage System* et *G-Link* version 2.4 (304). La technique de couplage utilisée pour agréger l'ensemble de ces sources de données est une technique probabiliste, préalablement décrite par Sayers et al. (305). Pour six variables, les prénom et nom de famille, la date de naissance et le sexe de l'individu, ainsi que le prénom de son père et de sa mère, un poids de correspondance global a été calculé entre les fichiers. Cette technique probabiliste permet de prendre en compte la concordance partielle entre les champs. En effet, ces champs ont été initialement colligés manuellement par des infirmières et des erreurs sur ces variables ont pu être introduites. Afin d'augmenter la probabilité de correspondance, des procédures de normalisation ont été entreprises en amont. Parmi elles, citons, une transformation en majuscules de toutes les chaînes de caractères des champs nom et prénom, une suppression de l'accentuation, une suppression des titres de civilité comme Mme ou M., l'uniformisation des Mac et Mc, et la suppression systématique des blancs en amont ou aval de la chaîne de caractères. Les prénoms et noms composés étaient également systématiquement

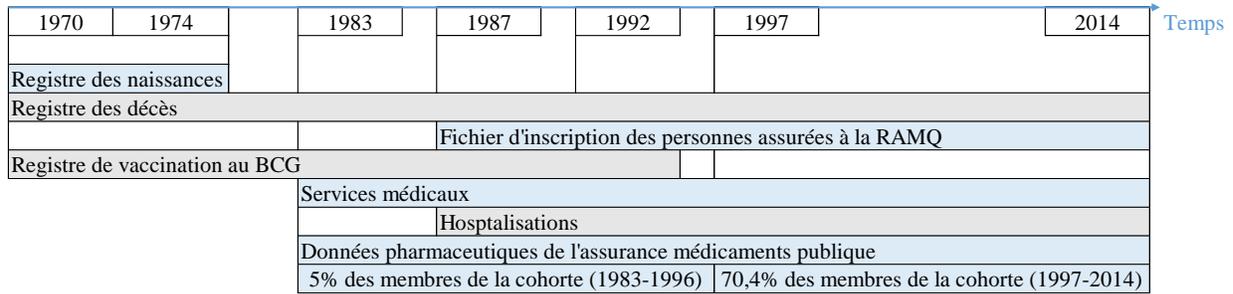
décomposés en plusieurs variables atomiques uniques afin d'être comparés. Pour les dates de naissance, cette normalisation transformait tous les champs en format AAAA/MM/JJ (Année, Mois, Jour). Enfin, le sexe était normalisé en F et H pour femme et homme. Néanmoins, une dissimilarité entre tous ces champs, pouvait encore être acceptée par exemple dans l'orthographe des prénoms et noms de famille entre deux jeux de données. Ce poids de correspondance global a été calculé à chaque paire d'enregistrements entre les deux jeux de données, en fonction de la concordance et de la fréquence des données. Par exemple, un poids plus élevé a été attribué en cas de concordance pour un nom de famille rare par rapport à un nom de famille très courant. Les paires d'enregistrements étaient ensuite rejetées si les poids de correspondance globaux étaient inférieurs à un seuil, acceptés s'ils étaient supérieurs à ce dernier et traités manuellement s'ils étaient intermédiaires (305). Ce couplage probabiliste a permis d'établir la population d'étude par un couplage deux à deux, en commençant par le FIPA et le registre des naissances. D'autres couplages ont ensuite été menés, avec la même technique, pour ajouter les informations provenant des registres de décès et de vaccination au BCG puisque ces deux bases de données ne contiennent pas le numéro d'assurance maladie (NAM) comme identifiant.

En théorie nous disposons de 400,611 individus suivis de 1983 à 2014 pour leur utilisation des services de santé et/ou pharmaceutiques. Néanmoins, en fonction des objectifs de recherche, la population d'étude et la période de suivi ont été définies selon les besoins. Par exemple, seule l'étude sur la sclérose en plaques a utilisé l'ensemble du suivi entre 1983 et 2014 (1<sup>er</sup> article). La période 1985-1993 a permis de circonscrire un suivi de 10 à 18 ans pour l'étude du diabète de type 1 dans l'enfance et adolescence (2<sup>e</sup> article) et la période 1997-2014 a servi à circonscrire un suivi de 22 à 44 ans pour l'analyse des diabètes de type 1, type 2 et LADA à l'âge adulte (3<sup>e</sup> article). Les populations d'étude sont dès lors décrites respectivement dans chaque manuscrit.

## **5.2 Sources de données**

Les sources de données agrégées sont le FIPA de 1983 à 2014, sauf pour les codes postaux à 3 positions disponibles à partir de 1987, les services médicaux (disponibles de 1983 à 2014) et pharmaceutiques de la RAMQ, les données sur les hospitalisations (MED-ECHO, 1987-2014), le registre des décès (1970-2014), le registre des naissances de 1970 à 1974, les données des recensements de Statistique Canada, effectués tous les cinq ans, ajoutées grâce aux codes postaux

à 3 positions (disponibles dans le FIPA) et le registre de vaccination au BCG dont les données prennent fin en 1992. La figure 5.1 résume la disponibilité de l'ensemble de ces données.



**Figure 5.1 : Périodes de couverture des différents registres et fichiers médico-administratifs et pharmaceutiques utilisés dans la Cohorte de naissance québécoise sur l'immunité et la santé, Québec, 1970-2014.**

Les services médicaux et hospitaliers provenaient respectivement des demandes de remboursement des médecins et des fiches d'hospitalisation (MED-ECHO). Les services médicaux contenaient les codes de diagnostic de la 9<sup>ème</sup> Classification statistique internationale des maladies et des problèmes de santé connexes (CIM-9) de 1983 à 2014. Les fiches d'hospitalisation contenaient les codes de diagnostic CIM-9 de 1983 à 2006, puis de la 10<sup>ème</sup> révision de la Classification statistique internationale des maladies (CIM-10) à partir d'avril 2006. L'ensemble des variables de la QBCIH et leur provenance en fonction des fichiers-source sont résumées au tableau 5.1. Le registre des naissances a ainsi permis, outre la date de naissance et le sexe (également contenus dans le FIPA), de relever des variables périnatales telles que le type de grossesse, simple ou gémellaire, le poids de naissance et la durée de gestation. Une variable composite du poids de naissance et de la durée de gestation, le poids de naissance pour l'âge gestationnel (selon Kramer et al. (306)) a été créée (303, 307). Le registre des naissances renfermait également des informations concernant le père et la mère comme les lieux de résidence des parents, leur âge respectif à la naissance de leur enfant et le nombre antérieur de naissances vivantes pour la mère. Cette dernière variable nous a ainsi permis de déterminer le nombre de frère(s) ou sœur(s) plus vieux que le sujet (303, 307).

Le registre des décès contenait l'âge de la personne décédée, la date et la cause médicale du décès.

Les variables utilisées du FIPA comprenaient la date de naissance, le sexe, les codes postaux à trois chiffres, et l'année du décès (le cas échéant).

Pour chaque service médical et hospitalier, la date et le code de diagnostic inhérent à la visite médicale ou au service hospitalier étaient consignés. L'ensemble des variables inhérentes à l'utilisation de ces services médicaux et/ou hospitaliers sont résumées au tableau 5.1. Le compte-rendu d'hospitalisation faisait état du code de diagnostic d'admission, du code de diagnostic primaire (ou principal) à la sortie, ainsi que des codes de diagnostic secondaires, jusqu'à 15 codes différents inhérents aux différents diagnostics et soins apportés durant l'hospitalisation principale.

Les services pharmaceutiques étaient disponibles pour 5% des membres de la cohorte entre 1983 et 1996 et pour 70,4% d'entre eux entre 1997 et 2014. Les périodes d'admissibilité à l'assurance médicaments pouvaient de plus être continues ou discontinues. Les variables inhérentes à ces services sont également résumées au tableau 5.1.

Le registre de vaccination BCG colligeait le statut vaccinal au BCG et l'année de vaccination au BCG.

**Tableau 5.1 : Variables utilisées de la Cohorte de naissance québécoise sur l'immunité et la santé et leur provenance en fonction des fichiers sources, Québec, Canada, entre 1970 et 2014**

Adapté de Rousseau et al. (303).

Source de données	Domaines	Variables
<b>Fichiers (périodes de couverture)</b>		
<b>Institut de la statistique du Québec</b>		
Registre des naissances (1970-1974)	Sociodémographiques	Date de naissance, sexe, lieu de résidence/de naissance
	Naissance	Poids de naissance, âge gestationnel, type de naissance (unique, multiple)
	Famille et parents	Nombre de naissances vivantes/mortinaissances, nombre total d'accouchements, âge des parents à la naissance de l'enfant, province/pays de naissance des parents
Registre des décès (1970-2014)	Décès	Âge de la personne décédée, date et cause médicale du décès
<b>Régie de l'assurance maladie du Québec (RAMQ)</b>		
Fichier d'inscription des personnes assurées (FIPA, 1987-2014)	Démographiques	Date de naissance, sexe, codes postaux à trois chiffres, année du décès (le cas échéant)
Services médicaux (1983-2014)	Utilisation des services médicaux	Clé unique identifiant un individu, date de service, codes de diagnostic CIM-9 entre autres de nos événements d'intérêt (sclérose en plaques et diabète, décrits dans le paragraphe inhérent à chaque maladie)

Source de données	Domaines	Variables
<b>Fichiers (périodes de couverture)</b>		
Hospitalisations (MED-ECHO, 1987-2014)	Données sur l'hospitalisation	Clé unique identifiant un individu, numéro séquentiel d'hospitalisation, dates d'admission et de sortie, type d'admission à l'hôpital, nombre total de jours d'hospitalisation, type de diagnostic hospitalier, codes de diagnostic CIM-9 et CIM-10 à partir d'avril 2006 pour entre autres nos maladies d'intérêt (sclérose en plaques et diabète), version de la CIM utilisée, codes de services médicaux hospitaliers, domaine de consultation, nombre total de consultations, nombre de jours passés dans le service, catégorie et spécialité du professionnel de la santé (service, intervention), date/lieu/type d'intervention, nombre total d'interventions
Services pharmaceutiques (1983-1996: 5% et 1997-20014: 70,4%)	Médicaments	Clé unique identifiant un individu, couverture d'assurance médicaments (dates de début et fin, nombre de jours et années des périodes d'admissibilité, nombre de périodes d'admissibilité et durées), date du service pharmaceutique, durée du traitement, classe AHFS, code de dénomination commune du médicament, code de la forme galénique, quantité de médicament, spécialité du prescripteur, spécialité du professionnel référent
<b>INRS-Institut Armand-Frappier</b>		
Registre de vaccination au BCG du Québec (1970-1992/6)	Information sur la vaccination	Statut vaccinal au BCG, année de vaccination au BCG

AHFS, American Hospital Formulary Service; CIM-9, Classification internationale des maladies – 9ème édition ; Med-Echo, Maintenance et exploitation des données pour l'étude de la clientèle hospitalière; RAMQ, Régie de l'assurance maladie du Québec.

### 5.3 Les expositions – vaccination au BCG et âge à la vaccination

Le registre de vaccination au BCG du Québec qui collige les données du programme volontaire de vaccination (1949-1976) a été préalablement numérisé en 2010 pour la période 1956-1992 (279). Avant numérisation, ce registre était conservé sous deux formats (figure 2.5) : certificats de vaccination sous la forme de microfilms (123 rouleaux, 1926-1955) puis des fiches papier (1956-1992). Les fiches papier étaient elles-mêmes classées sous 2 formes : des certificats de vaccination en papier classés en archives et des listes d'individus où chaque ligne résume une vaccination. Ces lignes colligent les prénom et nom de l'individu, le prénom du père, la date de naissance, le pré-test tuberculinique (type, date et réaction) si ce n'était pas un nouveau-né, les dates des précédentes vaccinations (s'il y en avait), le code de l'institution où est faite cette vaccination. Ces listes représentaient 140 volumes totalisant plus de 60 000 pages et environ 4,2 millions de lignes de données.

Afin d'évaluer la précision et la qualité des informations électroniques enregistrées, l'équipe de la Pr Rousseau a tout d'abord comparé un échantillon aléatoire (randomisé selon les volumes, pages puis lignes de 1956 à 1992) des listes papier (n=5 268, 0,1%) à la version numérisée. Puis, un

second échantillon aléatoire des certificats de vaccination en papier (n=4 972, 0,1%,) contenant approximativement 250 certificats pour chacune des 20 années de 1956 à 1975, a été comparé à la version électronique (279). Pour les deux échantillons, la qualité des données numérisées a été évaluée selon l'exhaustivité des informations enregistrées et la présence de divergences selon des variables nominatives (nom de famille, prénom, sexe, date de naissance, prénom du père) ou reliées à la vaccination au BCG (année/type/réaction du pré-test tuberculinique, année/type/établissement de la vaccination et année(s) de la 1<sup>ère</sup> et 2<sup>ème</sup> vaccination(s) antérieure(s)) (279). Parmi les 5 268 enregistrements sélectionnés du premier échantillon des listes papier, 281 étaient des doublons (5%). Parmi les 4 972 certificats de vaccination papier du second échantillon, 4 709 ont été retrouvés dans la version numérique, soit 94,7%. Pour chacune des variables nominatives ou reliées à la vaccination dans les deux échantillons, il y avait une très bonne concordance, oscillant entre 91,6% et 100%. Seules deux variables, prénom et sexe, dans le second échantillon, présentaient une concordance de respectivement 92,2% (IC95% :91,3 à 92,9) et 91,6% (IC95% : 90,8-92,4). Par exemple lors de la vaccination à la naissance, le prénom sur le certificat de vaccination n'était pas complété pour des enfants dont le nom n'avait pas encore été choisi par les parents. Lors de la compilation des listes, cette information manquante avait été remplacée dans certains cas par Marie/Joseph/Inconnu. Ces modifications systématiques ne représentant pas le prénom réel de la personne ont été notées et corrigées avant d'effectuer le jumelage avec les bases de données. Toutes les autres comparaisons, dans les deux échantillons, présentaient une concordance comprise entre 99% et 100%.

Le registre de vaccination au BCG contient les résultats des tests tuberculiques fait en amont, le statut vaccinal, la date de naissance au registre et l'année de vaccination, essentielles pour calculer l'âge à la vaccination. Le registre collige de plus l'ensemble des vaccinations effectuées, qui pouvaient être multiples pour certains individus. Pour un individu unique, le statut vaccinal et le nombre de vaccinations ont donc été établis en considérant toutes les fiches vaccinales du registre pour ce même individu. Le statut vaccinal par individu permet par la suite de dégager une première variable d'exposition binaire (vacciné/non-vacciné). L'année de vaccination et la date de naissance permettent de calculer un âge à la vaccination, mais en l'absence du jour et mois de vaccination, la seconde variable d'exposition ne peut être établie que selon une variable catégorielle (non-vacciné/vacciné avant 1 an/vacciné après un an d'âge). Enfin, le statut vaccinal incertain était attribué aux individus n'ayant pas une probabilité suffisante lors du couplage probabiliste entre le

registre de vaccination au BCG et les données nominatives de la RAMQ. Ce statut vaccinal a été traité différemment selon les articles. Pour l'article 2, sur le diabète de type 1 dans l'adolescence, ces individus ont été exclus de l'analyse. Pour les articles 1 et 3, portant respectivement sur la sclérose en plaques, et sur les différents phénotypes de diabète à l'âge adulte, ces individus ont été conservés dans l'analyse mais leur statut vaccinal incertain a été remplacé par un statut vaccinal imputé.

Le programme de vaccination du Québec, organisé sur la base du volontariat, permet un équilibre entre les vaccinés et les non-vaccinés et une comparaison aisée, comparativement à un programme de vaccination systématique qui recherche l'exhaustivité de la population cible.

## **5.4 Définition des maladies d'intérêt : sclérose en plaques et diabète**

### **5.4.1 Sclérose en plaques**

Pour la sclérose en plaques, la première définition utilisée a été celle de Marrie et al. (308-310), couramment appelée définition de Marrie. Elle identifie comme cas incident de sclérose en plaques, indépendamment de son phénotype clinique, tout individu ayant eu au moins 3 hospitalisations ou visites médicales pour la sclérose en plaques au cours du suivi. Cette définition a été appliquée à l'ensemble de l'échantillon éligible de l'étude (N=400,563) pour détecter en premier lieu l'ensemble des cas incidents de sclérose en plaques.

Le code de diagnostic de sclérose en plaques utilisé pour les visites médicales est le code CIM-9: 340 de 1983 à 2014. Les codes de diagnostic utilisés pour les hospitalisations sont successivement le code CIM-9: 340 à partir de 1983 puis le code CIM-10: G35 à partir d'avril 2006 jusqu'en 2014.

Plusieurs définitions médico-administratives existent afin d'identifier cette maladie (308-310). Néanmoins celle choisie de Marrie et al. a été validée à la fois dans des populations pédiatrique (309) et adulte (310). Le suivi de cohorte entre 1983 et 2014, d'individus nés entre 1970 et 1974, permettait en effet la détection de cas pédiatriques ( $\leq 16$  ans) et adultes dans notre population éligible d'étude.

La définition de Marrie a une sensibilité (96 %, IC à 95 % : 92,3-98,3) et une spécificité (99,5 %, IC à 95 % : 97,3 à 100,0) élevées (308-310). Cette définition a tout d'abord été élaborée et validée au Manitoba (310). Des individus catégorisés par cette définition dans les bases de données

médico-administratives du système de santé, ont été comparés à un échantillon aléatoire de 2 000 personnes ayant eu une maladie démyélinisante entre 1998 et 2006 et dont les diagnostics étaient extraits de leur dossiers médicaux (310). Cette première validation n'a cependant été opérée que dans une population adulte âgée de 24 à 65 ans (310). Une seconde étude, en cohorte prospective a été menée par Marrie et al. afin de valider cette même définition dans une population pédiatrique  $\leq 18$  ans (309). Cette étude compare le classement des cas obtenus selon la définition médico-administrative dans les bases de données du système de santé ontarien, au classement médical obtenu au sein de sa cohorte prospective (309). Tous les cas de sclérose en plaques des enfants inclus dans la cohorte ont été avérés par un diagnostic médical étayé par des IRM et établit selon les critères 2010 de McDonald (311), validés pour une utilisation dans la population pédiatrique (312). Cette même définition a enfin été validée sur des populations pédiatrique et adulte en Saskatchewan, en comparant à nouveau le classement de cette définition à partir des bases de données de santé de cette province aux diagnostics de référence établis dans des dossiers médicaux (n= 400) (308).

Une seconde définition, à ce jour non validée, a été développée dans le cadre de ce doctorat, basée sur les données pharmaceutiques de remboursement des médicaments modificateurs d'effet. Ces médicaments autorisés à partir de 1997 au Québec, ne pouvaient jusqu'en 2017 être prescrits qu'à des patients atteints de sclérose en plaques récurrente-rémittente (313). Cette seconde définition nous a donc permis d'identifier les cas de sclérose en plaques récurrente-rémittente, déjà circonscrits comme cas incidents de sclérose en plaques (sans phénotype) par la définition de Marrie mais également d'en identifier de nouveaux, non initialement détectés par cette première définition. Cette seconde définition ne pouvait s'appliquer que sur l'échantillon des individus disposant d'une couverture pharmaceutique publique. Néanmoins, cette couverture n'était pas uniforme sur l'ensemble du suivi (1983-2014). Entre 1983 et 1996, seuls 5% des membres de la cohorte bénéficiaient d'une assurance médicaments publique, alors qu'à partir de 1997 (début d'autorisation des médicaments modificateurs d'effet), cette couverture a été étendue à toute personne n'ayant pas d'assurance médicaments par son emploi. Dans la période 1997-2014, 70,4% des membres de la cohorte bénéficiaient ainsi d'une assurance médicaments publique, les données portant sur la couverture privée du remboursement des médicaments ne nous étant pas accessibles. Dans cette seule période (1997-2014), la seconde définition pharmaceutique a été appliquée afin d'identifier les individus ayant  $\geq 1$  remboursement pour des médicaments traitant exclusivement

la sclérose en plaques récurrente-rémittente. Puisque 50% des phénotypes initiaux de sclérose en plaques récurrente-rémittente restent sous cette forme clinique, les cas incidents détectés par la première définition de Marrie entre 1983 et 1996 et qui étaient par ailleurs consommateurs de ces médicaments sur la période 1997-2014 pouvaient être classés comme cas avérés de sclérose en plaques récurrente-rémittente durant la période 1983-1996.

Pour chaque individu identifié par la définition de Marrie comme cas incident dans la cohorte, la date de diagnostic a été définie comme la toute première date d'un service de soins hospitalier ou médical pour la sclérose en plaques ou pour une maladie démyélinisante préalable basée sur les codes CIM-9/10 à 4 chiffres : encéphalite, myélite et encéphalomyélite (323.4-323.6, 323.8-323.9 / G36.9), neuro-myélite optique (341.0 / G36.0), autres maladies démyélinisantes du système nerveux central (341.8 , 341.9 / G37.8, G37.9), et névrite optique (377.3 / H46). Cette approche a été suggérée et utilisée par les chercheurs ayant validé les algorithmes (99, 308-310) et dans d'autres études menées au Canada (314). Cette même date de diagnostic a été utilisée pour les cas identifiés par les deux définitions à la fois. Pour les cas identifiés uniquement par les médicaments (la seconde définition), la première date de médicament leur a été assignée. Toutefois, ces personnes pouvaient également avoir eu des services médicaux et/ou hospitalisations pour la sclérose en plaques sans toutefois rencontrer la définition. Dans ce cas, la date la plus précoce entre la première date de médicament et la date de service médical ou d'hospitalisation, soit pour la sclérose en plaques ou pour une maladie démyélinisante préalable, a été attribuée comme date de diagnostic.

L'analyse en deux périodes successives, 1983-1996 et 1997-2014, nous a également permis d'isoler les cas de sclérose en plaques selon l'âge de survenue; les formes pédiatriques (<16 ans) et précoces ( $\leq 21$  ans) dans la première période, puis plus tardives, entre 22 et 44 ans, dans la seconde période.

#### **5.4.2 Les diabètes**

Pour les diabètes, deux définitions validées pour l'utilisation de données médico-administratives ont tout d'abord été utilisées pour identifier les cas incidents, tous phénotypes confondus. Les codes de diagnostic utilisés dans le cadre de ces deux définitions sont les codes CIM-9 : 250.X et CIM-10 : E10-14, en fonction de la période et de la source des données considérées.

Dans l'article 2, portant sur le diabète de type 1, cette première définition était celle de Nakhla et al. (70), que nous appellerons définition de Nakhla. Parmi de nombreuses définitions médico-administratives (70, 315), celle de Nakhla identifie les cas de diabète si un individu avait  $\geq 4$  services médicaux pour le diabète sur une période de 2 ans (Sensibilité=90,4%, Spécificité=100%, Valeur Prédicative Positive= 93,2%) (70). Cette définition a été validée en comparant le classement de cette définition à partir des données médico-administratives du système de santé québécois, aux diagnostics de référence établis dans les dossiers médicaux des cliniques qui suivent ces patients diabétiques (70). Cette définition présentait deux avantages : 1) elle n'utilise pas les hospitalisations, seulement disponibles dans nos données à partir de 1987; et 2) elle avait récemment été validée sur la population pédiatrique québécoise (70). Cette définition est identique à celle Guttman et al. (315) précédemment utilisée par notre groupe pour circonscrire le diabète de type 1 (63). Pourtant, Nakhla et al. proposent une différence fondamentale, une séparation entre deux dates de consultation médicale d'au moins 30 jours et non pas un seul jour comme la précédente définition. Pour tous les individus identifiés par la définition de Nakhla, la date de diagnostic a été définie comme la toute première date des 4 services médicaux détectés sur la période de 2 ans. Le suivi de cohorte a été censuré à droite au jour précédant la date de leur 19<sup>ème</sup> anniversaire (au maximum 1993). Cette limitation supérieure de l'âge a permis de circonscrire au mieux les cas de diabètes de type 1, dont le pic d'incidence se situe à l'adolescence (69).

Dans l'article 3, portant sur les diabètes de type 1, type 2 et LADA à l'âge adulte, la première définition utilisée est celle du Système canadien de surveillance des maladies chroniques (CCDSS) pour la détection des cas de diabète. Cette définition a été validée dans la population canadienne adulte afin de détecter le diabète tous phénotypes confondus (316, 317), avec une bonne sensibilité (88,9–89,8%) et une spécificité élevée (97,5–97,7%) (318). Cette définition a été successivement utilisée dans les provinces du Manitoba, de l'Alberta, de la Saskatchewan et de l'Ontario. Là encore, sa validation a été effectuée en comparant les classements de cette définition à partir des données médico-administratives du système de santé ontarien, aux dossiers médicaux de référence (318). Basés sur cette définition, les cas incidents de diabète étaient les individus de la cohorte ayant  $\geq 1$  hospitalisation(s) ou  $\geq 2$  services médicaux pour le diabète sur une période de 2 ans, avec au moins 1 jour de séparation entre deux services de santé (319). Ici, l'utilisation des codes de diagnostic de diabète lors d'hospitalisation n'était pas problématique, puisque notre période d'étude a été circonscrite à l'âge adulte (22-44 ans), soit de 1997 à 2014. Cette période était donc

postérieure à 1987, date de disponibilité des données hospitalières. Afin d'exclure de potentiels cas de diabète gestationnels, la définition originale du CCDSS exclut des cas, toute femme âgée de 10 à 54 ans ayant des codes d'hospitalisation obstétricaux ou reliés à la grossesse (CIM-9/ICD-10 : 641-676, V27/ O1, O21-95, O98, O99, Z37) si ces derniers interviennent dans l'intervalle de 120 jours avant ou 180 jours après la date hospitalière ou de visite médicale utilisée utilisée dans la définition (319). Dans nos données, les codes de complication de grossesse étaient présents mais les codes d'accouchement normal et d'accouchement (ICD-9/ICD-10 : 650, V27/O80, Z37) n'étaient pas systématiquement présents. Cependant, l'utilisation des données des services médicaux plutôt que celles des seules hospitalisations nous a permis d'identifier toute grossesse avec complications. Conformément à la définition du CCDSS, la date de diagnostic a été définie comme la date de sortie d'hospitalisation ou la seconde date de visite médicale (intervenant dans la période de 2 ans).

Cette définition ne permet pourtant pas de distinguer les phénotypes cliniques. Pour ce faire, les diabètes de type 1, type 2 et LADA ont été caractérisés par une seconde et une troisième définition utilisant les données pharmaceutiques de remboursement des antidiabétiques oraux et /ou de l'insuline, ainsi que des délais entre le diagnostic et l'utilisation des premiers médicaments antidiabétiques. Ces deux définitions ont été développées dans le cadre de ce doctorat et ne sont à ce jour pas validées.

La seconde définition, pharmaceutique, de l'article 3 a utilisé l'ensemble de l'échantillon des individus couverts par l'assurance médicaments publique (70,4% des 400 611 individus initiaux de la cohorte ou 65,7% des individus éligibles pour la période d'incidence 1997-2014). Si un individu n'utilisait que de l'insuline sur l'ensemble de son suivi, il était alors catégorisé comme diabétique de type 1. Si un individu n'utilisait que des antidiabétiques oraux ou une combinaison d'antidiabétiques oraux et d'insuline sur l'ensemble de son suivi, il était alors catégorisé comme diabétique de type 2 (insulino-requérant ou non), car l'administration rapide d'insuline est essentielle lors de diabète de type 1. Si ces individus n'étaient préalablement pas détectés par la définition CCDSS, leur date de diagnostic était alors la date de la prise de leur premier médicament. S'ils étaient préalablement détectés, leur date de diagnostic restait celle attribuée par la définition CCDSS.

La troisième définition, basée sur l'information pharmaceutique, a été élaborée sur un sous-échantillon plus restreint (40,5% des sujets éligibles pour la période d'étude 1997-2014). Néanmoins, sur cet échantillon nous disposions des individus ayant une couverture de l'assurance médicaments publique sur  $\geq 90$  % du suivi, ou ayant le remboursement de leurs médicaments antidiabétiques oraux et/ou d'insuline dans leur première période de couverture. Dans cet échantillon, certains des cas de diabète de type 1 ou de type 2 précédemment détectés par la seconde définition ont été re-catégorisés en diabète LADA basé sur la connaissance du délai entre la date de diagnostic (donnée par la définition CCDSS) et la première administration d'insuline, ou du délai entre la prise du premier antidiabétique oral et de la première insuline (si cet individu n'était pas détecté par la définition CCDSS). La date de diagnostic était donc toujours celle donnée par la définition CCDSS ou celle de la première médication, toujours en excluant les cas potentiels de diabète gestationnel autour de ces premières dates.

Pour le calcul des délais de l'initiation de l'insuline pour cette troisième définition, deux courants de pensée s'affrontent. En 2005, l'*Immunology of Diabetes Society* (IDS) a établi que les cas de diabète LADA devaient être âgés de plus de 35 ans et qu'ils devaient recourir à l'insuline au-delà de 6 mois, voire plus, après leur diagnostic (6, 320). Cette limite d'âge, de plus de 35 ans, reste néanmoins débattue (6). Plus récemment, la *Japan Diabetes Society* (JDS), basée sur une comparaison des cas de diabète LADA avec des diabétiques de type 1 aigu, a proposé que l'âge d'apparition des diabètes LADA ne doit pas être restreint aux seuls individus de plus de 35 ans et que le diagnostic de diabète LADA pouvait être posé après une période initiale de 3 mois sans recours à l'insuline (6). Dans l'article 3, nous avons donc utilisé les deux délais susmentionnés pour identifier deux séries de cas de diabète LADA : 1) entre 6 mois et 1 an (définition IDS) et 2) entre 3 mois et 1 an (définition JDS). Les délais avant le recours à l'insuline étant encore fortement débattus (201), le délai maximum a été choisi à un an sur la base des observations cliniques d'une collaboratrice (Dr. H. Mansaray) de notre équipe. Buzzeti et al. a en effet décrit un délai maximal du recours à l'insuline pouvant aller jusqu'à 6 ans (201). Conformément aux évidences les plus récentes, nous n'avons pas appliqué de restrictions d'âge pour la définition IDS, alors que la définition JDS ne comportait déjà aucune restriction de l'âge au diagnostic.

## 5.5 L'analyse de survie et le modèle de régression proportionnel de Cox

Initialement, le terme de « survie » a été utilisé dans la littérature car typiquement l'évènement d'intérêt à surveiller dans une cohorte était le décès. Néanmoins, cet évènement d'intérêt peut également être l'apparition d'une maladie au cours du temps. Pour établir cette survie, il est donc nécessaire de connaître la date d'entrée dans la cohorte (par exemple dans notre cas la date de naissance) et la date de l'évènement considéré à un temps  $t$ . La différence entre ces deux dates définit un temps de suivi noté  $T$ , et la fonction de survie  $S(t)$  au temps  $t$  est la probabilité qu'un individu survive au-delà de la date  $t$ , donc qu'il décède ou qu'il ait l'évènement d'intérêt après cette date.

La fonction de survie s'écrit donc :

$$S(t) = P(T > t) \quad (1)$$

C'est une fonction décroissante de 1 à 0, soit la probabilité de ne pas subir l'évènement au temps  $t=0$ ,  $S(t_0)=1$ , ou inversement la probabilité de subir l'évènement quand  $t$  tend vers  $+\infty$  est  $S(t)=0$ .

Une des méthodes pour estimer cette fonction de survie est la méthode non-paramétrique de Kaplan-Meier (321). Elle permet d'estimer la survie dans un intervalle de temps où les données des sujets (i) débutent à une date fixe, par exemple leur date de naissance, et se termine à une dernière date d'observation (dite de point), après laquelle il n'y a plus aucune information disponible (dans notre cohorte, cette date est le 31 décembre 2014). Les données sont dites censurées à droite à cette date. La date de début de l'intervalle de temps peut également être différente de la date de naissance. Par exemple, dans notre cohorte la disponibilité des informations médico-administratives débute le 1<sup>er</sup> Janvier 1983. Dans ce cas de figure, les données sont dites tronquées à gauche en 1983.

Selon Kaplan-Meier (321), dans l'intervalle de temps  $[t_i, t_{i+1}[$  précédemment décrit, il est donc possible d'observer le nombre  $e_i$  de l'évènement d'intérêt, le nombre  $c_i$  de personnes censurées à droite et le nombre de sujets exposés  $n_i$  au risque de l'évènement juste avant  $t_{i+1}$ .

La probabilité de survie à la date  $t_{i+1}$  sachant que l'individu n'a pas subi l'évènement au temps  $t_i$  est estimée (321) par :

$$P[ti + 1|ti] = \frac{ni - ei}{ni} \quad (2)$$

L'estimation de la probabilité de ne pas subir l'évènement juste après  $t_{i+1}$  est le produit des probabilités conditionnelles de survie :

$$\hat{P}(ti) = \prod_{j=1}^i P [ ti + 1 | ti ] \quad (3)$$

et  $P(ti)$  est une variable aléatoire (321).

Le risque instantané de subir un évènement au temps  $t$  est ici noté  $RI(t)$ .

Il s'écrit (322) :

$$RI(t) = \lim_{dt \rightarrow 0} P [ (T < t + dt) | (T \geq t) ] / dt \quad (4)$$

Ce risque  $RI$  est une probabilité conditionnelle par unité de temps,

- $P [ (T < t + dt) | (T \geq t) ]$  est la probabilité de subir l'évènement après le temps  $t$ , sachant que le sujet ne l'a pas encore subi avant au temps  $t$ .
- $dt$  ou delta  $t$ , le dénominateur est la variation de temps entre  $(t)$  et  $(t+dt)$

$RI(t)$  peut également s'écrire comme la division de la dérivée première de la fonction de survie notée  $-S'(t)$  et de la fonction de survie  $S(t)$  au temps  $t$  (322), tel que :

$$RI(t) = -S'(t)/S(t) \quad (5)$$

$$\Rightarrow S(t) = e^{-\int RI(t).dt} \quad (6)$$

Si le risque instantané  $RI(t)$  de subir l'évènement au cours du temps est constant alors la fonction de survie peut s'écrire  $S(t) = e^{-C.t}$  avec  $C$  et  $t$ ,  $C$  une constante au cours du temps  $t$ , et la fonction  $S(t)$  suit une distribution exponentielle.

Le hazard ratio noté  $HR$  est le rapport du risque instantané entre les individus exposés divisé par le risque instantané parmi les individus non-exposés. Le  $HR$  est lui-même une fonction du temps qui est également instantané comme le rapport de ses risques instantanés et n'est donc pas constant au cours du temps. Il s'interprète de la même façon qu'un risque relatif.

### 5.5.1 Le modèle proportionnel de Cox

Proposé en 1972 par D.R. Cox (323), le modèle proportionnel de Cox est une régression couramment utilisée pour modéliser la survie et estimer l'effet d'une variable d'exposition principale sur cette survie, habituellement ajustée sur des co-variables d'exposition (324).

Soit  $Z$  le vecteur de  $p$  variables d'exposition :  $Z_1, \dots, Z_p$ . Pour un individu  $i$ , le vecteur des variables d'exposition s'écrit donc  $Z_i = (Z_{1i}, \dots, Z_{pi})$ . Les  $Z_i$  vecteurs de tous les individus inclus dans la cohorte peuvent, selon D.R. Cox, être des fonctions du temps  $t$ .

Il écrit ainsi la fonction de risque :

$$\lambda(t, Z) = \exp(Z_i\beta) \cdot \lambda_0(t) \quad (7)$$

Où les  $\beta$  sont les paramètres des variables  $Z_i$  à estimer et  $\lambda_0(t)$  est un paramètre inconnu donnant le risque de base lorsque l'ensemble des  $Z_i=0$ .

La modélisation de Cox permet d'obtenir les coefficients estimés des  $\beta$  des vecteurs de variables  $Z_i$  par la composante  $\exp(Z_i\beta)$ . La composante  $\lambda_0(t)$  n'est pas modélisée, c'est pourquoi ce modèle est dit semi-paramétrique.

En effet, pour deux individus  $i$  et  $j$ , le risque relatif peut s'écrire :

$$\lambda_i(t, Z) / \lambda_j(t, Z) = \exp(Z_i\beta) \cdot \lambda_0(t) / \exp(Z_j\beta) \cdot \lambda_0(t) = \exp(Z_i\beta) / \exp(Z_j\beta)$$

et est donc indépendant de  $\lambda_0(t)$  qui est éliminé les deux produits.

Les  $\beta$  sont donc estimés par une vraisemblance partielle (323), conditionnellement à la survenue de l'évènement d'intérêt au temps  $t$ , la probabilité  $P_i$  de subir l'évènement pour un sujet  $i$  au temps  $t$  par rapport à tous les autres individus  $j$ , tel que :

$$\begin{aligned} P_i(\beta) &= \exp(Z_i\beta) \cdot \lambda_0(t) / \sum \exp(Z_j\beta) \cdot \lambda_0(t) \\ &= \exp(Z_i\beta) / \sum \exp(Z_j\beta) \quad (\text{toujours indépendamment de } \lambda_0(t)) \end{aligned}$$

Le produit des  $P_i(\beta)$  pour chaque temps  $t$  où l'évènement d'intérêt survient est la vraisemblance partielle décrite par Cox :

$$P(\beta) = \prod_j P_j(\beta) = \prod_j \frac{\exp(Z_j\beta)}{\sum \exp(Z_j\beta)} \quad (8)$$

Les paramètres  $\beta$  sont constants au cours du temps ce qui implique que les risques sont proportionnels au cours du temps. C'est pourquoi le modèle est dit proportionnel. C'est une hypothèse forte qu'il est nécessaire de vérifier pour l'application du modèle de Cox sur un jeu de données.

Pour estimer cette vraisemblance partielle de Cox, la méthode utilisée par des logiciels statistiques courants (tels que SAS et R) est l'approximation de Breslow (324). Le test de Wald permet d'obtenir l'estimation des  $\beta$ , et l' $\exp(\beta)$  donne le HR spécifique de la variable considérée.

### **5.5.2 Les hypothèses du modèle de Cox et test de proportionnalité**

Le modèle de Cox n'est valide que si les paramètres  $\beta$  des effets des variables d'exposition sont constants au cours du temps. Cela doit être vérifié pour la variable d'exposition principale et pour les co-variables présentes dans le modèle. Ces deux premières hypothèses se traduisent en fait par un seul test : la proportionnalité des risques au cours du temps, qu'il est nécessaire de vérifier pour établir la validité du modèle.

Une troisième hypothèse est également implicitement formulée lors de l'utilisation de ce modèle : pour tout individu de la cohorte, il faut supposer qu'au cours du temps  $t$ , tous les individus partagent le même risque de base de subir l'évènement d'intérêt.

Enfin, la relation entre la fonction de risque et une variable quantitative d'exposition correspondante est forcément log-linéaire tel que  $HR_{1 \text{ à } n} = \exp(\beta_{1 \text{ à } n})$ . Dans les faits, il est donc nécessaire de vérifier que lorsqu'une variable quantitative  $X$  augmente d'une unité, quelle que soit la valeur de  $X$ , le HR correspondant à cette variable augmente également de  $\exp(\beta)$ . Cette hypothèse de log-linéarité se contournait souvent (par le passé) en mettant par exemple une variable âge en classes catégorielles puis de représenter graphiquement les logits de ces classes pour vérifier la linéarité de ces points.

Dans le présent doctorat, les formes fonctionnelles des variables continues d'exposition, soit l'âge des parents à l'accouchement, la durée de la gestation et le poids de naissance ont été explorées selon deux méthodes : 1) en polynômes fractionnaires dans R, puis 2) sous forme de variables « splines » linéaires, catégorielles, carrées, logarithmiques et cubiques restreintes afin de déterminer les meilleures formes à utiliser dans l'ajustement avec la macro SAS de l'Université de

Stanford (% SPECI) (325). Cette vérification a permis d'établir que ces variables quantitatives respectaient cette linéarité et pouvaient être utilisées telles qu'elles.

Il existe de nombreuses techniques, graphiques ou basées sur les résidus de la régression (e.g. : les résidus de Cox-Snell (326), de martingale (327)) pour vérifier la proportionnalité des risques instantanés. La technique des résidus de Schoenfeld (328) est aujourd'hui largement répandue. Elle consiste à modéliser la corrélation/interaction avec le temps des estimateurs utilisés.

Une méthode graphique :

Selon Cox, la fonction de risque s'écrit :  $\lambda(t, Z) = \exp(Z_i\beta) \cdot \lambda_0(t)$  (1)

Soit  $\beta(t)$  la fonction qui traduit la variation de  $\beta$  au cours du temps alors (1) devient :

$$\lambda(t, Z) = \exp(Z_i \cdot \beta(t)) \cdot \lambda_0(t) \quad (2)$$

Si  $\beta(t)$  est constant au cours du temps alors  $\beta(t)=\beta$  et (2)=(1) et tracer la fonction  $\beta(t)=\beta$  revient à une droite horizontale.

Les résidus de Schoenfeld :

Les résidus sont la somme des écarts entre les vraies valeurs observées dans un jeu de données et les valeurs théoriques données par la régression (326, 328).

En 1982 Schoenfeld (328) introduit les résidus  $r_{ik}$  et  $S_{ik}$  (standardisés) du même nom, pour des sujets  $i$  et  $j$  au  $k^{\text{ième}}$  instant  $t_k$  où leurs évènements d'intérêt sont observés tel que :

$$r_{ik} = X_{ik} - E(X_{ik}|R_i) \quad (9)$$

et

$$S_{ik} = \frac{\sum_{j \in A_{ik}} \exp(\hat{\beta} \cdot X_j)}{\sum_{j \in R_i} \exp(\hat{\beta} \cdot X_j)} \quad (10)$$

$E$  : espérance;  $A_{ik}$  est l'ensemble des identifiants tel  $j \in R_i$  et  $X_{jk} \leq X_{ik}$

Schoenfeld écrit « Puisque  $r_i$  est un vecteur, chaque  $r_{ik}$  doit être examiné graphiquement » (328).

Basé sur ces résidus standardisés, Patricia M. Grambsch et Terry M. Therneau ont montré que tracer le graphique de  $\beta(t) + S_{ik}$  au cours du temps permet de tester la non-proportionnalité

graphiquement (329). Il est également possible de tester l'hypothèse que cette courbe à une pente égale à zéro. Si ce test est statistiquement différent de zéro alors le risque est non-proportionnel (330). Néanmoins la p-valeur de ce test est dépendante de la taille de l'échantillon et du nombre d'évènements d'intérêt considérés. Plus cette taille et ce nombre sont importants, plus la valeur de p sera significative et ce test jugera alors à une non-proportionnalité pour une violation minimale de cette hypothèse (331). Enfin, le  $\beta(t)$  d'une variable considérée (par exemple l'exposition principale) peut être ajusté sur tous les autres  $\beta(t)$  des co-variables afin d'obtenir un test global de proportionnalité.

### **5.5.3 Le modèle de Cox cause-spécifique**

Un modèle de Cox cause-spécifique prend en compte l'évènement d'intérêt mais également le temps où il se produit. Ayant choisi cette approche pour l'ensemble de nos modèles, la date de décès et la cause étaient donc primordiaux, car la mort entre en compétition notamment pour nos évènements d'intérêt. Dans nos analyses, nous avons tout d'abord vérifié tous les décès et leurs causes. Aucun décès ne survenait par suite de nos évènements d'intérêt (sclérose en plaques ou diabète). Puis les décès (pour une autre cause que nos évènements d'intérêt), dans nos périodes d'incidence, ont été censurés à droite au jour de leur décès.

Le registre de décès et le FIPA apportaient tous deux ces variables essentielles à la construction de nos modèles (303).

## **5.6 Les co-variables et les facteurs de confusion**

### **5.6.1 Provenance des co-variables**

La provenance de nombreuses variables de la Cohorte de naissance québécoise sur l'immunité et la santé a été présentée au tableau 5.1. Nous décrivons ici plus particulièrement la définition de certaines variables utilisées dans l'ajustement de nos modèles de Cox.

L'origine ethnoculturelle des parents a été traduite par leurs lieux de naissance respectifs, dont l'information est enregistrée dans le fichier des naissances (303). Cette double origine a été utilisée comme proxy de l'ethnie des membres de la cohorte. En effet, entre 1940 et 1950, les parents de ces enfants nés au Québec étaient principalement des Caucasiens d'origine française ou britannique

qui représentait 94 % de la population de la province (332). Avant 1960, l'immigration au Québec était limitée et provenait majoritairement d'Europe. Parmi les autres groupes ethniques figuraient les Premières Nations, les Inuits et les Métis (~ 6 %), les Asiatiques (3 %) et les Africains (~ 1 %), ainsi que les personnes d'origine non déclarée (4,6 %) (332). Les parents nés à l'extérieur du Québec étaient dès lors considérés comme ayant possiblement d'autres origines ethnoculturelles. Néanmoins, au vu de la multiplicité des origines, deux variables binaires, une pour le père et une pour la mère, ont été construites, identifiant leurs lieux de naissance au Québec ou hors Québec.

Les 3 premiers chiffres du code postal, disponible de 1987 à 1994 et inclus au fichier FIPA de la RAMQ, ont permis de déterminer le lieu de naissance de l'individu (303, 307) et d'établir le statut rural / urbain, du lieu de résidence en 1987, information la plus proche de la date de naissance (1970-1974). Les données du recensement, tels que les revenus médians de la famille en 1991 et 1996 de Statistique Canada (307), ont également été ajoutées à nos bases de données grâce aux codes postaux à 3 positions. Seuls les indices de défavorisation matérielle et sociale, selon Pampalon et al. (333), ont été préalablement déterminés par la RAMQ à partir des codes postaux à 6 positions puis agrégés à nos données par le NAM de l'individu.

Les indices de défavorisation sociale et matérielle ont été développés par Pampalon et al. à l'Institut national de santé publique du Québec (INSPQ) à partir des données de recensement. Une analyse en composante principale a permis d'élaborer ces indices à partir de 6 indicateurs recueillis durant les recensements et utilisés par l'INSPQ dans cette analyse. Il s'agit de : 1) la proportion d'individus de 15 ans et plus divorcés, séparés ou veufs ; 2) la proportion de familles monoparentales; 3) la proportion d'individus parmi les 15 ans et plus occupant un emploi ; 4) le revenu moyen des sujets de 15 ans et plus ; 5) la proportion de personnes de 15 ans et plus vivant seules ; 6) la proportion d'individus de 15 ans et plus n'ayant pas obtenu de certificat d'étude ou de diplôme d'études secondaires. Pour les enfants (<15 ans) le revenu familial a été utilisé dans la construction de ces indices. L'analyse en composante principale dénombre 2 composantes, l'une matérielle, basée sur les indicateurs 1), 2) et 5) et l'autre sociale, basée sur les indicateurs 3), 4) et 6). Pour chaque secteur de dénombrement et composante principale les notes factorielles obtenues ont été découpées et ordonnées en quintiles (très favorisé, favorisé, moyennement favorisé, défavorisé, très défavorisé) (334). Différentes versions existent de ces indices (335). Puisque les codes postaux de résidence (secteurs de dénombrement) n'étaient disponibles qu'à partir de 1987 dans la cohorte utilisée, ces indices ont été obtenus pour le lieu de résidence en 1987 et en 1997 à

partir des données de recensement de 1991 et 1996 (les recensements sont effectués tous les 5 ans) (335, 336). L'année 1997 est également la première année de la seconde période d'analyse de l'article 1 et de la période de suivi de l'article 3. Il était donc opportun de construire ces indices pour 1987 et 1997.

## 5.6.2 Principe d'un graphe acyclique dirigé et de la confusion

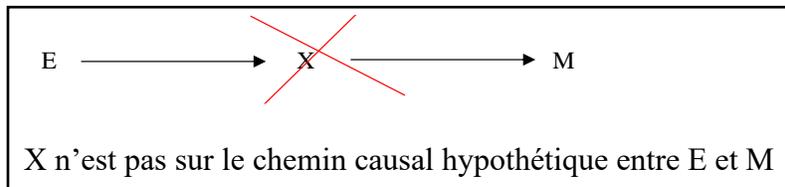
### 5.6.2.1 Notations d'un graphe acyclique dirigé et confusion

Un graphe acyclique dirigé permet de représenter graphiquement des hypothèses sur la relation entre deux ou plusieurs variables (337, 338). Les hypothèses de relation sont des lignes ou arêtes allant d'un nœud à un autre. Un graphe simple représente l'hypothèse de relation entre E et M (nos deux nœuds), où E est notre exposition principale et M une de nos maladies d'intérêt. Nous supposons tout d'abord qu'il existe une association entre E et M :

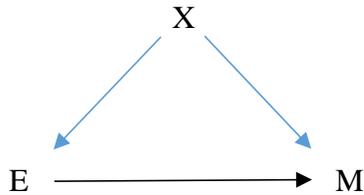


Un graphe est dit « acyclique » s'il n'y a pas de boucles de rétroaction ; une variable ne peut pas être son propre descendant. Ces bords sont dits orientés ou dirigés, c'est-à-dire qu'ils ont une seule pointe de flèche indiquant leur effet. Nous supposons ici que l'exposition au BCG induit une protection contre la survenue de notre maladie d'intérêt.

Soit X une troisième variable indépendante de E et M, qui est à la fois une cause de E et de M tel que trois conditions :



Alors X est une variable de confusion dans la relation  $E \rightarrow M$  et cela se traduit par deux flèches orientées de X vers E et M (337, 338), tel que :



### 5.6.2.2 Structure causale, origine des graphes acycliques dirigés et application dans l'étude

Le fait que nos graphes soient « dirigés » toujours dans un seul sens, sans possibilité de boucle rétroactive, et dits « acycliques », signifie qu'ils ont une structure causale. Chaque variable survient à un moment précis dans le temps et le fait que X cause E, ou X cause M ou que E cause M signifie que X précède E et M, et que E précède M dans le temps. Il n'y a pas de retour en arrière et il existe un ordre temporel dans le graphe qui est dit causal.

Cette structure causale a été développée dans le cadre de l'inférence causale généralisée par le Dr Rubin dans les années 1970, plus récemment décrite par Hernan (338). Le but de l'inférence est de déterminer si les résultats d'une étude sont extrapolables à la population cible et la causalité respecte l'idée intuitive qu'une cause précède toujours une conséquence et lie la variable en amont comme la cause directe de l'évènement d'intérêt en aval. Cette variable devient la cause directe de M.

Néanmoins dans le cadre de notre étude, le BCG ne peut être la seule cause directe de survenue de notre maladie M. L'exposome de l'individu induit par exemple une pluralité de variables ascendantes causant à terme la survenue de M. Ces variables pointent vers M mais pas forcément vers E. Elles ne sont donc pas toutes des variables de confusion. Le graphe acyclique dirigé nous permet donc de représenter et hiérarchiser les liens entre les variables. Secondairement, il permet de déterminer le nombre minimal de variable d'ajustement (pour la confusion) afin que l'estimation de la relation entre E et M soit le moins biaisée possible (339), tout en conservant un intervalle de confiance étroit qui permette une conclusion et une inférence à la population cible de l'étude.

### **5.6.3 Choix des facteurs de confusion**

Toutes les variables d'ajustement de nos modèles multivariés de Cox ont été établies comme des variables de confusion potentielle. Elles étaient toutes à la fois liées à notre exposition principale et à nos événements d'intérêt. Leurs liaisons, relations et hiérarchisations, les unes par rapport aux autres, ont été préalablement identifiées sur un examen minutieux de la littérature et la réalisation de graphes acycliques dirigés (337, 338) sous le logiciel Dagitty afin de déterminer le nombre minimal de variables d'ajustement (pour la confusion) conduisant à des estimateurs les moins biaisés possibles (339). Les graphes acycliques ainsi que les tableaux de revues de littérature sont annexés à chaque contribution. Tous les graphes acycliques et les revues de littérature en référence ont été formulés sur un examen complet de toutes les variables potentiellement liées à notre exposition principale ou à nos événements d'intérêt et non pas seulement sur un lien statistique avec les seules variables en présence dans la cohorte.

A l'examen de ces différents graphes acycliques dirigés, aucune variable identifiée comme facteur de confusion n'était indisponible dans nos jeux de données médico-administratives, ce qui est souvent un frein potentiel dans l'utilisation de ce type de données.

### **5.6.4 Interaction avec le sexe**

Un terme d'interaction entre l'exposition principale et le sexe a tout d'abord été introduit dans nos modèles de Cox univariés, puis ce même terme d'interaction a secondairement été testé en multivarié afin de tester une potentielle modification de l'effet.

## **5.7 Les données manquantes et techniques d'imputation**

Il existe de nombreuses techniques d'imputation multiple (e.g Multiple Imputation by Chained Equations (MICE), Predictive Mean Matching (PMM), MultiVariate Normal Imputation (MVNI), score de propension) (340). Nous ne détaillerons ici que la méthode choisie et utilisée pour nos données : l'imputation multiple par équations chaînées dite « MICE » en anglais. Cette technique à l'avantage d'être flexible et semi-paramétrique (341) puisqu'il est possible de pouvoir spécifier le modèle choisi, c'est-à-dire la fonction de lien qui peut être paramétrique tel que linéaire, logistique, ou polytomique, ou non-paramétrique comme PMM, afin d'imputer chaque variable

présentant des données manquantes dans une base de données. Le terme « chaîné » signifie également que chacune des régressions utilisées par variable présentant des données manquantes, se fait conditionnellement aux autres variables prédictrices incluses dans l'imputation. Pour comprendre son principe, il est avant tout nécessaire de décrire les mécanismes de construction des données manquantes.

### 5.7.1 Les mécanismes des données manquantes

Trois mécanismes de données manquantes sont couramment distingués (342-344) :

1. Les données manquantes sont générées de façon complètement aléatoire dites *Missing completely at random* (MCAR). La probabilité que des données soient manquantes pour une variable ne dépend ni des variables observées ni des variables non observées du jeu de données. Ceci est relativement rare.
2. Les données sont manquantes de façon aléatoire, dites *Missing at random* (MAR). La probabilité d'absence n'est pas totalement aléatoire mais elle est uniquement liée à une ou plusieurs autres variables observées du jeu de données.
3. La probabilité d'absence des données manquantes est non-aléatoire et la probabilité d'absence est liée à une ou plusieurs autres variables non observées du jeu de données. Les données manquantes sont dites *Missing not at random* (MNAR), ou manquantes de façon non aléatoire. Par exemple, la probabilité d'absence de réponse à une question posée dépend de la question elle-même. Un sujet sensible est en effet souvent évité voir éludé, car la réponse est dérangeante, comme son niveau de revenu ou sa sexualité qui ne se partagent pas avec tous.

Pour effectuer une imputation MICE, il faut présupposer que les données manquantes sont MAR, et qu'il est possible de générer une imputation sur les variables en présence du jeu de données. Il faut donc au minimum utiliser toutes les variables introduites dans nos régressions multiples de Cox. Néanmoins, Graham souligne l'importance d'introduire des variables auxiliaires informatives pour le mécanisme de données manquantes et qui permet également de réduire la perte d'information de l'imputation (biais d'information des données manquantes et attrition) et d'en augmenter la puissance statistique (345). Les variables auxiliaires sont des variables qui ne font pas partie du modèle d'intérêt, mais qui sont fortement corrélées aux variables qui le composent (345). L'ajout de ces variables auxiliaires permet de combler le biais d'information

généralisé par les données manquantes elles-mêmes mais également de réduire le biais d'estimation de ces données manquantes si leur mécanisme est finalement MNAR plutôt que MAR (345). Cet ajout permet également de restaurer partiellement la puissance statistique perdue par les données manquantes (345). L'attrition est la mesure de l'effet du manque de données sur le biais d'estimation qui peut être traduite par un pourcentage de l'erreur standard de ce biais, tel que :

$$\frac{100X(\text{estimation moyenne du paramètre} - \text{valeur dans la population})}{SE} \quad (11)$$

où SE=l'erreur type de l'estimation, ou l'écart type de la distribution d'échantillonnage pour le paramètre estimé en question.

### 5.7.2 Les principes de l'imputation « MICE »

Au lieu de remplacer une donnée manquante par une seule valeur (imputation simple), l'imputation multiple remplace cette donnée par plusieurs valeurs estimées (344). Pour chacune des valeurs estimées, il est possible de calculer une variance (344). Ceci est la première phase, la phase dite d'imputation.

Soit  $Z$  un vecteur contenant des variables observées  $O$  et des variables incomplètes  $I$  tel que  $Z=(O_1, \dots, O_n, I_1, \dots, I_n)$ , une régression (par variable) estime les paramètres  $\beta$ , tel que par exemple pour la première variable :  $P_1(I_1|O)=P(I_1/ O_1, \dots, O_n)$

La distribution des paramètres  $\beta_n$  avec  $Z_{1 \text{ à } n}=(O_{1 \text{ à } n}, I_{1 \text{ à } n})$  est une chaîne de Markov (344). Cette étape permet d'obtenir des estimateurs des données manquantes (et leur variabilité) conditionnellement aux variables observées mais aussi en fonction des nouvelles observations générées par les précédentes régressions de  $P_1$  à  $P_n$  (344).

Le nombre de bases de données à générer fait débat. Il est variable et dépend de la quantité de données manquantes présentes dans un jeu de données, selon Graham (345). Selon lui, pour 50% de données manquantes, il ne faudrait pas moins de 40 tables d'imputation différentes. En effet, plus ce nombre est important, plus il est possible d'obtenir des estimateurs et leurs variances (internes) sur chacune de ces tables ainsi que la variance inter-tables. Ceci est la phase d'estimation. Cette multiplication des estimations permet donc en théorie d'obtenir des estimateurs

et variances agrégés non biaisés. En pratique il est souvent communément accepté que 3 à 5 tables d'imputation sont suffisantes face à une infinité qui pourraient être générées, et que la multiplication au-delà de 5 n'est pas nécessaire pour obtenir une meilleure précision (344). Harel explique même plus récemment que le nombre nécessaire de tables augmente bien à mesure que la quantité d'informations manquantes augmente, mais que malheureusement, il n'y a pas de seuil clair qui soit généralisable à tous les cas de figures (342). Dans nos 3 articles, nos seuils ont été de 5 et 7 tables face à une quantité de données manquantes ne dépassant pas au maximum 9 à 13% de nos données.

Une troisième phase consiste ensuite à « fusionner » les différents estimateurs et leurs variances sur chaque base de données obtenue afin d'obtenir un estimateur par variable contenant des données manquantes prenant en compte la variance interne de l'estimateur généré à chaque base de données itérées, et la variance externe (inter-bases de données itérées).

La méthode d'imputation multiple par équations chaînées est accessible dans le module FSC (Fully Conditional Specification) de SAS (341), ou le package MICE de R (346).

## **5.8 Le redressement des échantillons**

### **5.8.1 Utilisation du score de propension et de l'IPW pour le redressement**

Les analyses des articles 2 et 3 ont été produites à partir d'échantillons restreints de la population éligible, définis par une couverture du système public pour les services pharmaceutiques. Cette couverture est habituellement offerte aux populations plus défavorisées, c'est-à-dire n'ayant pas accès à une assurance privée via un employeur. Afin de compenser de potentiels biais de sélection dus à ces restrictions, des pondérations ont été utilisées pour redresser les résultats obtenus afin qu'ils soient représentatifs de la population d'étude.

Pour tous les individus  $i$  de notre population éligible, nous étudions une association tel que :

$Y_i = \beta_0 + \beta_1.X_{i,1} + \dots + \beta_n.X_{i,n}$  où  $Y_i$  est notre événement d'intérêt et  $X_{i,1}$  à  $n$  le vecteur de variables explicatives.

En restreignant notre échantillon,  $Y_i$  n'est pas toujours observé alors que les variables explicatives du modèle,  $X_{i(1 \text{ à } n)}$ , et des variables sociodémographiques auxiliaires,  $X_{i(0 \text{ à } z)}$ , le sont toutes pour

l'ensemble de l'échantillon éligible et de l'échantillon restreint (ce dernier étant inclus dans l'échantillon éligible).

Soit  $R_i=1$  et  $R_i=0$  quand  $Y_i$  est observé ou non, et  $P(R_i=1) = \pi_i$  la probabilité d'observation de  $R_i=1$ .

Soit un individu  $i$ , qui a ou non une couverture médicaments publique et un ensemble de variables observées  $X_{(1...o...z)}$ , il est aisé d'obtenir la probabilité de  $P(R_i|X_{i(1...o...z)})$  par un score de propension qui s'exprime par une régression logistique tel que :

$$P(R_i|X_{i(1 \text{ à } z)}) = \beta_0 + \beta_1 \cdot X_{i,1} + \dots + \beta_{iz} \cdot X_{iz}$$

Le biais de sélection peut être corrigé par une pondération selon la probabilité inverse (*inverse probability weights*, IPW) (338), qui est la probabilité inverse pour chaque individu  $i$  d'observer  $R_i$  sur la base des caractéristiques des variables explicatives choisies  $X_{i(1 \text{ à } z)}$ , tel que:

$$\hat{\mu}_{ipw} = \frac{1}{P(R|X)} \quad (12)$$

Hernan et Robins (2020) démontrent que cette pondération est proche de l'estimateur de Horvitz et Thompson (1952) (338):

$$\hat{\mu}_{ipw} = \frac{\sum_{i=1}^n R_i \cdot \frac{Y_i}{\pi_i}}{\sum_{i=1}^n \frac{R_i}{\pi_i}} \quad (13)$$

mais que cette pondération présente plusieurs inconvénients (338).

L'objectif est en effet de créer une pseudo-population dans laquelle le vecteur de co-variables  $X$  observées est identique dans les deux groupes où  $Y$  est observé/non observé.

Néanmoins, la pseudo-population obtenue est très souvent le double de l'échantillon réel selon Hernan et Robins (2020) car les individus surreprésentés reçoivent une pondération  $\leq 1$ , et ceux sous-représentés, une pondération  $>1$ , pouvant même atteindre 2, et même plus, pour des individus extrêmes (très peu représentés). La somme de ces poids augmente donc artificiellement la taille de l'échantillon (à 2 en moyenne) et cela conduit à une sous-estimation de la variance et des intervalles de confiance plus restreints.

Pour pallier une telle situation, Hernan et Robins (2020) conseillent d'utiliser des poids stabilisés ou *Stabilized Inverse Probability Weights* (SIPW) (338), en ajoutant au numérateur la probabilité marginale qu'a un individu  $i$  d'appartenir à son groupe d'exposition, tel que :

$$\hat{\mu}_{sipw} = \frac{P(R = 1)}{P(R|X)} \quad (14)$$

De tels poids présentent l'avantage de ne jamais dépasser 1 et la pseudo-population est de la même taille que l'échantillon initial car la somme de ces poids n'excède pas la valeur 1 (en moyenne). Ceci permet d'obtenir des estimations correctes de la variance (338).

Dans le présent doctorat, la construction successive des poids IPW puis SIPW nous a permis d'observer les recommandations de Hernan et Robins (2020) (338). Seuls les poids SIPW ont été utilisés pour le redressement des HRs des contributions 2 et 3.

## 5.9 Résumé du plan d'analyse

Les trois articles du présent doctorat utilisent les mêmes expositions, mais des événements d'intérêt, des périodes de suivi et des critères d'inclusion différents, d'où le choix d'utiliser ou non des pondérations de redressement afin de rendre nos résultats inférables à la population cible des études. L'ensemble du plan d'analyse est résumé au tableau 5.2.

**Tableau 5.2 : Résumé du plan d'analyse**

Objectif	Exposition	Événement d'intérêt	Période de suivi (âges)	Critères d'inclusion	Pondération
1a)	Vaccination BCG Âge au BCG	Sclérose en plaques	1983-2014 (9-44 ans)	Aucun (sujets décédés avant le 1 <sup>er</sup> janvier 1983 ont été exclus)	Non
1b)	Vaccination BCG Âge au BCG	Sclérose en plaques (RRMS et autres phénotypes)	2 périodes 1983-1996 (≤21 ans) 1997-2014 (22-44 ans)	Aucun (sujets décédés avant le 1 <sup>er</sup> janvier 1983 ont été exclus)	Oui
2	Vaccination BCG Âge au BCG	Diabète de type 1	1985-1993 (10-18 ans)	Cas incident de la période considérée (exclusion des cas prévalents entre 1983 et 1984)	Non
3a	Vaccination BCG Âge au BCG	Diabète de type 1 et type 2, sans caractérisation des LADA	1997-2014 (22-44ans)	Couverture des services pharmaceutiques entre 1997 et 2014 et ne pas avoir eu un diabète ou être	Oui

				décédé avant le 1 <sup>er</sup> Janvier 1997	
3b	Vaccination BCG Âge au BCG	Diabètes de type 1, 2 et LADA	1997-2014 (22-44 ans)	Couverture des services pharmaceutiques entre 1997 et 2014 et ne pas avoir eu un diabète ou être décédé avant le 1 <sup>er</sup> Janvier 1997	Oui

BCG, Bacille Calmette-Guérin; LADA, Diabète auto-immun latent de l'adulte

## 5.10 Éthique

L'étude a été approuvée par les comités d'éthique de l'Institut national de la recherche scientifique (Projet Sclérose en plaques, CER-15-377; Projet Diabète, CÉR-09-204), de l'Institut de la statistique du Québec et de la Régie de l'assurance maladie du Québec, et l'accès aux données a été accordé par la Commission d'accès à l'information du Québec (CAI #110267-S).

Les données de la cohorte sont hébergées et sécurisées sur les serveurs de l'Institut de la statistique du Québec. Elles ont été accessibles physiquement en Centre d'accès aux données de recherche dans les locaux du Centre interuniversitaire québécois de statistiques sociales de l'Université de Montréal, puis à distance grâce à un accès sécurisé durant la pandémie de SARS-CoV-2.

Les analyses présentées ont toutes fait l'objet d'une autorisation avant divulgation et respectent les règles de confidentialités de ces organismes en matière de recherche, notamment aucune statistique incluant moins de 5 individus.

## **6 Article 1. Bacillus Calmette-Guerin vaccination and multiple sclerosis: a population-based birth cohort study in Quebec, Canada**

---

### **Vaccination au Bacille Calmette-Guérin et survenue de sclérose en plaques : une étude de cohorte de naissance basée sur la population Québécoise, Canada**

**Auteurs :** Philippe Corsenac<sup>1</sup>, Marie-Élise Parent<sup>1</sup>, Christina Wolfson<sup>2</sup>, Nathalie Arbour<sup>3</sup>, Pierre Duquette<sup>3</sup>, Andrea Benedetti<sup>2,4</sup>, Hugues Richard<sup>1</sup>, Simona Stäger<sup>5</sup>, Marie-Claude Rousseau<sup>1</sup>.

<sup>1</sup> Unité d'épidémiologie et biostatistiques, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique (INRS), Laval, Canada ;

<sup>2</sup> Department of Epidemiology, Biostatistics and Occupational Health, School of Population and Global Health, McGill University, Montreal, QC, Canada;

<sup>3</sup> Centre de recherche du CHUM, Département des neurosciences, Université de Montréal, Montréal, QC, Canada;

<sup>4</sup> Respiratory Epidemiology and Clinical Research Unit, Research Institute of the McGill University Health Centre; Montreal, QC, Canada;

<sup>5</sup> Immunologie et virologie, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique (INRS), Laval, Canada ;

**Titre de la revue :** European Journal of Neurology ; 2022 Feb 15.

DOI: 10.1111/ene.15290.

### **Contribution des auteurs :**

Philippe Corsenac a conçu la méthodologie analytique de l'étude, effectué la revue de la littérature, les analyses statistiques et rédigé toutes les versions du manuscrit. Marie-Élise Parent a participé à la conceptualisation de la Cohorte de naissance québécoise sur l'immunité et la santé (QBCIH), co-encadré les travaux de thèse de doctorat de Philippe Corsenac, contribué au manuscrit et approuvé la version finale. Christina Wolfson a apporté son expertise dans les domaines de l'épidémiologie et de la sclérose en plaques, a contribué au manuscrit et a approuvé le manuscrit final. Nathalie Arbour a fourni une expertise en immunologie, en particulier en immunopathologie de la sclérose en plaques, a contribué au manuscrit et a approuvé le manuscrit final. Pierre Duquette a fourni une expertise clinique dans le domaine de la sclérose en plaques, contribué au manuscrit et approuvé le manuscrit final. Andrea Benedetti a participé à la conceptualisation de la QBCIH, fourni une expertise statistique, contribué au manuscrit et approuvé la version finale. Hugues Richard était responsable de la préparation et de la gestion des données, a fourni une expertise statistique, contribué au manuscrit et approuvé le manuscrit final. Simona Stäger a fourni une expertise en immunologie, contribué au manuscrit et approuvé la version finale et encadré le doctorat de Philippe Corsenac dans sa partie immunologie. En tant que chercheur principal, Marie-Claude Rousseau a conceptualisé et dirigé le développement de la QBCIH, obtenu et géré les fonds de recherche. Elle a supervisé la recherche doctorale de Philippe Corsenac, y compris la revue de la littérature et les analyses statistiques, a revu toutes les ébauches du manuscrit et a approuvé le manuscrit final. Philippe Corsenac et Marie-Claude Rousseau sont les garants de ce travail et, à ce titre, ont eu un accès complet à toutes les données de l'étude et assument la responsabilité de l'intégrité des données et de l'exactitude de l'analyse des données.

## 6.1 Mise en contexte

Six études cas témoins (54-59), regroupées dans une méta-analyse (60), ont étudié l'association entre la vaccination au BCG et la sclérose en plaques. Ces études mettaient en évidence une absence d'association (odds ratio=0.96, 0.69-1.34) (60). Néanmoins, aucune ne distinguait les phénotypes principaux de sclérose en plaques récurrente-rémittente et progressive primaire, ni même l'âge au diagnostic, indicateur de la forme clinique. Aucune ne visait à déterminer la présence d'effet différentiels selon le sexe, ni d'un effet potentiel de l'âge lors de la vaccination. Ces études se composaient, de plus, de petits échantillons (54-59) et utilisaient pour l'exposition principale, un statut vaccinal au BCG auto-rapporté (54-57, 59), reconnu comme ayant une validité limitée (347).

L'article 1, intitulé : *Bacillus Calmette-Guerin vaccination and multiple sclerosis: a population-based birth cohort study in Quebec, Canada*, s'intéresse tout d'abord à estimer l'association entre la vaccination au BCG et la survenue de sclérose en plaques, toutes formes cliniques confondues, en palliant aux limites méthodologiques des précédentes études. Afin d'étudier les formes cliniques plus précisément, cet article utilise en complément des données médico-administratives, les données pharmaceutiques de la Régie de l'assurance maladie du Québec et circonscrit les cas de sclérose en plaques selon l'âge au diagnostic puis selon la forme clinique prépondérante, la sclérose en plaques récurrente-rémittente qui représente 85% des formes initiales. Cet article met enfin en évidence la nécessité de futures analyses, devant absolument distinguer les phénotypes de sclérose en plaques et circonscrire plus particulièrement les formes de sclérose en plaques progressive primaire et progressive secondaire.

## 6.2 Abstract

**Background:** The Bacillus Calmette-Guerin (BCG) vaccine could reduce the incidence of multiple sclerosis (MS) through immunomodulation. Previous studies, presenting some limitations, reported no association. We re-examined this association in a large cohort focusing on relapsing-remitting MS (RRMS).

**Methods:** The cohort included 400,563 individuals, and was linked with the Quebec provincial BCG vaccination registry and administrative health data. Individuals were followed-up from 1983 to 2014 and then within period 1 (1983-1996) and period 2 (1997-2014), for the occurrence of MS. Incident MS cases were defined as those with  $\geq 3$  hospital or physician claims for MS. Subjects with  $\geq 1$  drug reimbursement for MS disease-modifying therapies were classified as RRMS. Cox proportional hazards regression was used to estimate hazard ratios (HR) over the follow-ups, adjusting for potential confounders. Possible effect modification due to sex was assessed.

**Results:** A total of 178,335 (46%) individuals were BCG vaccinated. There were 274 (0.06%) incident MS cases identified in 1983-1996, and 1,433 (0.4%) in 1997-2014. No association was found with RRMS, either in period 1 (adjusted HRs= 0.96, 95% confidence interval: 0.63-1.45; 96 cases) or in period 2 (HR<sub>adj</sub>= 1.02, 0.85-1.23; 480 cases). The remaining MS cases, for whom the phenotype was unknown, were positively associated with BCG over the entire follow-up (HR<sub>adj</sub>= 1.25, 1.10-1.41; 1,131 cases) and in period 2 (HR<sub>adj</sub>=1.33, 1.17-1.52; 953 cases). No interaction with sex was found.

**Conclusion:** Findings suggest that BCG vaccination does not decrease the risk of RRMS, and that future studies should consider phenotypes of MS.

### 6.3 Introduction

Multiple sclerosis (MS) is a chronic neurodegenerative disease with variable courses (180).

The disease begins with one of two distinct clinical forms, relapsing-remitting (RRMS, 85-90% of cases) or primary progressive (PPMS, 10-15%) (1-3).

The disease etiology is unclear but MS would be an autoimmune disease, even though the identification of autoantigens targeted by T and B cells remains incompletely resolved (21).

RRMS and PPMS differ in terms of prodromal periods (348), sociodemographic characteristics (3, 101, 349), environmental risk factors, and genetic susceptibility (2, 348, 350). RRMS and PPMS patients also exhibit different central nervous system inflammation characteristics (23)

The Bacillus Calmette-Guerin (BCG) vaccine enhances the activity of innate and adaptive immune cells exhibiting regulatory properties, as observed in both animal models and humans (37, 44, 45, 351), thereby curtailing chronic inflammation.

A meta-analysis (60) of six case-control studies (54-59) on BCG vaccination and MS, without distinction of phenotypes, found no evidence of an association (summary odds ratio=0.96, 95% confidence interval (CI): 0.69-1.34). These studies had some methodological limitations including small sample sizes (54-59) and self-reported BCG vaccination status (54-57, 59), recognized to have limited validity (347).

Furthermore, BCG vaccination could have different effects according to age at vaccination, especially during the first year of life (47, 48) due to better heterologous innate stimulation and proinflammatory cytokine production in earlier than in postponed vaccination. No previous studies targeted this aspect.

To address the aforementioned shortcomings, we assessed the association between BCG vaccination (or age at vaccination) and MS, first overall and then distinguishing RRMS cases identified through use of disease modifying therapies (DMTs), in a large birth cohort.

## **6.4 Methods**

### **Study design and data sources**

The study used the Quebec Birth Cohort for Immunity and Health (QBCIH), previously described (303). It was established through probabilistic linkage of the birth registry and the Healthcare Registration file using surname, first name, sex, date of birth, and father's first name as record identifiers between datasets.

Probabilistic linkage was also used to add data from the death registry (covering 1970-2014) and Quebec BCG vaccination registry. Data from physician claims (covering 1983-2014), the hospitalization database (covering 1987-2014), and the public drug insurance database (available in 1983-1996 for 5% of the cohort, and in 1997-2014 for 70.4%) were imported using a unique identifier available in these datasets and in the Healthcare Registration File.

### **Study population and follow-up**

Eligible subjects were born in Quebec between 1970 and 1974 after at least 32 weeks gestation, documented in the birth registry and covered by the public health system. The cohort initially included 400,611 individuals, representing 90.4% of eligible persons for whom the initial probabilistic linkage was successful (Figure 6.1).

Analyses were conducted on the entire follow-up period (1983-2014) and then divided in two time periods, 1983-1996 (period 1) and 1997-2014 (period 2) for two reasons. First,

we wanted to study pediatric and adult cases separately, but there were too few pediatric cases (up to 16 years of age). We thus decided to analyze pediatric and early adulthood cases occurring until 1996 together, as a group of cases diagnosed at a younger age. Second, prescription drug reimbursements were only available in period 2, which coincided with the availability of DMTs in Canada. A large proportion of the study population was covered by the public drug insurance (70.4%, median duration=6.9 years, Figures 6.1 and 6.S1).

### **Outcome definition**

Different definitions allowed the identification of incident MS overall and incident RRMS. The first definition, validated by Marrie et al. in Canadian pediatric and adult populations, has high sensitivity (96%, 95% CI: 92.3-98.3) and specificity (99.5%, 95% CI: 97.3-100.0) (308-310). Using this definition, referred to hereafter as ‘Marrie’s definition’, incident MS cases were identified as individuals who had  $\geq 3$  hospital or physician claims for MS over the follow-up (308-310). Physician claims and hospitalizations with MS as the diagnostic code (International Classification of Disease 9<sup>th</sup> revision (ICD-9): 340 and ICD-10: G35) were used. Dates of diagnosis were defined as the first ever date of a healthcare service for MS or demyelinating diseases [based on 4-digit ICD-9/10 codes: encephalitis myelitis and encephalomyelitis (323.4-323.6, 323.8-323.9 / G36.9), neuromyelitis optica (341.0 / G36.0), other demyelinating diseases of the central nervous system (341.8, 341.9 / G37.8, G37.9), optic neuritis (377.3 / H46)].

The second, our definition, identified individuals with  $\geq 1$  DMT claims as definite RRMS cases between 1997 and 2014, (median duration of the public drug insurance coverage = 12.6 years for DMT users). Included DMTs are presented in Table 6.S1. During this period, DMTs could only be prescribed to RRMS patients (313). All patients who received a DMT

in 1997-2014 and had a diagnosis date of MS in period 1 (1983-1996) were also identified as definite RRMS as of their date of diagnosis. For cases identified only with our definition, the date of diagnosis corresponded to the date of first DMT reimbursement or a prior date of health service with a demyelinating disease diagnostic code. Cases which were not identified as RRMS were considered as “unspecified”.

### **Exposure definition**

An organized voluntary program for BCG vaccination was conducted in the province of Québec, Canada, from 1949 to the mid-1970s (279), targeting newborns and schoolchildren (282). All vaccination certificates were centralized in a provincial BCG vaccination registry, which has been computerized, and assessed for completeness and accuracy (279, 282). BCG vaccination status was ascertained by presence in the registry, and those unlinked were considered as unvaccinated. Vaccination status was first defined as a binary variable (yes / no) and then categorized according to age at BCG vaccination (not vaccinated / vaccinated < 1 year old / vaccinated  $\geq$  1 year old). When the vaccination status, either binary or categorical, was uncertain due to unsatisfactory linkage to the registry, it was imputed.

### **Statistical analysis**

Cox proportional hazards models were estimated with age as the time scale to assess the association between BCG vaccination and incident overall MS, RRMS and unspecified MS.

Subjects were followed over the entire follow-up, and then over two consecutive time periods. Data were left-truncated at 1983, or at 1997 according to the period. The follow-up was right-censored at 31<sup>st</sup> December 1996, or 31<sup>st</sup> December 2014, depending on the

period. Age at entry over the maximum follow-up and in period 1 was between 8 and 12 years old, while age at entry in period 2 was between 22 and 26 years old. Follow-up ended at the date of diagnosis of MS, death, or on 31<sup>st</sup> December 1996 (period 1) or 2014 (entire follow-up, period 2), whichever came first. The proportionality assumption was ascertained by examining tests and graphs for the Schoenfeld standardized residuals.

Since sex-differential non-specific effects of the BCG vaccine have already been observed, the potential differential effect of BCG vaccination by sex was assessed by introducing an interaction term (BCG vaccination status\*sex) in the multivariable models (46).

Potential confounders were selected according to a directed acyclic graph (DAG) (337) presented in Figure 6.2, based on a literature review (Table 6.S3). Multiple imputation by chained equations was applied (n=5 data sets) using the distributions of other informative variables included in the dataset. Continuous variables were treated as linear, after confirming linearity of the logit.

Study samples were weighted through three stabilized inverse probability weighting procedures, one for each follow-up period, using propensity scores (338) to balance covariates between the subsamples of subjects covered or not by the drug insurance (Figure 6.S1). We then examined whether weighting balanced the measured covariates between those with or without drug insurance (352) (Figures 6.S2 to 6.S19).

We investigated associations within the subgroup with unknown MS phenotypes. First, we estimated interaction terms between BCG vaccination and year of birth. Second, the sociodemographic characteristics were described by MS subgroup and compared across them.

All analyses were performed with the SAS software version 9.4.

### **Analytical sample**

The study population included 400,563 individuals within the overall follow-up and the first analytical period (1983-1996), and 397,666 people within the second period (1997-2014) (Figures 6.1 and 6.S1).

### **Ethics approval and access to data**

The study was approved by the ethics committees of Institut National de la Recherche Scientifique, Institut de la Statistique du Québec and Régie de l'assurance maladie du Québec, and access to data was granted by the Commission d'accès à l'information du Québec.

### **6.5 Results**

None of the sociodemographic variables differed according to BCG vaccination status, except for areas of residence in 1987 and 1997, and parents' birthplace. The female to male ratio was 0.95 in the analytical sample (Table 6.1).

Among vaccinated subjects (n=178,133, 44.5% of the analytical sample), the vast majority were vaccinated under one year of age (87.8%) and received only one dose (93.9%).

### **Multiple sclerosis prevalence and incidence**

A total of 1,707 incident MS cases were identified, including 576 RRMS and 1,131 unspecified MS (Figure 6.1). The incident MS cases included 48 RRMS cases only identified by our definition (Figure 6.S1). There were 274 incident MS cases identified in 1983-1996 (mean age at diagnosis 21.7, SD=3.0 years), whereas 1,433 were identified in 1997-2014 (mean age at diagnosis 33.4, SD=5.1 years). The period prevalence of MS over the overall follow-up (1983-2014), based on subjects identified with our two definitions,

was 0.43%, 95% CI: 0.42-0.43%. The incidence rate of MS for the entire follow-up was 10.1 per 100,000 person-years (95% CI: 9.3-10.9). In period 1 (1983-1996), this rate was 4.9 cases per 100,000 person-years (95% CI: 4.3-5.5), including 1.7 cases per 100,000 person-years (95% CI: 1.4-2.1) for RRMS. In period 2 (1997-2014), the incidence rate of MS was 12.7 cases per 100,000 person-years (95% CI: 11.8-13.6), including 4.3 cases per 100,000 person-years (95% CI: 3.7-4.8) for RRMS.

### **Proportionality assumption**

Hazard ratios were globally proportional over time (P-values ranging between 0.18 and 0.99 through the 5 imputed datasets), except when the variable-specific hazards for some covariates were not proportional. Some multivariable models thus required exclusion of material deprivation in 1987 and/or number of older siblings (Tables 6.2 and 6.3).

### **Association between BCG vaccination and multiple sclerosis**

#### *All cases*

The risk of MS associated with BCG vaccination was increased when considering all cases ( $HR_{adj} = 1.16$ , 95% CI: 1.05-1.28, 1707 cases, Table 6.2), reflecting the increased risk among subjects vaccinated before one year of age. When analyzing the two periods (Table 6.3), no association emerged for period 1, comprised of 274 younger cases. However, there was evidence of a positive association with BCG vaccination for older subjects in period 2 ( $HR_{adj} 1.22$ , 95% CI: 1.09-1.36, 1433 cases).

### *RRMS cases*

There was no association between BCG vaccination and RRMS in the entire follow-up period ( $HR_{adj} = 1.01$ , 95% CI: 0.85-1.20; 576 cases). Nor were there associations in period 1 ( $HR_{adj} = 0.96$ , 95% CI: 0.63-1.45, 96 cases) or in period 2 ( $HR_{adj} = 1.02$ , 95% CI: 0.85-1.23; 480 cases).

### *Cases with unknown phenotypes*

BCG vaccination was positively associated with unspecified MS ( $HR_{adj} = 1.25$ , 95% CI: 1.10-1.41; 1,131 cases). While no association was apparent in period 1 (178 cases), an elevated risk was found in period 2 ( $HR_{adj} = 1.33$ , 95% CI: 1.17-1.52; 953 cases) (Table 6.3). Again, the increased risk was observed among subjects vaccinated before 1 year of age.

We observed no interaction between BCG vaccination and year of birth (P-values of interaction terms ranged 0.27-0.32 through the 5 imputed datasets). The sociodemographic characteristics by MS subgroup presented no notable differences (Table 6.S4); cases of MS with unknown phenotype diagnosed older were not sociodemographically different from those diagnosed younger, or from the RRMS cases. Risk estimates of unspecified MS cases were also uniform according to age at diagnosis.

Findings did not differ by sex. P-values for interaction terms between BCG vaccination status and sex ranged between 0.12 and 0.97 through the 5 imputed datasets.

## 6.6 Discussion

We observed no association between either BCG vaccination or age at vaccination, and the incidence of overall MS or RRMS, before 21 years old. A positive association was found for later-onset overall MS, but not for RRMS. In the subgroup with unknown phenotypes, risk estimates were elevated when considering the entire follow-up and among cases diagnosed after age 21.

The age-adjusted prevalence of MS in Manitoba in 2006 was estimated at 0.23% (95% CI: 0.22–0.24) (310). From 1998 to 2006, the average age- and sex-adjusted annual incidence of MS per 100,000 population was 11.4 (95% CI: 10.7–12.0) (310). Using the Marrie definition, the age- and sex-standardized prevalence of MS in Saskatchewan in 2013 was estimated at 0.31% (95% CI: 0.30-0.32) and an average annual adjusted incidence per 100,000 between 2001 and 2013 at 16.5 (95% CI: 15.8-17.2) in overall population (mean age at diagnosis 42.8, SD=13.8 years) (308). Unfortunately, the time periods, and the age of the subjects, were different. As such, it is difficult to conclude except to highlight that our period prevalence and incidence seem to be approximately in the same range as those from other studies conducted in Canada (308, 310).

Six previous case-control studies on the association between BCG vaccination and MS (54-59) reported no association, but methodological limitations likely hampered the ability to detect associations. The largest study included 140 MS cases (58). Overall, the associations ranged from 0.70 to 1.30, none differing from the null value, and a meta-analysis reported a summary OR of 0.96 (95% CI 0.69–1.34) (60). None of these studies addressed specific phenotypes of MS or age at diagnosis.

In these studies, BCG exposure assessment was either based on vaccination records or annual reactions to tuberculin (54), and on self-reported vaccination status (55-59). Self-reports have been shown to have poor validity (347) for determining vaccination status, and thus can lead to misclassification bias.

The present study used a reliable source to determine exposure, a provincial BCG vaccination registry, similarly to Andersen et al. (54). They found a null risk estimate (0.70, 95% CI: 0.37-1.33) with a wide confidence interval, likely due to limited statistical power (92 MS cases). Our results for overall MS cases diagnosed before 21, and for RRMS cases notwithstanding the age at diagnosis, are in line with this finding.

With respect to outcome assessment, previous studies ascertained MS incidence through a population-based register (54) or clinically-defined diagnoses (55-59), but none addressed specific phenotypes. Although clinical data represents the gold standard for identifying MS, the present study is the first to apply a validated definition, which was previously used in the Canadian provinces of Saskatchewan and Manitoba for identifying pediatric and adult incident MS cases from administrative health data (308-310). An original contribution of our work was to use information on DMT utilization from the public drug insurance to identify RRMS. Culpepper et al. (353) demonstrated better sensitivity and specificity for identifying MS from administrative health data when including DMT claims in the definition. We applied Marrie's and our definition separately, and took advantage of the fact that the vast majority (70.4%) of the cohort members were covered for drug insurance when DMTs arrived on the market. Our definition improved sensitivity over Marrie's. Indeed, it enabled the identification of 48 additional incident MS cases, classified as RRMS, representing 2.8% of all cases. Although not validated, our definition picked up

576 RRMS cases, of which 92% (528 cases) were also identified using Marrie's definition (Figure 6.S1).

None of the previous studies on BCG vaccination and MS considered an exhaustive list of potential confounders (54-59). Most used age- (54-58) and/or sex- (54-56, 58) and/or area of residence- (55, 56) matched controls as well as univariable analyses to estimate their ORs. In the present study, we used a DAG (337) to formalize and justify the selection of potential confounders.

### **Putative immunological mechanisms**

Previous clinical trials concerning BCG vaccination of patients with MS demonstrated decreased disease activity (40, 354, 355). It was suggested that such induced protection could be mediated by tumor necrosis factor (TNF- $\alpha$ ) (40, 298), a complex cytokine. TNF exists in soluble or in membrane bound forms and each form has a greater affinity for TNF receptor 1 (TNFR1) and TNFR2, respectively. While TNFR1 signaling triggers apoptosis and chronic inflammation, TNFR2 signaling plays a key role in physiological processes in the central nervous system such as cell survival and myelination. These authors also suggested that TNFR2 signaling could favor regulatory T lymphocytes (40, 298). In contrast, others proposed that trained immunity is the breeding ground for the development of autoimmune diseases like MS (299). This hypothesis may support in part our findings on the unspecified MS subgroup. We propose that BCG vaccination at an early age enhances innate immune responses through greater pro-inflammatory cytokine production. However, elevated cytokine responses by innate cells could improve regulatory responses and avoid breach of self-tolerance (21). Finally, we speculate that if viruses are involved in the development of MS (21), BCG could promote non-specific antiviral defenses (43,

75, 289) through trained innate cells and T helper 1 responses, accelerating viral clearance and curtailing chronic inflammation (355).

### **Limitations**

RRMS phenotype was only identified by the use of DMTs among people covered by public drug insurance. The vast majority (70.4%) of cohort members had drug insurance coverage for a portion of the second follow-up (1997-2014). Thus, 30% of study subjects were not covered during this follow-up, which prevented the identification of RRMS in this subgroup. Moreover, the median cumulative duration of drug insurance among those covered during 1997-2014 was 6.9 years, compared to 12.6 years for subjects identified as having at least one DMT prescription. The latter were thus covered by the public drug insurance for a longer period, and were more deprived socially and materially (Table 6.S4). Public insurance is offered to individuals who do not have a private plan, either purchased or offered by employers. To address this issue, we used inverse probability weighting and achieved a balance of sociodemographic characteristics in the subset of those who had drug insurance versus others.

RRMS cases identified may not be representative of all RRMS cases. The former were indeed subjects who were taking DMTs through public drug insurance. A third of Canadian MS patients do not take or stop DMTs in spite of a RRMS diagnosis and physician prescriptions, either by choice (for instance, young women planning a pregnancy or breast feeding) or because of adverse effects (356). Among study subjects with drug insurance coverage, 40.5% of incident MS cases identified using Marrie's definition had at least one DMT prescription (528/1,305, Figure 6.S1). A recent study in the US suggested that only 35% of new cases of MS were taking DMTs 1-5.5 years after diagnosis (357). Thus among

persons with drug insurance coverage in our study, the proportion of RRMS cases under DMT was within the range of expected values. Despite this limitation, we identified 576 RRMS cases, allowing for the first time to estimate risk relating BCG vaccination to a specific MS phenotype.

Administrative health data were not available before 1983. Between 1983 and 1996, only 274 incident MS cases were detected, in line with the lower incidence rate in younger people. Before 1983, when cohort members were even younger (before 9 to 13 years old), few cases would have been expected. The impact of this limitation on the validity of our risk estimates and on statistical power was likely minimal.

In the earlier follow-up (1983-1996), drug data were not available. However, all DMTs used in our definition (Table 6.S1 and 6.S2) were approved in Canada only in period 2 (1997-2014). This period coincided with the advent of the public drug insurance coverage.

We could not differentiate between RRMS, SPMS and PPMS within the unspecified MS subgroup (n=953 in period 2). We investigated whether the excess risk was concentrated in a specific birth cohort or in a subgroup of cases with later MS onset, assuming that some of the older ones might be PPMS. If the sociodemographic characteristics (3, 101, 349) and pathogenesis of RRMS differ from PPMS through two types of inflammation (23), we may hypothesize different associations with BCG vaccination. The main distinctive characteristics available to us for distinguishing between RRMS and PPMS patients among those who did not take DMTs is an expected older age at diagnosis in the latter subgroup and different sex ratios. Our results suggest that the increased risk observed in the unspecified MS subgroup cannot be attributed to an identifiable subset of MS cases, such as PPMS.

## **Strengths**

The present study is based on a vast and highly representative population-based birth cohort (303). With 1,707 incident MS cases, it is the largest study to date on this topic. The available data allowed us to count on a reliable assessment of exposure to BCG through a provincial registry. Moreover, we could ascertain MS outcomes using a validated definition and a new one based on DMTs utilization. For the first time, we studied RRMS specifically, the most common MS phenotype. This is important given the different clinical courses of MS phenotypes, which would be expected to have distinct determinants.

We applied a formal process to carefully identify covariates for adjustment using a DAG (337). Nevertheless, we observed marginal differences between crude and adjusted hazard ratios, suggesting that these factors did not confound the associations under study.

Finally, several sub-analyses were undertaken to ensure the robustness of our findings, and to inform their interpretation.

## **6.7 Conclusions**

Within a large birth cohort, we observed no association between either BCG vaccination or age at vaccination and the incidence of RRMS, including pediatric, young adult and adult cases. An increased risk of later-diagnosis MS in the subgroup of cases with unknown phenotypes was observed, but interpretation is hampered by the probable heterogeneity of MS subtypes represented. However, this increased risk points out to the need to conduct further research within clinical cohorts or MS registries that allow for the identification of specific MS phenotypes.

## **ACKNOWLEDGEMENTS**

We gratefully acknowledge the *Institut de la statistique du Québec*, the *Régie de l'assurance maladie du Québec* and the personnel of the data access centre (*Centre d'accès aux données de recherche de l'Institut de la statistique du Québec, CADRISQ*) for their technical help and support.

## **CONFLICT OF INTEREST**

The authors declare that they have no conflict of interest.

## **AUTHOR CONTRIBUTIONS**

P.C. conceived the analytical methodology for the multiple sclerosis study, performed the literature review, conducted statistical analyses, and drafted all versions of the manuscript.

M.E.P. participated in the conceptualization of the birth cohort (QBCIH), co-supervised P.C.'s doctoral thesis work, contributed to the manuscript, and approved the final version.

C.W. provided expertise in the field of multiple sclerosis, contributed to the manuscript, and approved the final manuscript. N.A. provided expertise in immunology, especially in

immunopathology of multiple sclerosis, contributed to the manuscript, and approved the final manuscript. P.D. provided clinical expertise in the field of multiple sclerosis,

contributed to the manuscript, and approved the final manuscript. A.B. participated in the conceptualization of the birth cohort (QBCIH), provided statistical expertise, contributed

to the manuscript, and approved the final version. H.R. was responsible for data preparation and management, provided statistical expertise, contributed to the manuscript, and

approved the final manuscript. S.S. provided expertise in immunology, contributed to the manuscript, and approved the final version. As principal investigator, M.C.R.

conceptualized and led the development of the QBCIH, obtained and managed the research funds. She supervised P.C.'s doctoral research including the literature review and statistical analyses, reviewed all manuscript drafts, and approved the final manuscript. P.C. and M.C.R. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

### **ROLE OF THE FUNDER/SPONSOR**

The funders had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

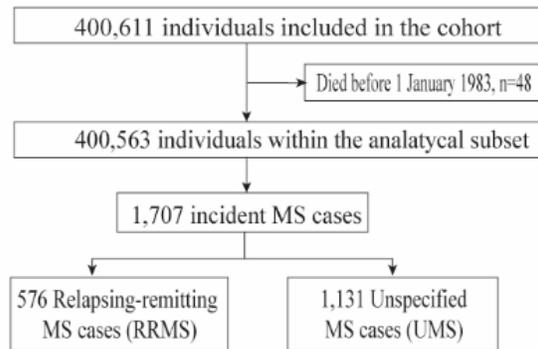
### **DATA AVAILABILITY STATEMENT**

The data that support the findings of this study are governed by strict data privacy considerations, require governmental bodies' approval, and so are not publicly available. For enquiries, please contact the last author.

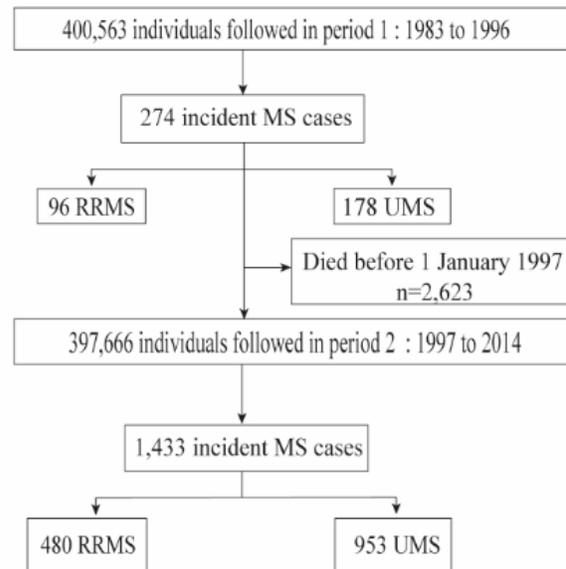
### **FUNDING INFORMATION**

Establishment of the QBCIH was supported by an infrastructure grant from the Canada Foundation for Innovation & the Québec Ministry of Education, Leisure and Sports (#12532), research grants from the Canadian Institutes of Health Research (#MOP-97777, #MCH-97593, #MOP-142705, #PJT-159647), Fonds de recherche du Québec-Santé (FRQS, #16227), and Multiple Sclerosis Society of Canada (#2435). Drs. Benedetti and Parent are recipients of career awards from the FRQS. P. Corsenac is the recipient of a PhD fellowship from FRQS.

A) Entire follow-up

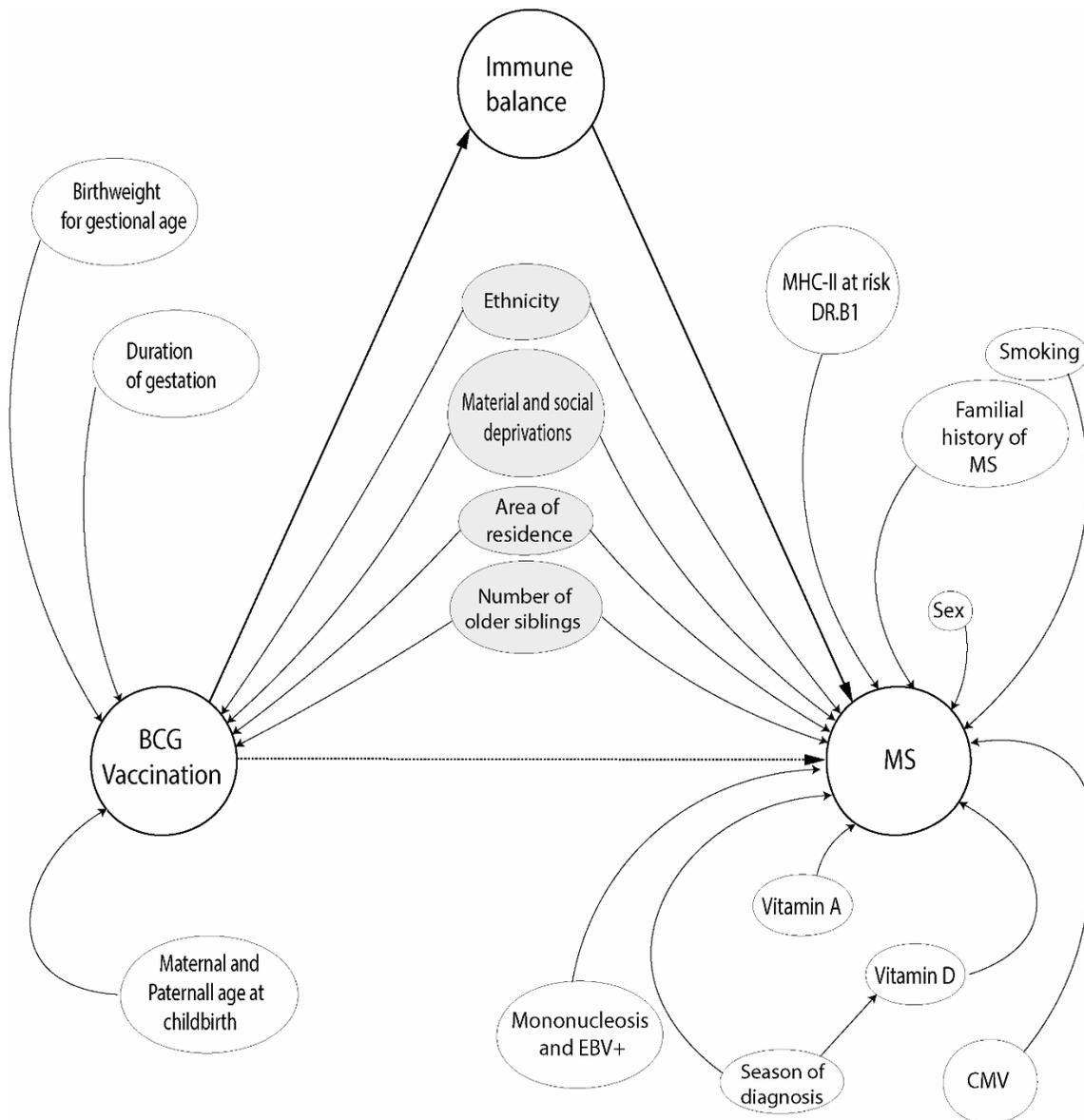


B) Follow-up divided in two periods



**Figure 6.1: Flowchart of participants' inclusions and exclusions in the analytical subset from the Quebec Birth Cohort on Immunity and Health (1983-2014), Quebec, Canada.**

**Legend for Figure 6.1:** Unspecified multiple sclerosis (MS) cases comprises unknown clinical forms of MS. Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



**Figure 6.2: Directed acyclic graph of the association between BCG vaccination and multiple sclerosis.**

**Legend for Figure 6.2:** Grey shading indicates confounding factors for the causal relationship between BCG vaccination and multiple sclerosis.

**Table 6.1: Description of the study population according to BCG vaccination status, Quebec Birth Cohort on Immunity and Health, Quebec, Canada.**

Variables	Overall study population		Vaccinated population		Unvaccinated population		Subset with uncertain vaccination	
	N	% or Mean $\pm$ Std	N	% or Mean $\pm$ Std	N	% or Mean $\pm$ Std	N	% or Mean $\pm$ Std
	N = 400 563 (100%)		N =178 335 (44.5%)		N = 209 705 (52.4%)		N = 12 523 (3.1%)	
<b>Year of birth</b>								
1970	83 187	20.8	39 977	22.4	40 397	19.3	2 813	22.5
1971	80 837	20.2	34 855	19.5	43 156	20.6	2 826	22.6
1972	76 724	19.2	32 449	18.2	41 673	19.9	2 602	20.8
1973	77 201	19.3	33 298	18.7	40 842	19.5	3 061	24.4
1974	82 614	20.6	37 756	21.2	43 637	20.8	1 221	9.8
<b>Sex</b>								
Male	205 950	51.4	91 305	51.2	108 246	51.6	6 389	51.0
Female	194 623	48.6	87 030	48.8	101 459	48.4	6 134	49.0
<b>Duration of gestation (in weeks)</b>								
Mean $\pm$ Std	400 556	39.2 $\pm$ 1.7	178 334	39.2 $\pm$ 1.7	209 699	39.2 $\pm$ 1.7	12 523	39.2 $\pm$ 1.7
Missing	7	-	1	-	6	-	0	-
<b>Birthweight (in grams)</b>								
Mean $\pm$ Std	400 530	3240 $\pm$ 508	178 327	3260 $\pm$ 491	209 681	3223 $\pm$ 524	12 522	3259 $\pm$ 487
Missing	33	-	8	-	24	-	1	-
<b>Number of older siblings</b>								
None	167 645	41.9	76 436	42.9	85 617	40.8	5 592	44.7
1 or more	221 215	55.2	100 032	56.1	114 595	54.6	6 588	52.6
Missing	11 703	2.9	1 867	1.0	9 493	4.5	343	2.7
<b>Area of residence 1987 <sup>†</sup></b>								
Rural	126 651	31.6	71 520	40.1	51 086	24.4	4 045	32.3
Urban	270 864	67.6	106 126	59.5	156 357	74.6	8 381	66.9
Missing	3 048	0.8	689	0.4	2 262	1.1	97	0.8
<b>Area of residence 1997 <sup>†</sup></b>								
Rural	96 501	24.1	53 558	30.0	39 924	19.0	3 019	24.1

Urban	302 148	75.4	124 441	69.8	168 241	80.2	9 466	75.6
Missing	1 914	0.5	336	0.2	1 540	0.7	38	0.3
<b>Maternal age at childbirth (years)</b>								
Mean ± Std	391 145	26.6±5.2	177 027	26.5±5.1	201 922	26.7±5.3	12 196	26.3±5.4
Missing	9 418	2.4	14 641	8.2	17 939	8.6	327	2.6
<b>Paternal age at childbirth (years)</b>								
Mean ± Std	382 219	29.8±5.9	174 841	29.3±5.8	196 151	29.7±6.0	11 227	29.3±6.0
Missing	18 344	4.6	11 272	6.3	14 547	6.9	1 296	10.3
<b>Maternal birth place as proxy of ethnicity</b>								
In Quebec	341 434	85.2	168 960	94.7	161 005	76.8	11 469	91.6
Outside	48 106	12.0	7 612	4.3	39 809	19.0	685	5.5
Missing	11 023	2.8	1 763	1.0	8 891	4.2	369	2.9
<b>Paternal birth place as proxy of ethnicity</b>								
In Quebec	329 375	82.2	165 463	92.8	153 384	73.1	10 528	84.1
Outside	51 054	12.7	8 793	4.9	41 602	19.8	659	5.3
Missing	20 134	5.0	4 079	2.3	14 719	7.0	1 336	10.7
<b>Material deprivation 1987</b>								
Very privileged	66 468	16.6	20 653	11.6	43 970	21.0	1 845	14.7
Privileged	77 423	19.3	32 933	18.5	41 943	20.0	2 547	20.3
Moderately privileged	77 985	19.5	36 792	20.6	38 717	18.5	2 476	19.8
Deprived	77 419	19.3	37 569	21.1	37 476	17.9	2 374	19.0
Very deprived	80 543	20.1	41 428	23.2	36 461	17.4	2 654	21.2
Missing	20 725	5.2	8 960	5.0	11 138	5.3	627	5.0
<b>Material deprivation 1997</b>								
Very privileged	67 781	16.9	23 540	13.2	42 241	20.1	2 000	16.0
Privileged	76 422	19.1	32 583	18.3	41 500	19.8	2 339	18.7
Moderately privileged	80 651	20.1	38 115	21.4	39 975	19.1	2 561	20.5
Deprived	80 398	20.1	39 573	22.2	38 241	18.2	2 584	20.6
Very deprived	78 695	19.6	39 400	22.1	36 675	17.5	2 620	20.9
Missing	16 616	4.1	5 124	2.9	11 073	5.3	419	3.3
<b>Social deprivation 1987</b>								
Very privileged	84 429	21.1	40 807	22.9	41 189	19.6	2 433	19.4
Privileged	77 252	19.3	37 638	21.1	37 266	17.8	2 348	18.7
Moderately privileged	82 113	20.5	38 993	21.9	40 606	19.4	2 514	20.1

Deprived	73 018	18.2	29 442	16.5	41 152	19.6	2 424	19.4
Very deprived	63 026	15.7	22 495	12.6	38 354	18.3	2 177	17.4
Missing	20 725	5.2	8 960	5.0	11 138	5.3	627	5.0
<b>Social deprivation 1997</b>								
Very privileged	75 220	18.8	34 213	19.2	38 800	18.5	2 207	17.6
Privileged	75 462	18.8	36 951	20.7	36 170	17.2	2 341	18.7
Moderately privileged	74 855	18.7	35 369	19.8	37 120	17.7	2 366	18.9
Deprived	76 140	19.0	33 171	18.6	40 558	19.3	2 411	19.3
Very deprived	82 270	20.5	33 507	18.8	45 984	21.9	2 779	22.2
Missing	16 616	4.1	5 124	2.9	11 073	5.3	419	3.3
<b>Income 1991 †</b>								
Less than \$35,333	95 454	23.8	47683	26.7	44660	21.3	3 111	24.8
\$35,333 to \$40,066	99 943	25.0	49136	27.6	47811	22.8	2 996	23.9
\$40,067to \$47,058	97 557	24.4	45901	25.7	48509	23.1	3 147	25.1
\$47,059 or more	98 005	24.5	33 141	18.6	61 827	29.5	3 037	24.3
Missing	9 604	2.4	2 474	1.4	6 898	3.3	232	1.9
<b>Income 1996 §</b>								
less of \$36,608	95 133	23.7	42 348	23.7	49 649	23.7	3 136	25.0
\$36,608 to \$41,608	97 597	24.4	47 441	26.6	47 349	22.6	2 807	22.4
\$41,609 to \$48,172	96 548	24.1	48 192	27.0	45 118	21.5	3 238	25.9
\$48,173 and more	96 909	24.2	36 083	20.2	57 801	27.6	3 025	24.2
Missing	14 376	3.6	4 271	2.4	9 788	4.7	317	2.5
<b>Vaccination status</b>								
Unvaccinated	209 705	52.4	0	0	209 705	100	0	0.0
Vaccinated	178 335	44.5	178 335	100	0	0	0	0.0
Uncertain	12 523	3.1	0	0	0	0	12 523	100.0
<b>Age at BCG vaccination</b>								
Unvaccinated	209 705	52.4	0	0	209 705	100	0	0.0
< 1 year old	156 497	39.1	156 497	87.8	0	0	0	0.0
≥ 1 year old	21 838	5.4	21 838	12.2	0	0	0	0.0
Uncertain vaccination	12 523	3.1	0	0	0	0	12 523	100.0
<b>Number of BCG vaccinations</b>								
Unvaccinated	209 705	52.4	0	0	209 705	100	0	0.0
1	167 458	41.8	167 458	93.9	0	0	0	0.0

$\geq 2$	10 877	2.7	10 877	6.1	0	0	0	0.0
Uncertain vaccination	12 523	3.1	0	0	0	0	12523	100.0

† Determined using the second character of the subjects' residential postal code (0: rural,  $\geq 1$ : urban).

‡ Based on 1991 census.

§ Based on 1996 census.

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 6.2: Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident multiple sclerosis between January 1<sup>st</sup> 1983 and December 31<sup>st</sup> 2014, Quebec Birth Cohort on Immunity and Health (1983-2014), Quebec, Canada.**

Type of MS	Subjects with MS N	Person-years	Crude Hazard Ratio (95% CI)	Adjusted Hazard Ratio (95% CI)
<b>Total analytical subset : 1983 to 2014</b>				
Total incident cases	1707	16 909 983		
<b>BCG vaccination status</b>				
Unvaccinated	819	8 848 351	1 (Ref)	1 (Ref)
Vaccinated	888	8 061 632	<b>1.19 (1.08-1.31)</b>	<b>1.16 (1.05-1.28)<sup>†</sup></b>
<b>Age at BCG vaccination</b>				
Unvaccinated	819		1 (Ref)	1 (Ref)
< 1 year old	793	7 064 376	<b>1.21 (1.09-1.33)</b>	<b>1.18 (1.06-1.31)<sup>†</sup></b>
≥ 1 year old	95	997 256	0.98 (0.78-1.22)	0.96 (0.77-1.21) <sup>†</sup>
Relapsing-Remitting	576	16 909 983		
<b>BCG vaccination status</b>				
Unvaccinated	295	8 848 351	1 (Ref)	1 (Ref)
Vaccinated	281	8 061 632	1.01 (0.86-1.19)	1.01 (0.85-1.20) <sup>‡</sup>
<b>Age at BCG vaccination</b>				
Unvaccinated	295	8 848 351	1 (Ref)	1 (Ref)
< 1 year old	256	7 064 376	1.06 (0.89-1.26)	1.05 (0.88-1.26) <sup>‡</sup>
≥ 1 year old	25	997 256	0.86 (0.58-1.28)	1.05 (0.88-1.26) <sup>‡</sup>
Unspecified <sup>§</sup>	1131	16 909 983		
<b>BCG vaccination status</b>				
Unvaccinated	524	8 848 351	1 (Ref)	1 (Ref)
Vaccinated	607	8 061 632	<b>1.29 (1.14-1.44)</b>	<b>1.25 (1.10-1.41)<sup>†</sup></b>

<b>Age at BCG vaccination</b>				
Unvaccinated	524	8 848 351	1 (Ref)	1 (Ref)
< 1 year old	541	7 064 376	<b>1.29 (1.14-1.47)</b>	<b>1.25 (1.09-1.43)</b> †
≥ 1 year old	62	997 256	1.04 (0.79-1.36)	1.02 (0.77-1.34) †

BCG, Bacillus Calmette–Guerin; CI, confidence interval.

† Models were adjusted for maternal and paternal birthplace as a proxy of ethnicity, social deprivation in 1987, number of older siblings, and area of residence in 1987. Material deprivation in 1987 was excluded as confounder because specific hazard was not proportional.

‡ Models were adjusted for maternal and paternal birthplace as a proxy of ethnicity, material and social deprivation in 1987, number of older siblings, and area of residence in 1987.

§ Unknown clinical form of MS, comprising relapsing-remitting- and primary progressive MS

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 6.3: Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident multiple sclerosis between January 1<sup>st</sup> 1983 and December 31<sup>st</sup> 2014, over two periods, Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.**

Type of MS	Subjects with MS	Person-years	Crude Hazard Ratio	Adjusted Hazard Ratio	Subjects with MS	Person-years	Crude Hazard Ratio	Adjusted Hazard Ratio <sup>†</sup>
	N		(95% CI)	(95% CI)	N		(95% CI)	(95% CI)
	<b>Period 1: 1983 to 1996</b>				<b>Period 2: 1997 to 2014</b>			
Total incident cases	274	5 595 901			1 433	11 283 033		
<b>BCG vaccination status</b>								
Unvaccinated	146	2 929 945	1 (Ref)	1 (Ref)	673	5 903 119	1 (Ref)	1 (Ref)
Vaccinated	128	2 665 956	0.94 (0.74-1.20)	0.94 (0.74-1.21) <sup>¶</sup>	760	5 379 914	<b>1.25 (1.13-1.39)</b>	<b>1.22 (1.09-1.36)</b>
<b>Age at BCG vaccination</b>								
Unvaccinated	146	2 929 945	1 (Ref)	1 (Ref)	673	5 903 119	1 (Ref)	1 (Ref)
< 1 year old	109	2 338 004	0.90 (0.70-1.16)	0.89 (0.68-1.17) <sup>¶</sup>	684	4 712 725	<b>1.28 (1.15-1.43)</b>	<b>1.25 (1.12-1.40)</b>
≥ 1 year old	19	327 952	1.13 (0.69-1.84)	1.13 (0.69-1.86) <sup>¶</sup>	76	667 189	0.94 (0.73-1.21)	0.93 (0.72-1.20)
Relapsing-Remitting	96	5 595 901			480	11 283 033		
<b>BCG vaccination status</b>								
Unvaccinated	51	2 929 945	1 (Ref)	1 (Ref)	244	5 903 119	1 (Ref)	1 (Ref)
Vaccinated	45	2 665 956	0.93 (0.63-1.39)	0.96 (0.63-1.45) <sup>†</sup>	236	5 379 914	1.03 (0.86-1.23)	1.02 (0.85-1.23)
<b>Age at BCG vaccination</b>								
Unvaccinated	51	2 929 945	1 (Ref)	1 (Ref)	244	5 903 119	1 (Ref)	1 (Ref)
< 1 year old	45 <sup>§</sup>	2 338 004	0.99 (0.66-1.49)	1.02 (0.66-1.56) <sup>†</sup>	211	4 712 725	1.07 (0.88-1.29)	1.06 (0.87-1.29)
≥ 1 year old		327 952	0.63 (0.21-1.87)	0.63 (0.21-1.90) <sup>†</sup>	25	667 189	0.92 (0.61-1.40)	0.93 (0.61-1.42)
Unspecified ‡	178	5 595 901			953	11 283 033		
<b>BCG vaccination status</b>								
Unvaccinated	95	2 929 945	1 (Ref)	1 (Ref)	429	5 903 119	1 (Ref)	1 (Ref)
Vaccinated	83	2 665 956	0.95 (0.70-1.27)	0.94 (0.69-1.28) <sup>†</sup>	524	5 379 914	<b>1.38 (1.21-1.56)</b>	<b>1.33 (1.17-1.52)</b>

<b>Age at BCG vaccination</b>								
Unvaccinated	95	2 929 945	1 (Ref)	1 (Ref)	429	5 903 119	1 (Ref)	1 (Ref)
< 1 year old	68	2 338 004	0.85 (0.61-1.18)	0.83 (0.58-1.19) <sup>†</sup>	473	4 712 725	<b>1.41 (1.23-1.61)</b>	<b>1.36 (1.18-1.56)</b>
≥ 1 year old	11	327 952	1.40 (0.79-2.47)	1.39 (0.79-2.47) <sup>†</sup>	51	667 189	0.96 (0.70-1.32)	0.93 (0.68-1.28)

BCG, Bacillus Calmette–Guerin; CI, confidence interval.

<sup>†</sup> Models were adjusted for maternal and paternal birthplace as a proxy of ethnicity, material and social deprivation in 1987, number of older siblings, and area of residence in 1987.

<sup>‡</sup> Unknown clinical form of MS, comprising relapsing-remitting and primary progressive MS

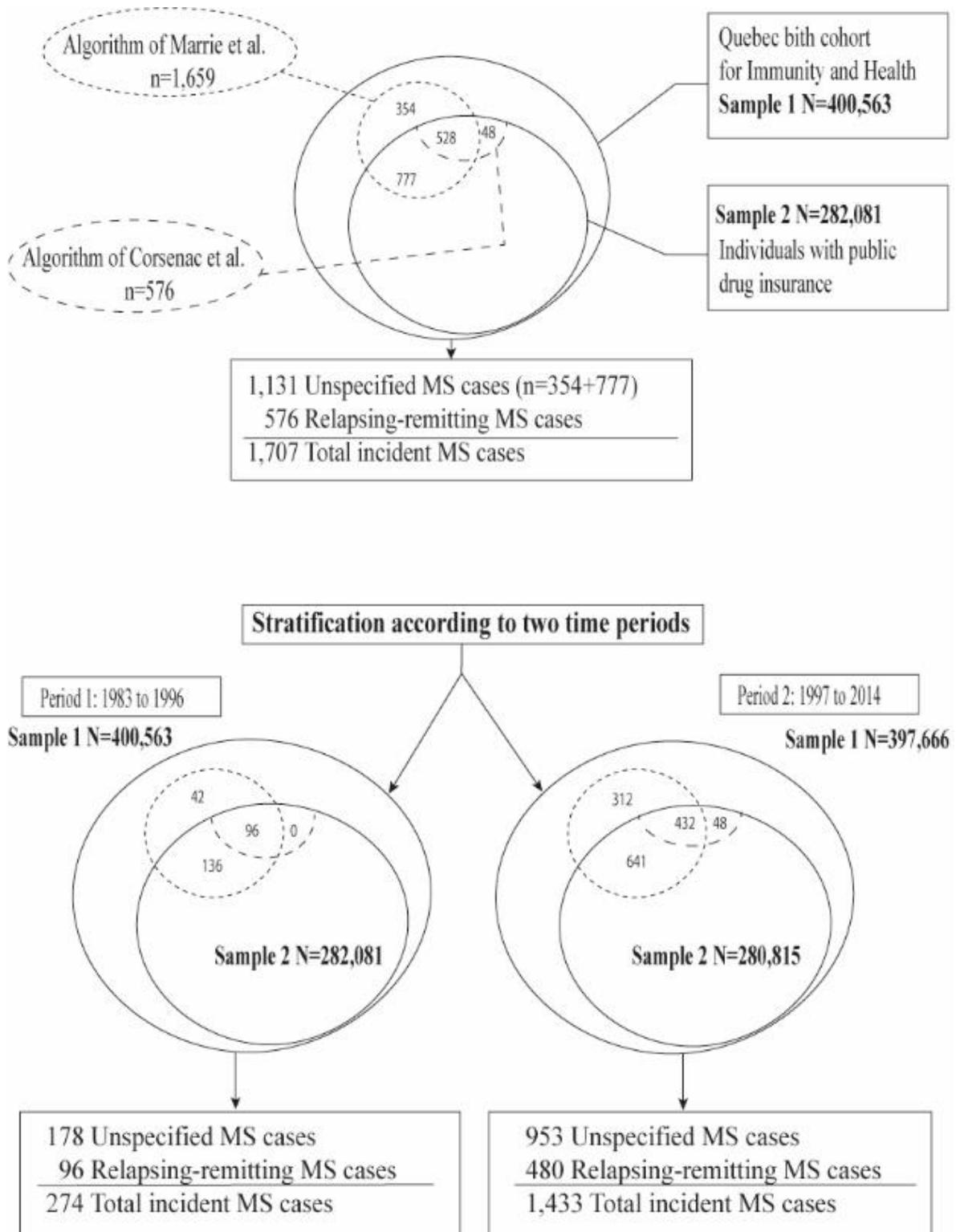
<sup>§</sup> One of the two categories had a total number of subjects <5. Both categories were grouped to respect disclosure standards of *Institut de la statistique du Québec*.

<sup>¶</sup> Models were adjusted for maternal and paternal birthplace as a proxy of ethnicity, social deprivation in 1987, and area of residence in 1987. Number of older siblings and material deprivation in 1987 were excluded as confounders because their specific hazards were not proportional.

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

## 6.8 Supporting information

BCG and multiple sclerosis, Corsenac et al. 2022 European Journal of Neurology



**Figure 6.S1** : MS definitions and proportion of participants included in the analytical subsets from the Quebec Birth Cohort on Immunity and Health (1983-2014) and from drug insurance coverage, Quebec, Canada.

Unspecified multiple sclerosis (MS) cases represented unknown clinical forms of MS.

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 6.S1:** Disease-modifying therapies used in our definition for identifying cases of relapsing-remitting MS. These were authorized for use in Quebec between 1997 and 2014, and included in the public drug insurance database.

Therapeutic classes	International Nonproprietary Name (INN)	INN codes	References
<b>First generation and first-line treatments</b>			
Immunomodulators	Glatiramer acetate	47253	(358)
	Interferon bêta-1a	47280, 46422 †	(359)
	Interferon bêta-1b	47185	(360, 361)
<b>Second-line treatments with disease course modifiers</b>			
Anti-CD52 monoclonal antibody	Alemtuzumab	48016	(362)
Immunomodulators (second generation)	Dimethylfumarate	47954	(363)
Sphingosine-1-Phosphate receptor modulator	Fingolimod	47857	(364)
Anti-Alpha4 integrin monoclonal antibody	Natalizumab	47668	(365)
Inhibitor of Pyrimidine de novo synthesis	Teriflunomide	47989	(366)

† RebifMD was not authorized in the Quebec market in this period, but it received an Exception drug status in 1998 (367).

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 6.S2** : Factors considered in the directed acyclic graph of a causal association between BCG vaccine and the onset of MS, and supporting references

<b>Variables</b>	<b>Causal association with BCG</b>	<b>References</b>	<b>Causal association with MS</b>	<b>References</b>
<b>Season of diagnosis</b>	No	-	Yes	(163)
<b>Material and social deprivation</b>	Yes	Income (307); education (368-370); socioeconomic status (369-372)	Yes	(373, 374)
<b>Birth weight for gestational age</b>	Yes	(307)	No	-
<b>Duration of gestation</b>	Yes	Post-term (307)	No	-
<b>Number of older siblings</b>	Yes	(307)	Yes	(375)
<b>Area of residence</b>	Yes	(307)	Yes, pollution marker	(376)
<b>Maternal age at childbirth</b>	Yes	(307)	No	-
<b>Paternal age at childbirth</b>	Yes	(307)	No	-
<b>Smoking (passive + active)</b>	No	-	Yes	(151)
<b>Familial history of MS</b>	No	-	Yes	(377, 378)
<b>Ethnicity</b>	Yes	(307)	Yes	(109)
<b>Sex</b>	No	(282, 307)	Yes	(51)
<b>Lack of Vitamin D</b>	No	-	Yes	(121)
<b>MHC-II at risk: DR.B1</b>	No	-	Yes	(82, 111, 379)
<b>CMV</b>	No	-	Yes	(19)
<b>Mononucleosis or EBV+</b>	No	-	Yes	(380)
<b>Lack of vitamin A</b>	No	-	Yes	(381, 382)

**Table 6.S3:** Non-weighted description of the study population (N= 397,666) in the first imputation dataset within period 2 (1997-2014), according to clinical forms of multiple sclerosis (MS) in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

<i>Variables</i>	<i>Overall study population</i>		<i>Relapsing-Remitting MS</i>		<i>Unspecified MS</i>		<i>Unspecified MS, age at diagnosis &lt;29.5</i>		<i>Unspecified MS, age at diagnosis 29.5-33.4</i>		<i>Unspecified MS, age at diagnosis 33.5-44.5</i>	
	N = 397 666 (100%)		N = 480 (100%)		N = 953 (100%)		N = 239 (100%)		N = 238 (100%)		N = 476 (100%)	
	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std
<b>Year of birth</b>												
1970	82 519	20.8	119	24.8	204	21.4	34	14.23	53	22.27	117	24.6
1971	80 146	20.2	91	19.0	201	21.1	35	14.64	55	23.11	111	23.3
1972	76 163	19.2	92	19.2	180	18.9	54	22.59	43	18.07	83	17.4
1973	76 684	19.3	87	18.1	184	19.3	64	26.78	40	16.81	80	16.8
1974	82 154	20.7	91	19.0	184	19.3	52	21.76	47	19.75	85	17.9
<b>Sex</b>												
Male	203 844	51.3	122	25.4	229	24.0	49	20.50	58	24.37	122	25.63
Female	193 822	48.7	358	74.6	724	76.0	190	79.50	180	75.63	354	74.37
<b>Duration of gestation (in weeks)</b>												
Mean ± Std	397 666	39.2±1.7	480	38.9±1.9	953	39.2±1.7	239	39.1±1.8	238	39.3±1.6	476	39.2±1.7
<b>Birthweight (in grams)</b>												
Mean ± Std	397 666	3240±508	480	3210±508	953	3207±492	239	3206±511	238	3183±491	476	3222±483
<b>Number of older siblings</b>												
none	171 920	43.2	180	37.5	403	42.3	110	46.0	105	44.1	188	39.5
1 or more	225 746	56.8	300	62.5	550	57.7	129	54.0	133	55.9	288	60.5
<b>Area of residence 1987<sup>a</sup></b>												
Rural	126 116	31.7	170	35.4	315	33.1	77	32.2	81	34.0	157	33
Urban	271 550	68.3	310	64.6	638	66.9	162	67.8	157	66	319	67

**Area of residence 1997 †**

Rural	95 740	24.1	140	29.2	249	26.1	53	22.2	56	23.5	140	29.4
Urban	301 926	75.9	340	70.8	704	73.9	186	77.8	182	76.5	336	70.6

**Maternal age at childbirth (years)**

Means ± Std	397 666	26.6±5.2	480	26.8±5.4	953	26.6±5.1	239	26.5±4.6	238	26.4±5.1	476	26.8±5.3
-------------	---------	----------	-----	----------	-----	----------	-----	----------	-----	----------	-----	----------

**Paternal age at childbirth (years)**

Means ± Std	397 666	29.4±5.9	480	29.7±6.1	953	29.3±5.7	239	29.3±5.4	238	29.1±5.8	476	29.4±5.8
-------------	---------	----------	-----	----------	-----	----------	-----	----------	-----	----------	-----	----------

**Maternal birth place as proxy of ethnicity**

In Quebec	347 903	87.5	428	89.2	855	89.7	209	87.5	211	88.7	435	91.4
Outside	49 763	12.5	52	10.8	98	10.3	30	12.5	27	11.3	41	8.6

**Paternal birth place as proxy of ethnicity**

In Quebec	343 749	86.4	423	88.1	848	89.0	211	88.3	208	87.4	429	90.1
Outside	53 917	13.6	57	11.9	105	11.0	28	11.7	30	12.6	47	9.9

**Material deprivation 1987**

Very privileged	69 723	17.5	70	14.6	141	14.8	39	16.3	32	13.5	70	14.7
Privileged	81 065	20.4	88	18.3	183	19.2	45	18.8	50	21.0	88	18.5
Moderately privileged	81 488	20.5	108	22.5	226	23.7	67	28.0	48	20.2	111	23.3
Deprived	81 000	20.4	86	17.9	199	20.9	48	20.1	55	23.1	96	20.2
Very deprived	84 390	21.2	128	26.7	204	21.4	40	16.7	53	22.3	111	23.3

**Material deprivation 1997**

Very privileged	70 801	17.8	65	13.5	139	14.6	37	15.5	39	16.4	63	13.2
Privileged	78 910	19.8	90	18.8	202	21.2	54	22.6	41	17.2	107	22.5
Moderately privileged	83 454	21.0	105	21.9	203	21.3	54	22.6	51	21.4	98	20.6
Deprived	82 945	20.9	97	20.2	228	23.9	57	23.9	60	25.2	111	23.3
Very deprived	81 556	20.5	123	25.6	181	19.0	37	15.5	47	19.8	97	20.4

**Social deprivation 1987**

Very privileged	88 385	22.2	98	20.4	208	21.8	58	24.3	52	21.85	98	20.6
Privileged	80 767	20.3	91	19.0	200	21.0	43	18.0	55	23.11	102	21.4

Moderately privileged	85 872	21.6	119	24.8	223	23.4	52	21.8	58	24.37	113	23.7
Deprived	76 588	19.3	86	17.9	166	17.4	45	18.8	33	13.87	88	18.5
Very deprived	66 054	16.6	86	17.9	156	16.4	41	17.2	40	16.81	75	15.8
<b>Social deprivation 1997</b>												
Very privileged	78 070	19.6	100	20.8	177	18.6	47	19.7	37	15.55	93	19.5
Privileged	77 959	19.6	92	19.2	213	22.4	53	22.2	51	21.43	109	22.9
Moderately privileged	77 378	19.5	95	19.8	177	18.6	41	17.2	42	17.65	94	19.8
Deprived	78 844	19.8	81	16.9	195	20.5	53	22.2	68	28.57	74	15.6
Very deprived	85 415	21.5	112	23.3	191	20.0	45	18.8	40	16.81	106	22.3
<b>Income 1991 ‡</b>												
Less than \$35,333	97 300	24.5	139	29.0	237	24.9	53	22.18	59	24.8	125	26.3
\$35,333 to 40,066	101 238	25.5	127	26.5	243	25.5	65	27.20	61	25.6	117	24.6
\$40,067 to 47,058	99 096	24.9	112	23.3	263	27.6	67	28.03	69	29.0	127	26.7
\$47,059 and more	100 032	25.2	102	21.3	210	22.0	54	22.59	49	20.6	107	22.5
<b>Income 1996 §</b>												
Less than \$36,608	98 279	24.7	144	30.0	257	27.0	64	26.8	58	24.4	135	28.4
\$36,608 to 41,608	99 944	25.1	126	26.3	230	24.1	58	24.3	63	26.5	109	22.9
\$41,609 to 48,172	99 229	25.0	105	21.9	260	27.3	68	28.5	64	26.9	128	26.9
\$48,173 and more	100 214	25.2	105	21.9	206	21.6	49	20.5	53	22.3	104	21.9
<b>Vaccination status</b>												
Unvaccinated	208 272	52.4	244	50.8	429	45.0	109	45.6	113	47.5	207	43.5
Vaccinated	189 394	47.6	236	49.2	524	55.0	130	54.4	125	52.5	269	56.5
<b>Age at BCG vaccination</b>												
Unvaccinated	208 272	52.4	244	50.8	429	45.0	109	45.6	113	47.5	207	43.5
< 1 year old	166 106	41.8	211	44.0	473	49.6	120	50.2	114	47.9	239	50.2

≥ 1 year old	23 288	5.9	25	5.2	51	5.4	10	4.2	11	4.6	30	6.3
<b>Number of BCG vaccination</b>												
Unvaccinated	208 355	52.4	244	50.8	429	45.0	109	45.6	113	47.5	207	43.5
1	177 665	44.7	214	44.6	491	51.5	124	51.9	116	48.7	251	52.7
≥ 2	11 646	2.9	22	4.6	33	3.5	6	2.5	9	3.8	18	3.8

† Determined using the second character of subjects' residential postal code (0: rural, ≥1: urban).

‡ Based on 1991 census.

§ Based on 1996 census.

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 6.S4:** Description of the study population according to drug insurance coverage and by clinical form of MS in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

Variables	Overall study population		Study population covered by medication insurance		RRMS within population covered by medication insurance		Unspecified MS within population covered by medication insurance		Study population not covered by medication insurance	
	N = 400 563 (100%)		N = 282 081 (100%)		N = 576 (100%)		N = 777 (100%)		N = 118 482 (100%)	
	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std	N	F (%) or Mean ± Std
<b>Year of birth</b>										
1970	83 187	20.8	55 033	19.5	146	25.3	162	20.8	28 154	23.8
1971	80 837	20.2	55 213	19.6	112	19.4	158	20.3	25 624	21.6
1972	76 724	19.2	53 945	19.1	111	19.3	147	18.9	22 779	19.2
1973	77 201	19.3	56 353	20.0	104	18.1	161	20.7	20 848	17.6
1974	82 614	20.6	61 537	21.8	103	17.9	149	19.2	21 077	17.8
<b>Sex</b>										
Male	205 950	51.4	143 722	51.0	146	25.3	183	23.6	62 218	52.5
Female	194 623	48.6	138 359	49.0	430	74.7	594	76.4	56 264	47.5
<b>Duration of gestation (in weeks)</b>										
Mean ± Std	400 556	39.2±1.7	282 075	39.2±1.7	576	39±1.9	777	39.2±1.8	118 481	39.2±1.7
Missing	7	-	6	-	-	-	-	-	1	-
<b>Birthweight (in grams)</b>										
Mean ± Std	400 530	3240±508	282 057	3229±512	576	3222±500	777	3185±478	118 473	3267±498
Missing	33	-	24	-	-	-	-	-	9	-
<b>Number of older siblings</b>										
none	167 645	41.9	116 493	41.3	212	36.8	329	42.3	51 152	43.2
At least 1	118 922	29.7	81 564	28.9	347	60.2	416	53.5	37 358	31.5
2 or more	102 293	25.5	75 002	26.6	17	3	32	4.1	27 291	23.0

Missing	11 703	2.9	9 022	3.2	-	-	-	-	2 681	2.3
<b>Area of residence 1987 †</b>										
Rural	126 651	31.6	94 054	33.3	198	34.4	263	33.8	32 597	27.5
Urban	270 864	67.6	186 532	66.1	376	65.3	510	65.8	84 332	71.2
Missing	3 048	0.8	1 495	0.5	2	0.3	4	0.5	1 553	1.3
<b>Area of residence 1997 †</b>										
Rural	96 501	24.1	72 510	25.7	164	28.5	206	26.5	23 991	20.2
Urban	302 148	74.4	208 905	74.1	411	71.4	569	73.2	93 243	78.7
Missing	1 914	0.5	666	0.2	1	0.2	2	0.3	1 248	1.1
<b>Maternal age at childbirth (years)</b>										
Mean ± Std	391 145	26.6±5.2	274 683	26.5±5.3	476	26.9±5.5	777	26.2±5.3	116 462	26.9±4.9
Missing	9 418	2.4	7 398	2.6	100	-	-	-	2 020	1.7
<b>Paternal age at childbirth (years)</b>										
Mean ± Std	382 219	29.8±5.9	266 877	29.5±6.1	476	29.8±6.1	777	29.2±5.9	115 342	29.5±5.5
Missing	18 344	4.6	15 204	5.4	100	-	-	-	3 140	2.7
<b>Maternal birth place as proxy of ethnicity</b>										
In Quebec	341 434	85.2	245 236	86.9	501	87.0	665	85.6	96 198	81.2
Outside	48 106	12.0	28 414	10.1	60	10.4	75	9.7	19 692	16.6
Missing	11 023	2.8	8 431	3.0	15	2.6	27	3.5	2 592	2.2
<b>Paternal birth place as proxy of ethnicity</b>										
In Quebec	329 375	82.2	235 257	83.4	490	85.1	647	83.3	94 118	79.4
Outside	51 054	12.7	30 476	10.8	62	10.8	73	9.4	20 578	17.4
Missing	20 134	5.0	16 348	5.8	24	4.2	57	7.3	3 786	3.2
<b>Material deprivation 1987</b>										
Very privileged	66 468	16.6	40 709	14.4	83	14.4	102	13.1	25 759	21.7
Privileged	77 423	19.3	52 401	18.6	100	17.4	140	18.0	25 022	21.1
Moderately privileged	77 985	19.5	55 279	19.6	126	21.9	168	21.6	22 706	19.2
Deprived	77 419	19.3	56 967	20.2	108	18.8	160	20.6	20 452	17.3
Very deprived	80 543	20.1	62 936	22.3	137	23.8	177	22.8	17 607	14.9

Missing	20 725	5.2	13 789	4.9	22	3.8	30	9.9	6 936	5.9
<b>Material deprivation 1997</b>										
Very privileged	67 781	16.9	42 082	14.9	85	14.8	95	12.2	25 699	21.7
Privileged	76 422	19.1	51 609	18.3	103	17.9	148	19.0	24 813	20.9
Moderately privileged	80 651	20.1	57 551	20.4	122	21.2	161	20.7	23 100	19.5
Deprived	80 398	20.1	59 999	21.3	120	20.8	187	24.1	20 399	17.2
Very deprived	78 695	19.6	62 300	22.1	138	24.0	160	20.6	16 395	13.8
Missing	16 616	4.1	8 540	3.0	8	1.4	26	3.3	8 076	6.8
<b>Social deprivation 1987</b>										
Very privileged	84 429	21.1	57 558	20.4	111	19.3	163	21.0	26 871	22.7
Privileged	77 252	19.3	53 711	19.0	101	17.5	151	19.4	23 541	19.9
Little privileged	82 113	20.5	57 232	20.3	143	24.8	170	21.9	24 881	21.0
Deprived	73 018	18.2	51 695	18.3	98	17.0	131	16.9	21 323	18.0
Very deprived	63 026	15.7	48 096	17.1	101	17.5	132	17.0	14 930	12.6
Missing	20 725	5.2	13 789	4.9	22	3.8	30	3.9	6 936	5.9
<b>Social deprivation 1997</b>										
Very privileged	75 220	18.8	51 007	18.1	116	20.1	143	18.4	24 213	20.4
Privileged	75 462	18.8	52 390	18.6	106	18.4	148	19.0	23 072	19.5
Moderately privileged	74 855	18.7	52 889	18.7	116	20.1	136	17.5	21 966	18.5
Deprived	76 140	19.0	54 700	19.4	105	18.2	160	20.6	21 440	18.1
Very deprived	82 270	20.5	62 555	22.2	125	21.7	164	21.1	19 715	16.6
Missing	16 616	4.1	8 540	3	8	1.4	26	3.3	8 076	6.8
<b>Income 1991 †</b>										
Less than \$35,333	95 454	23.8	73 433	26.0	156	27.1	197	25.4	22 021	18.6
\$35,333 to \$40,066	99 943	25.0	72 724	25.8	151	26.2	201	25.9	27 219	23.0
\$40,067 to \$47,058	97 557	24.4	68 846	24.4	134	23.3	212	27.3	28 711	24.2
\$47,059 and more	98 005	24.5	62 694	22.2	128	22.2	156	20.1	35 311	29.8
Missing	9 604	2.4	4 384	1.6	7	1.2	11	1.4	5 220	4.4
<b>Income 1996 §</b>										

Less than \$36,608	95 133	23.7	73 750	26.1	173	30.0	224	28.8	21 383	18.0
\$36,608 to \$41,608	97 597	24.4	71 964	25.5	143	24.8	184	23.7	25 633	21.6
\$41,609 to \$48,172	96 548	24.1	68 655	24.3	126	21.9	219	28.2	27 893	23.5
\$48,173 and more	96 909	24.2	62 082	22.0	125	21.7	137	17.6	34 827	29.4
Missing	14 376	3.6	5 630	2.0	9	1.6	13	1.7	8 746	7.4
<b>Vaccination status</b>										
Unvaccinated	209 705	52.4	145 345	51.5	295	51.2	364	46.8	64 360	54.3
Vaccinated	178 335	44.5	127 430	45.2	268	46.5	378	48.6	50 905	43.0
Uncertain vaccination	12 523	3.1	9 306	3.3	13	2.3	35	4.5	3 217	2.7
<b>Age at BCG vaccination</b>										
Unvaccinated	209 705	52.4	145 345	51.5	295	51.2	364	46.8	64 360	54.3
Uncertain vaccination	12 523	3.1	9 306	3.3	13	2.3	35	4.5	3 217	2.7
< 1 year old	156 497	39.1	111 909	39.7	240	41.7	343	44.1	44 588	37.6
≥ 1 year old	21 838	5.5	15 521	5.5	28	4.9	35	4.5	6 317	5.3
<b>Number of BCG vaccination</b>										
Unvaccinated	209 705	52.4	145 345	51.5	295	51.2	364	46.8	64 360	54.3
1	167 458	41.8	119 816	42.5	243	42.2	360	46.3	47 642	40.2
≥ 2	10 877	2.7	7 614	2.7	25	4.3	18	2.3	3 263	2.8
Uncertain vaccination	12 523	3.1	9 306	3.3	13	2.3	35	4.5	3217	2.7

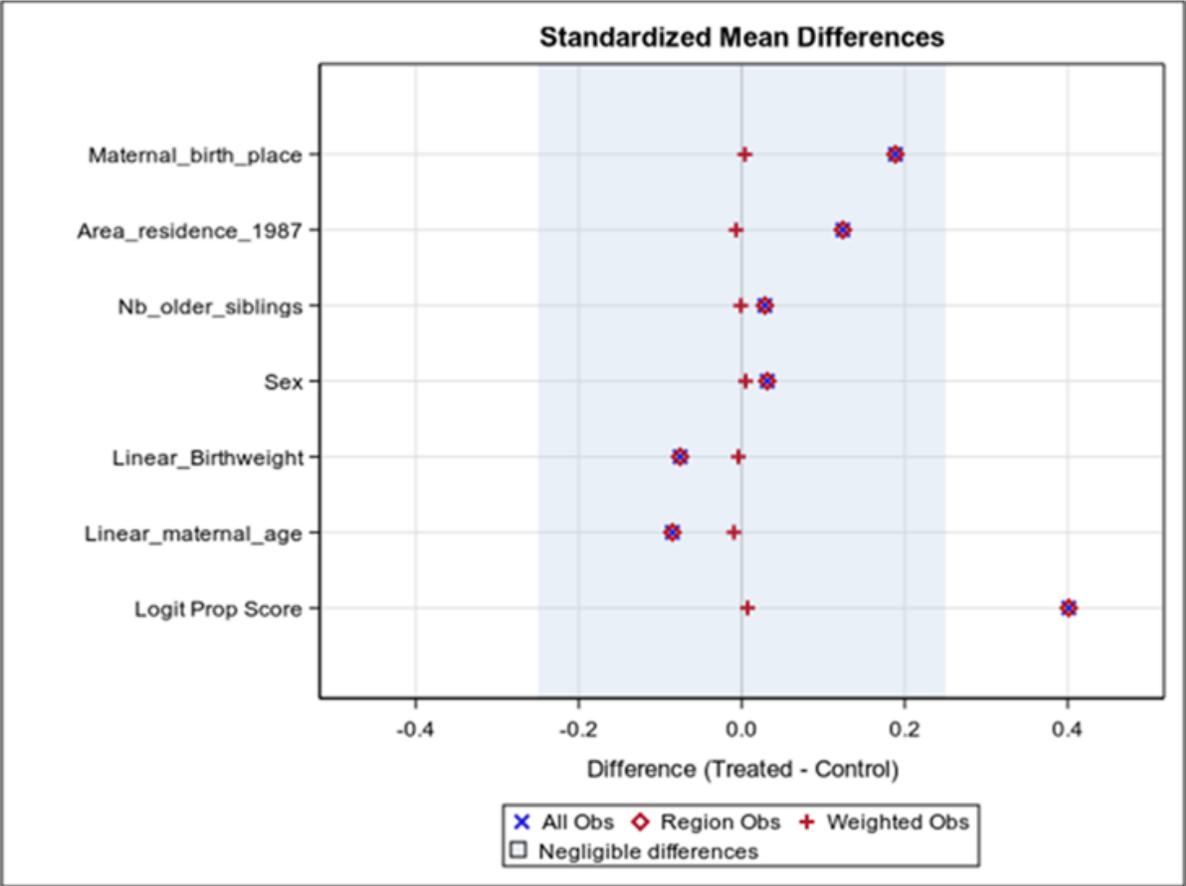
† Determined using the second character of subjects' residential postal code (0: rural, ≥1: urban).

‡ Based on 1991 census.

§ Based on 1996 census.

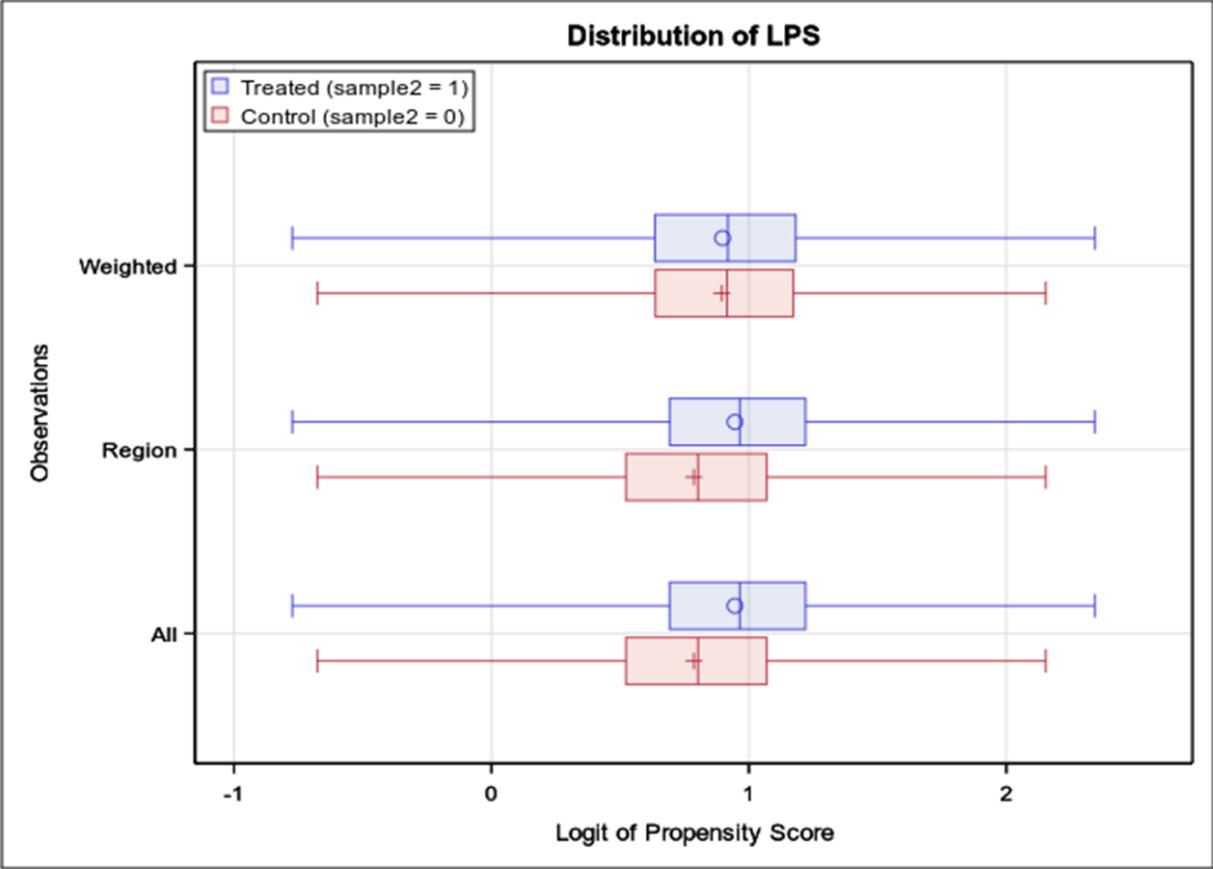
Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du

Québec is not responsible for compilations or interpretation of result



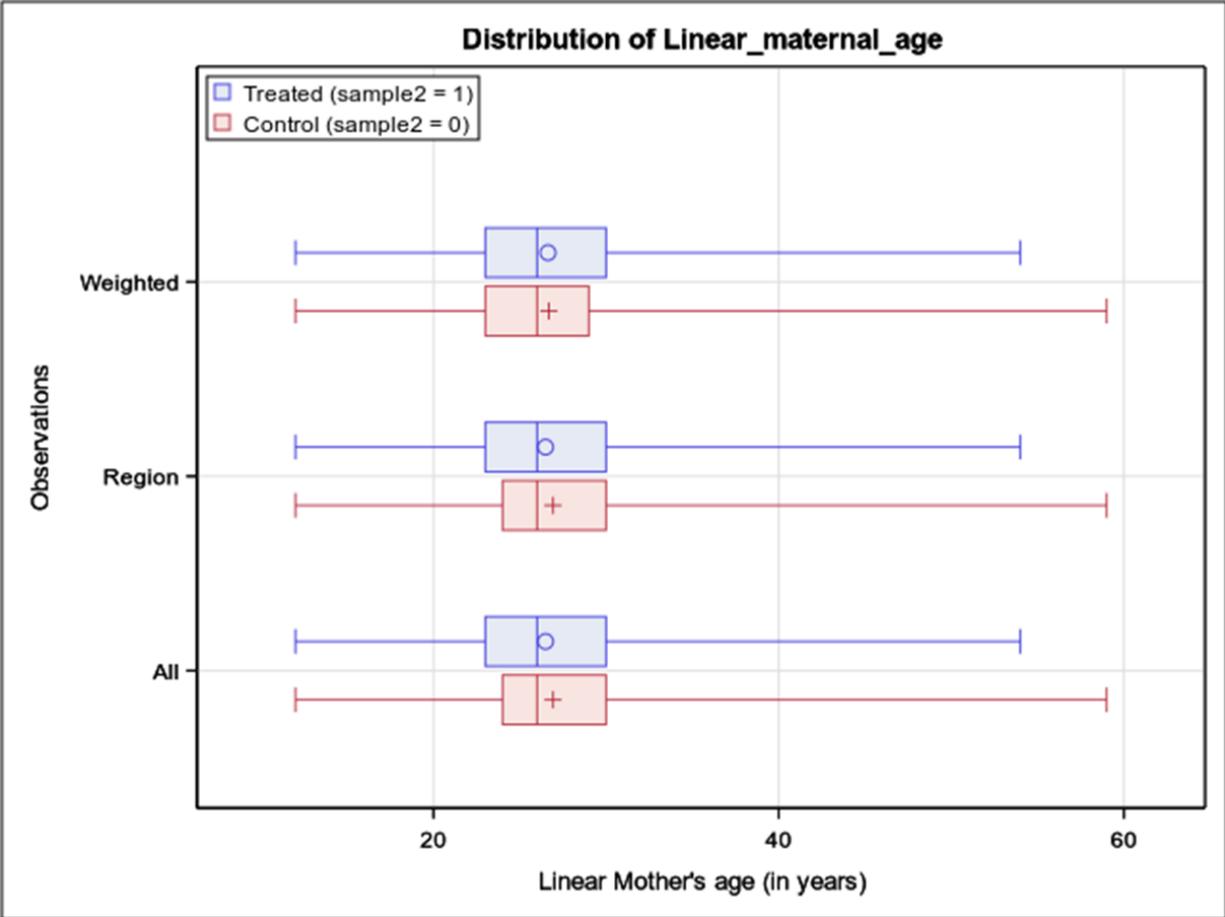
**Figure 6.S2:** Standardized mean difference of continuous and binary variables used in the propensity score to calculate weights in the analytical subset for period 1 (1983 to 1996), from the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



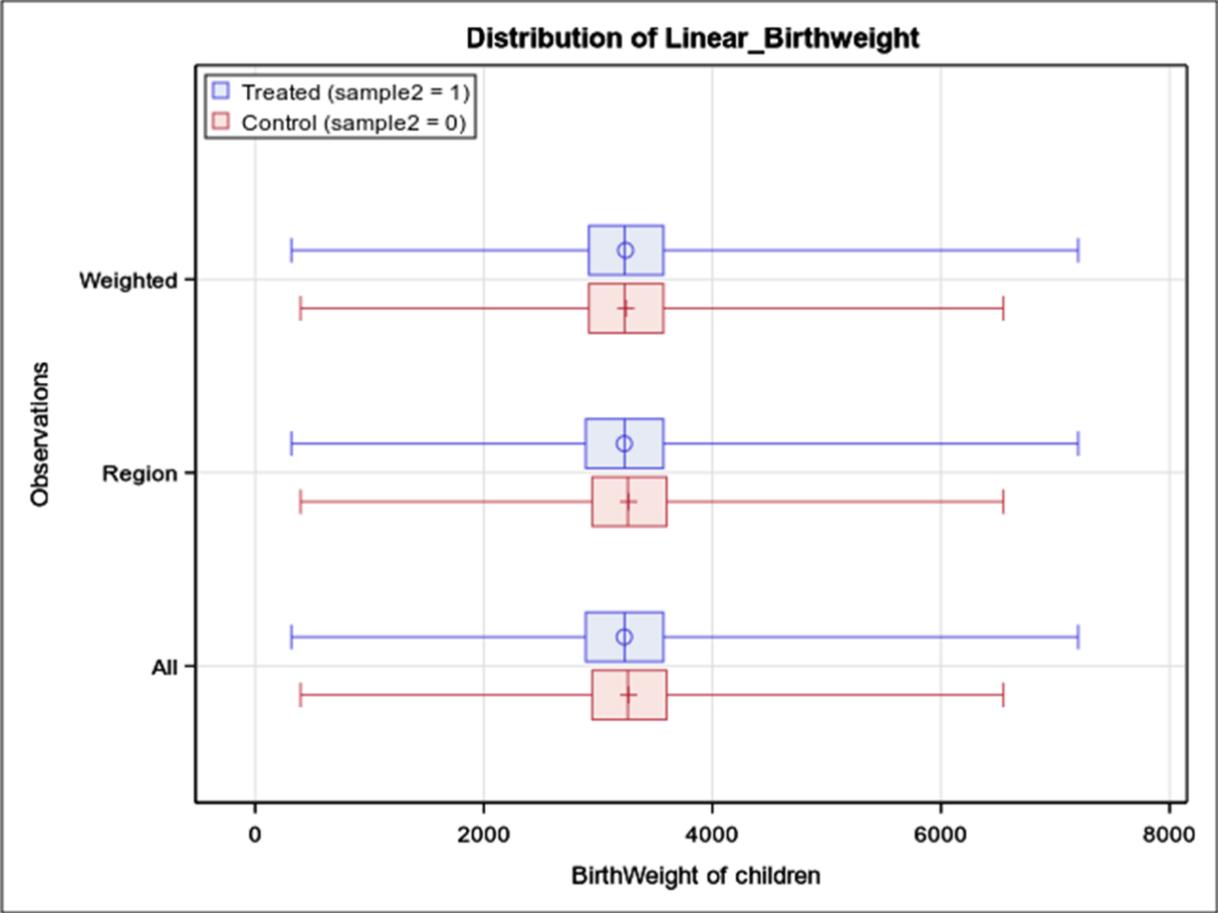
**Figure 6.S3:** Distribution of propensity score logit calculated in the analytical subset for period 1 (1983 to 1996) in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



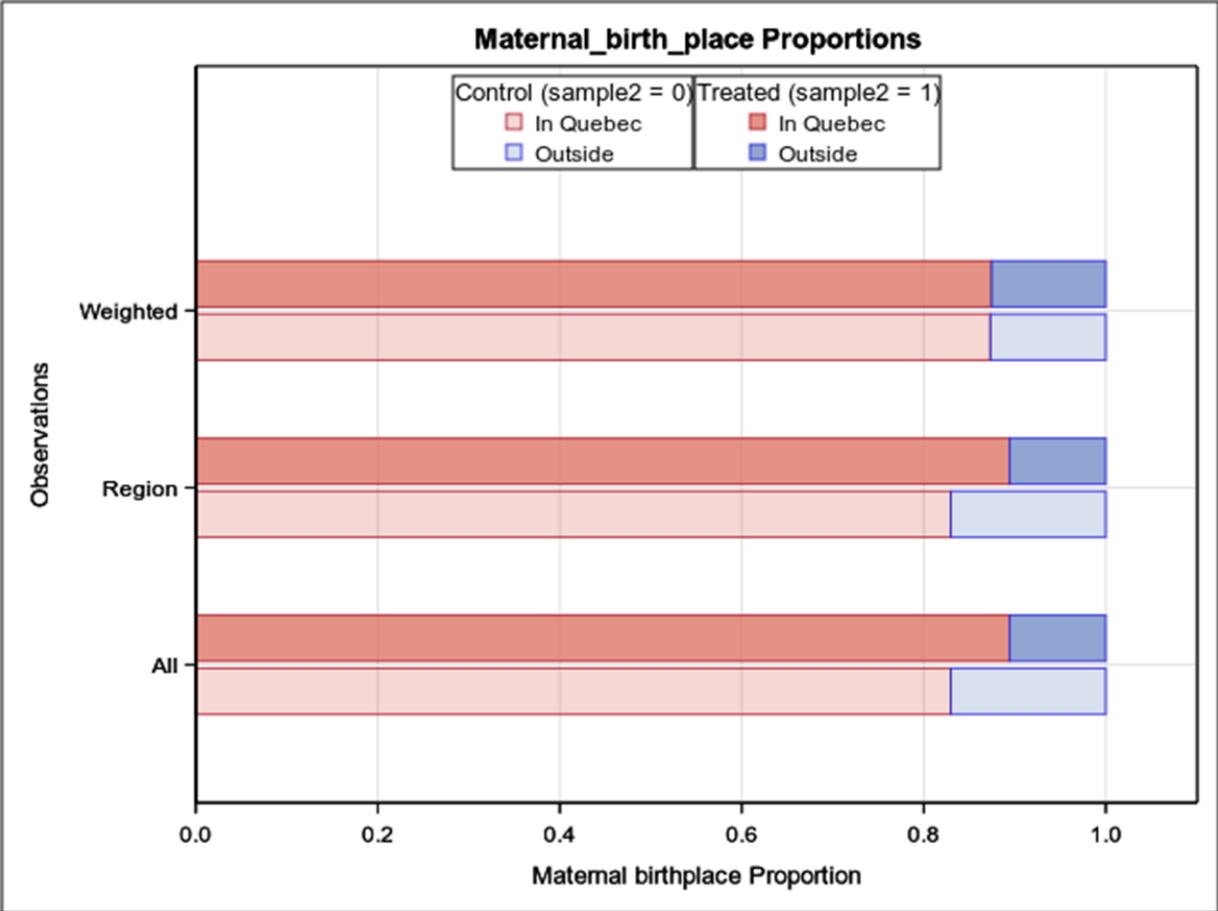
**Figure 6.S4:** Boxplot of maternal age at childbirth (continuous) before (labelled as region) and after (labelled as weighted), applying weights in the analytical subset for period 1 (1983 to 1996) in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



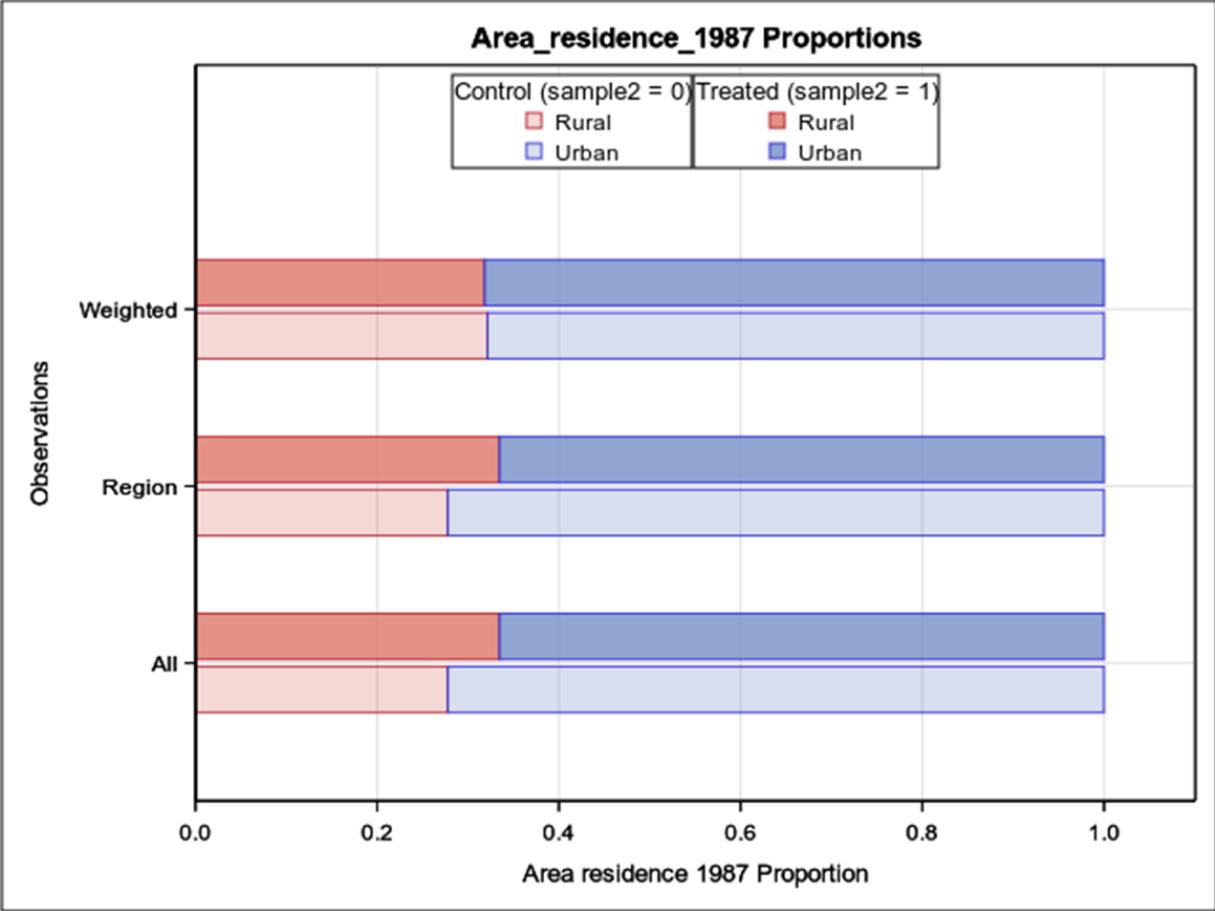
**Figure 6.S5:** Boxplot of birthweight (continuous) before (labelled as region) and after (labelled as weighted), applying weights in the analytical subset for period 1 (1983 to 1996) in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



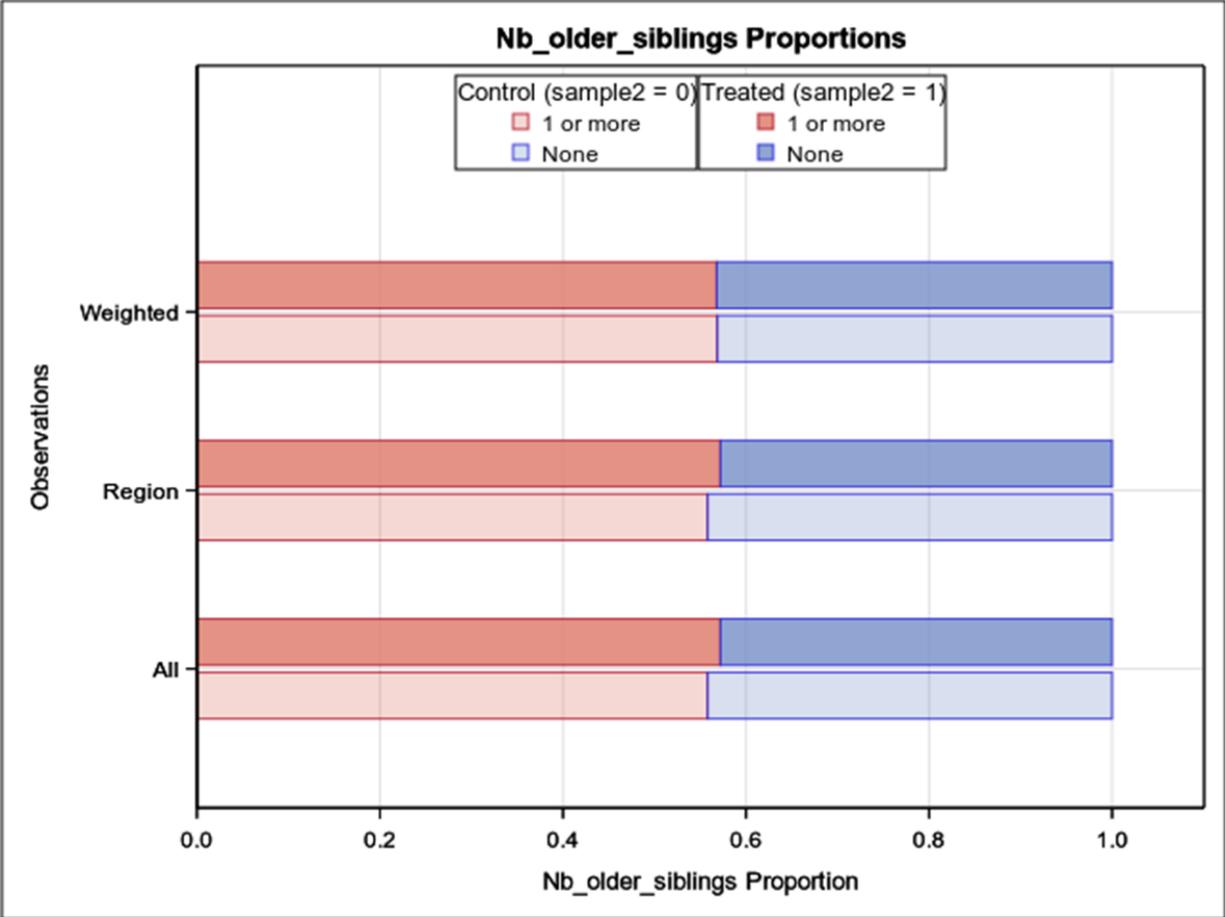
**Figure 6.S6:** Chart of the distribution of maternal birthplace before (labelled as region) and after (labelled as weighted), applying weights in the analytical subset for period 1 (1983 to 1996) in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



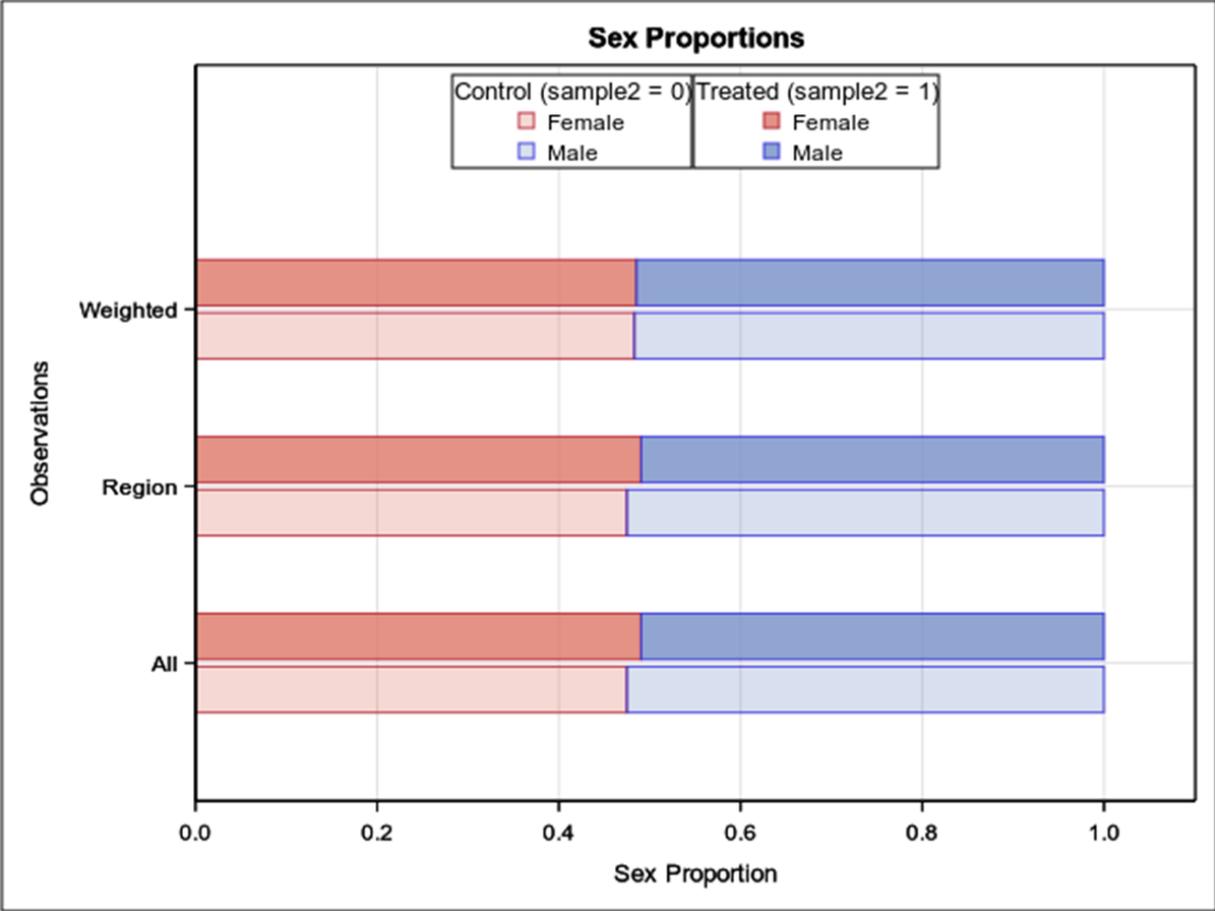
**Figure 6.S7:** Chart of the distribution of area of residence in 1987 before (labelled as region) and after (labelled as weighted), applying weights in the analytical subset for period 1 (1983 to 1996) in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



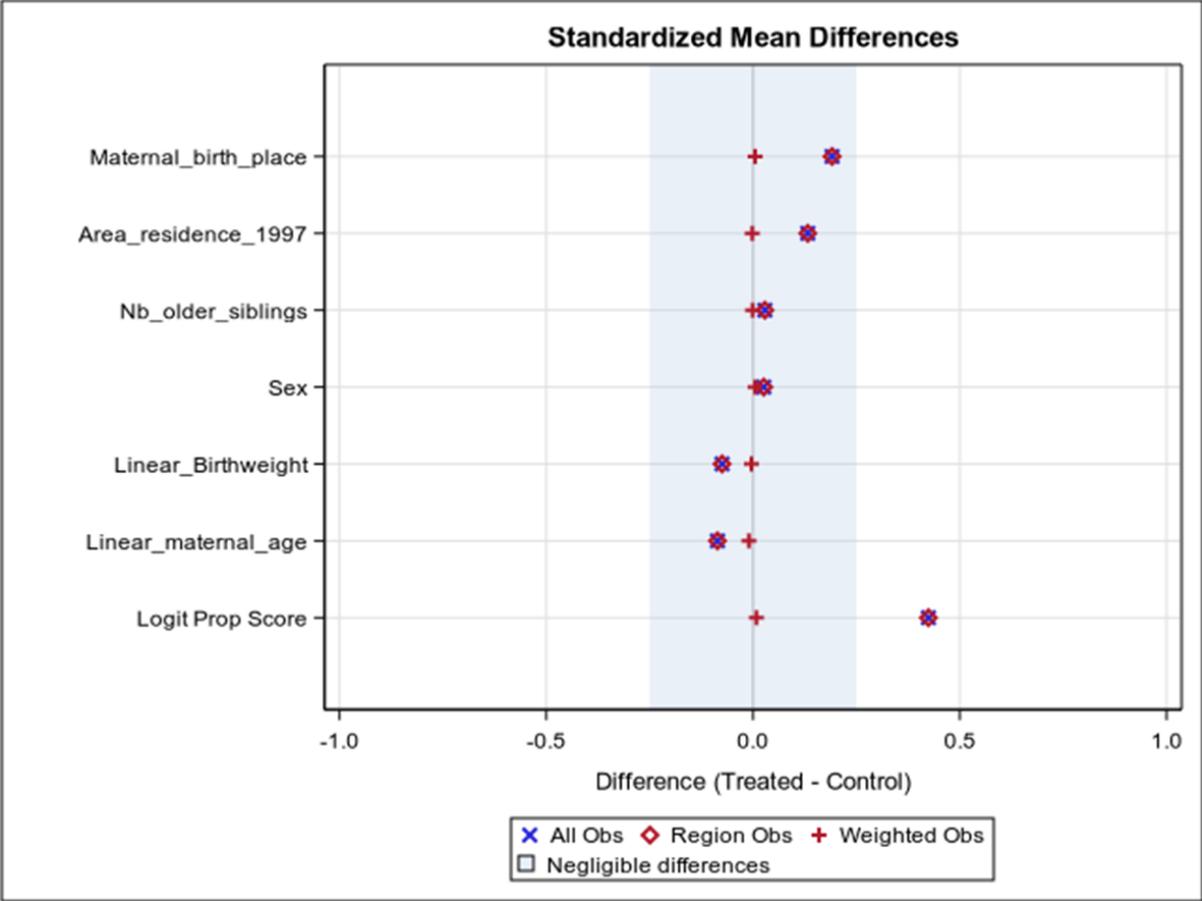
**Figure 6.S8:** Chart of the distribution of number of older siblings proportions before (labelled as region) and after (labelled as weighted), applying weights in the analytical subset for period 1 (1983 to 1996), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



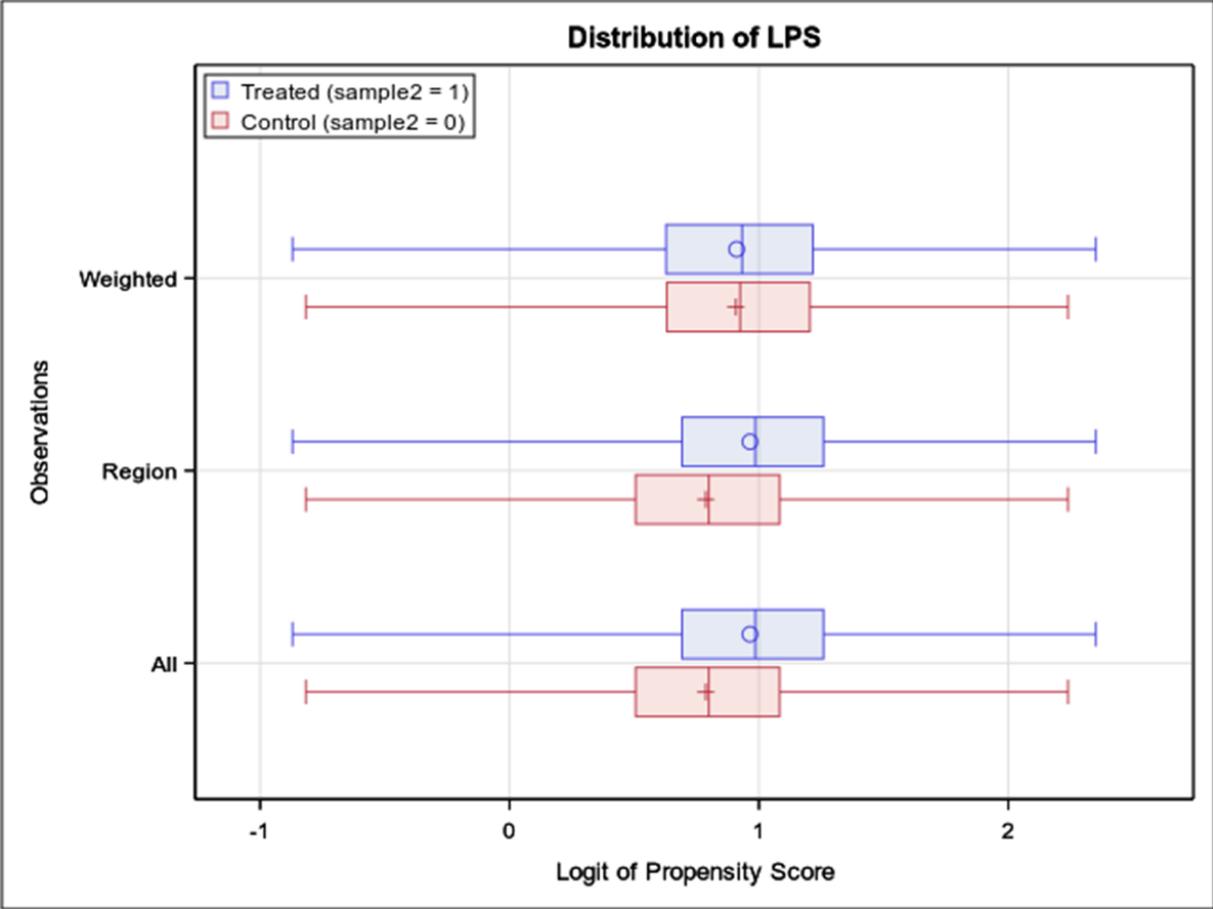
**Figure 6.S9:** Chart of the distribution of sex proportions before (labelled as region) and after (labelled as weighted), applying weights in the analytical subset for period 1 (1983 to 1996), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



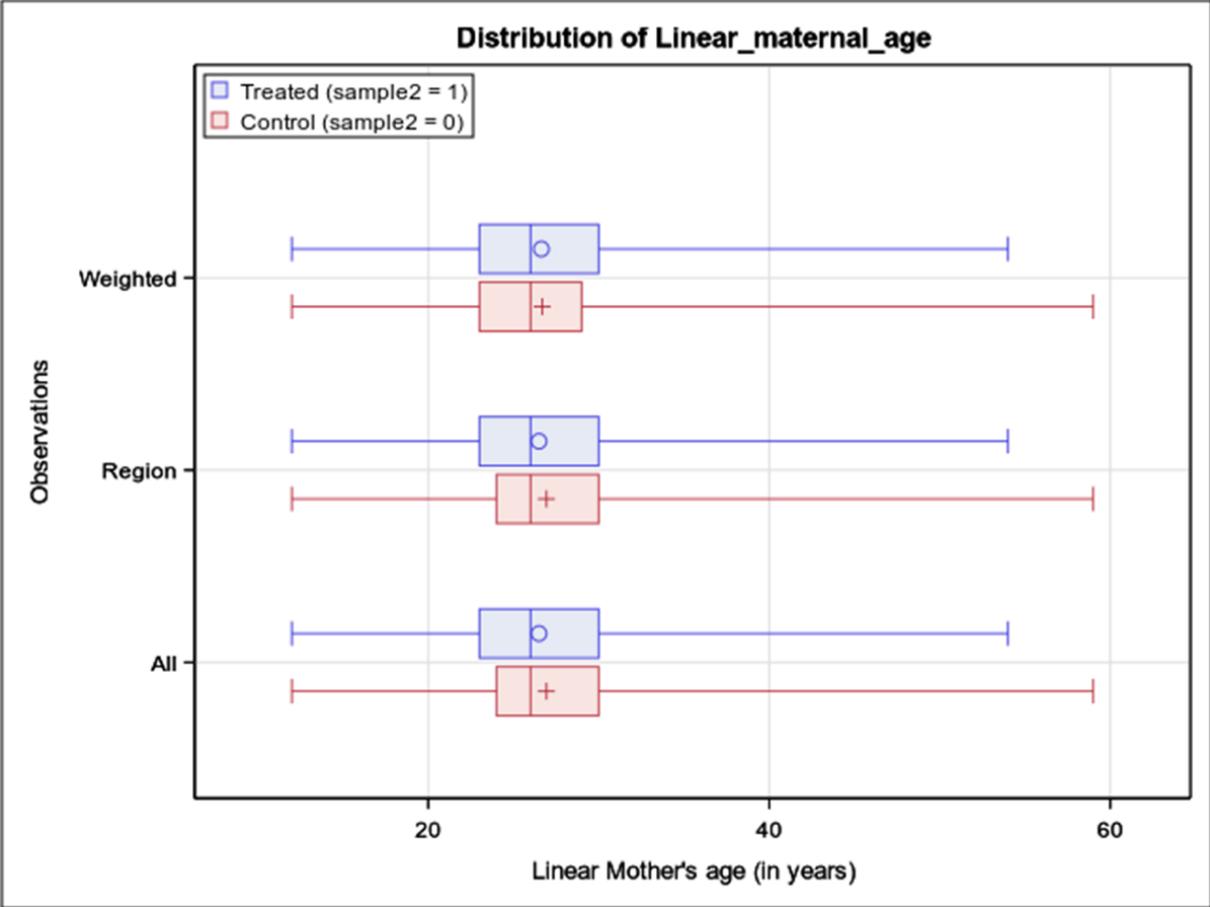
**Figure 6.S10:** Standardized mean difference of continuous and binary variables used in the propensity score to calculate weights in the analytical subset for period 2 (1997 to 2014), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



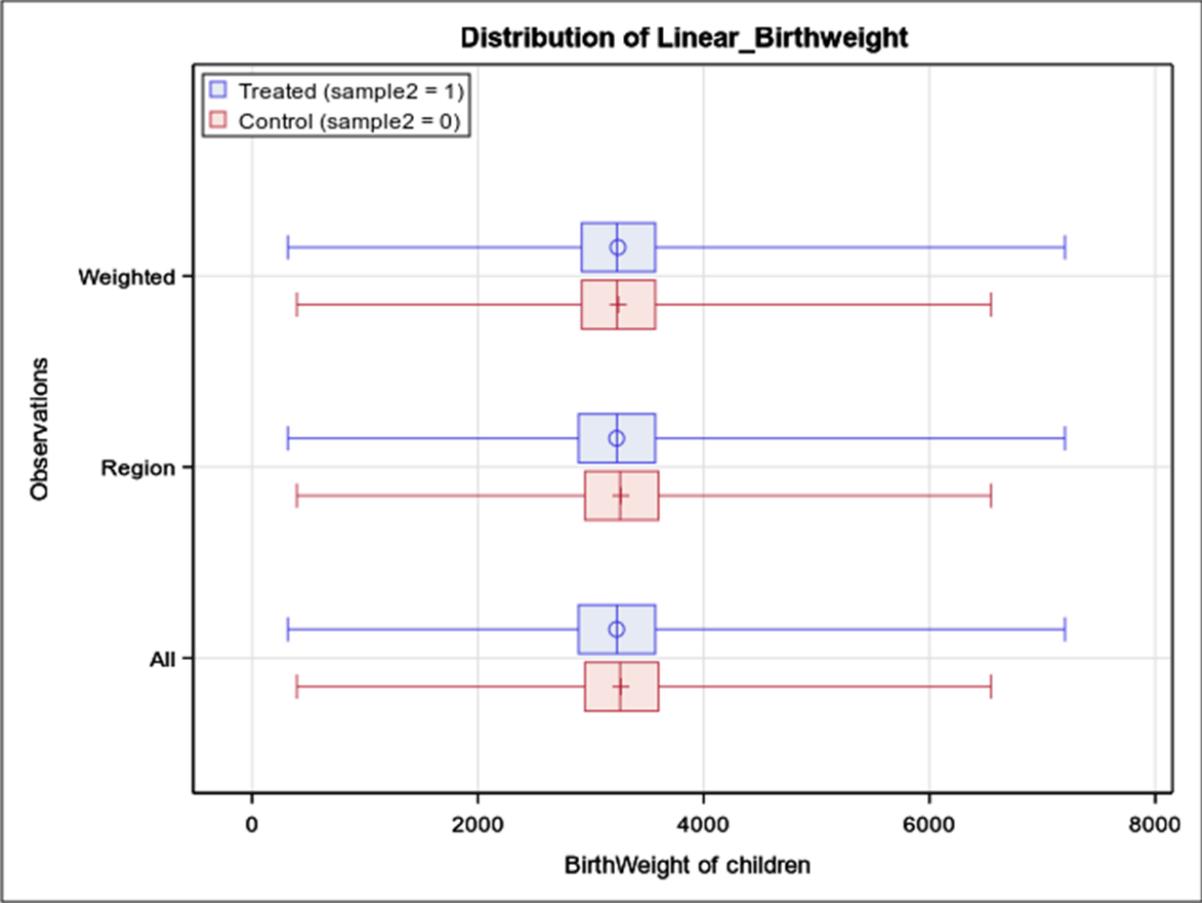
**Figure 6.S11:** Distribution of propensity score logit calculated in the analytical subset for period 2 (1997 to 2014), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



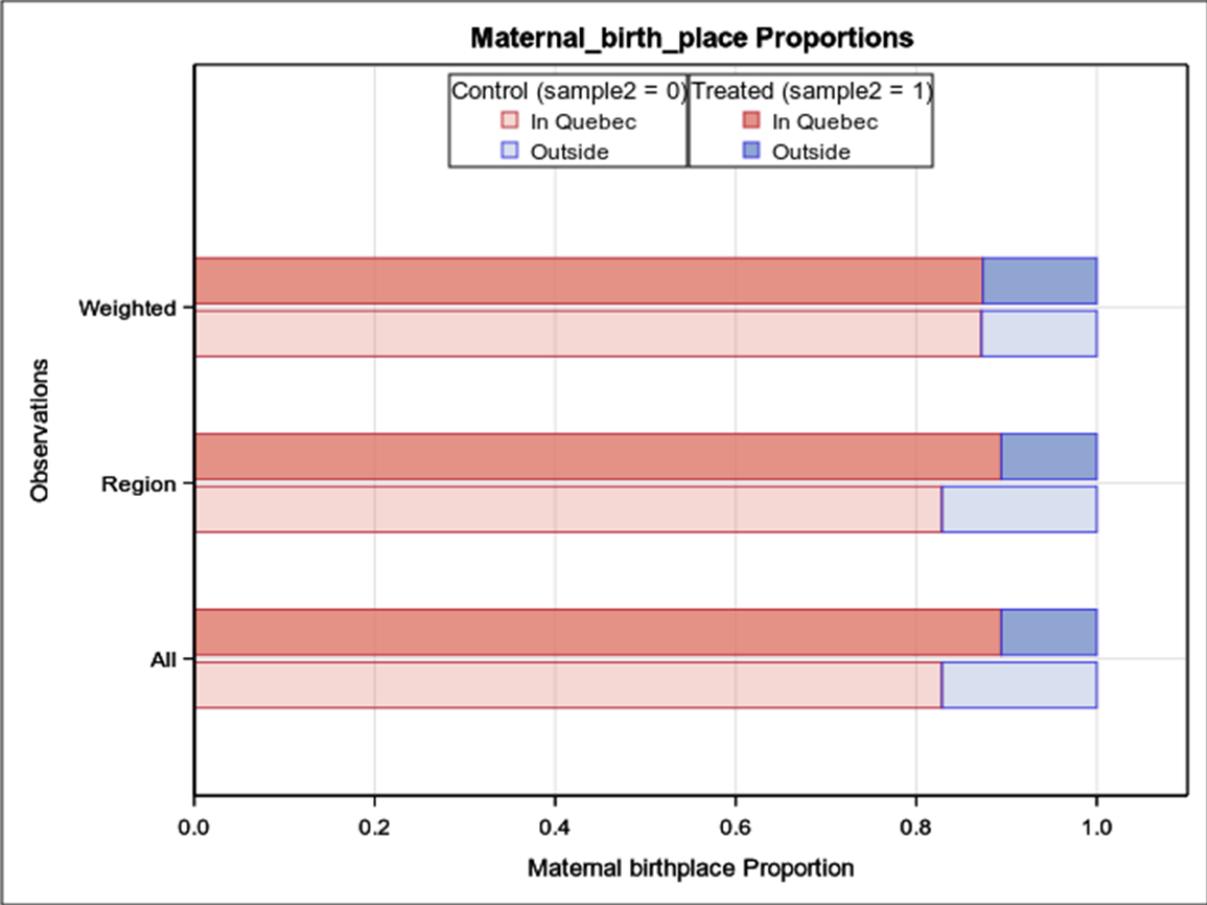
**Figure 6.S12:** Boxplot of maternal age at childbirth (continuous) before (labelled as region) and after (labelled as weighted) applying weights in the analytical subset for period 2 (1997 to 2014), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



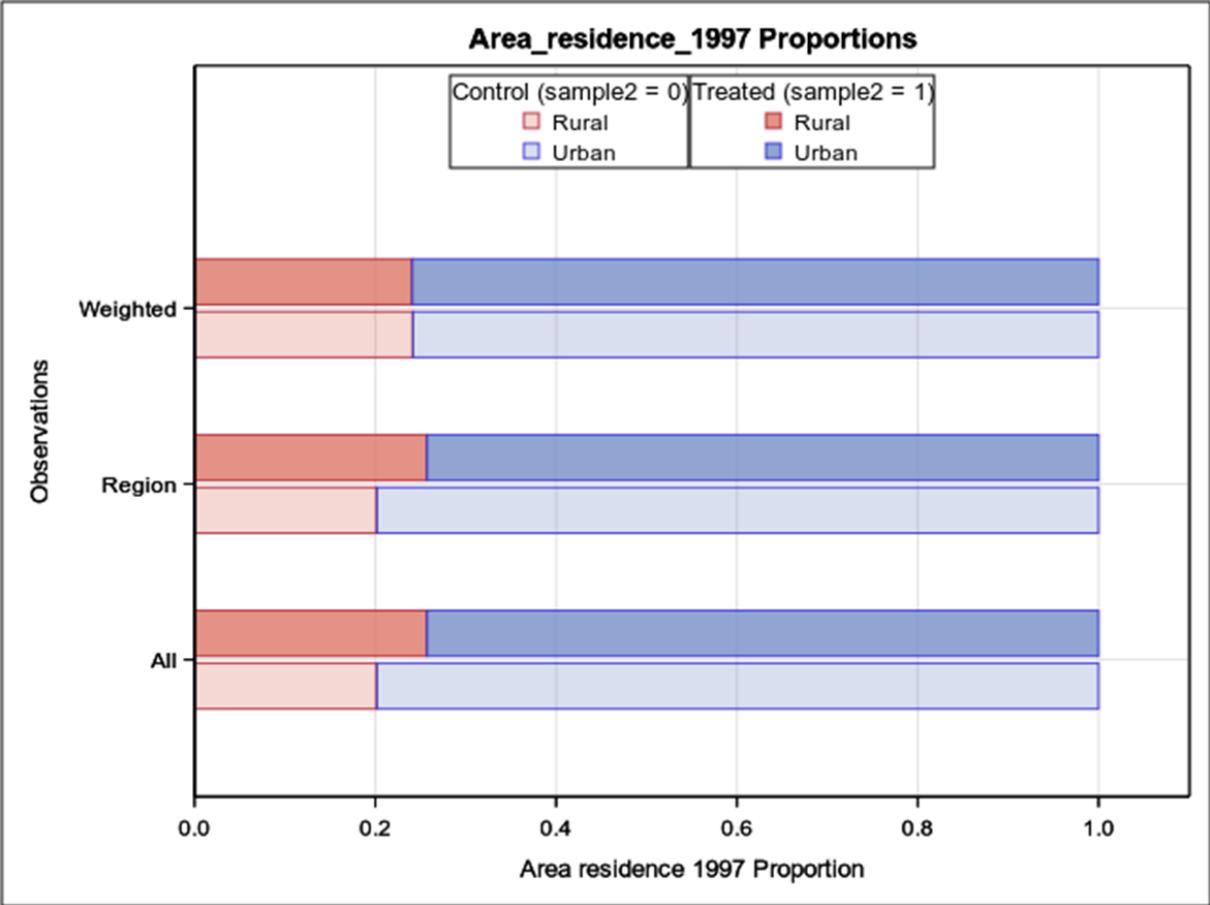
**Figure 6.S13:** Boxplot of birthweight (continuous) before (labelled as region) and after (labelled as weighted) applying weights in the analytical subset for period 2 (1997 to 2014), from the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



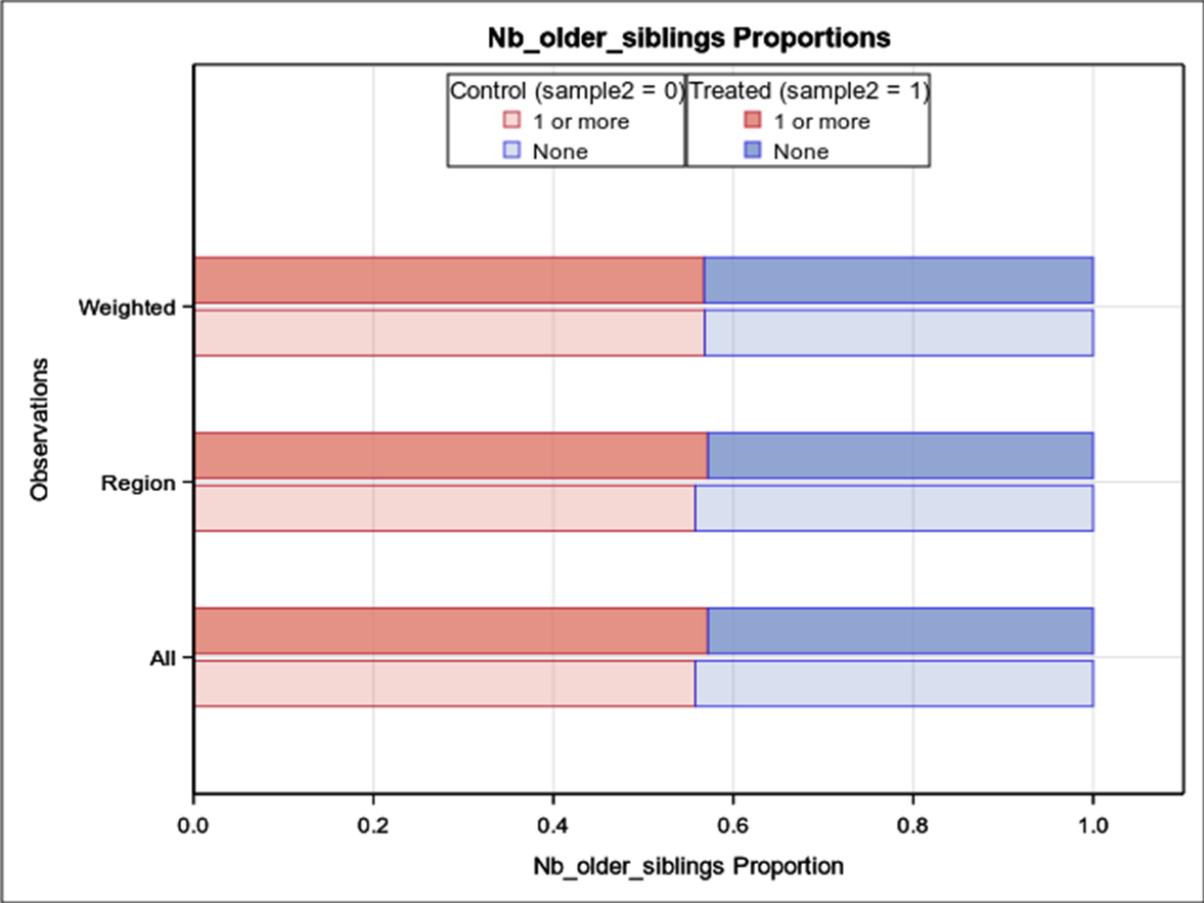
**Figure 6.S14:** Chart of the distribution of maternal birthplace proportions before (labelled as region) and after (labelled as weighted) applying weightings in the analytical subset for period 2 (1997 to 2014), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



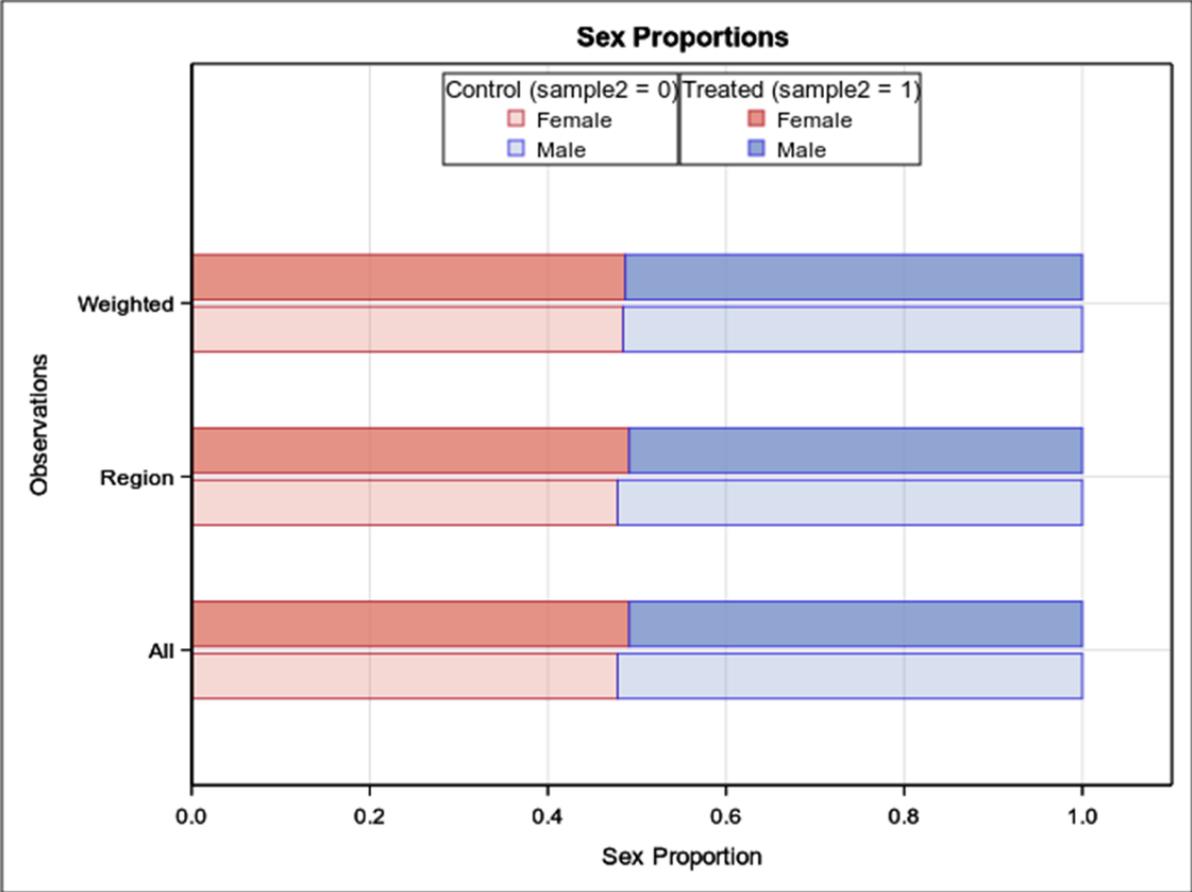
**Figure 6.S15:** Chart of the distribution of area of residence in 1997 proportions before (labelled as region) and after (labelled as weighted) applying weights in the analytical subset for period 2 (1997 to 2014), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



**Figure 6.S16:** Chart of the distribution of number of older siblings proportions before (labelled as region) and after (labelled as weighted) applying weights in the analytical subset for period 2 (1997 to 2014), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



**Figure 6.S17:** Chart of the distribution of sex proportions before (labelled as region) and after (labelled as weighted) applying weights in the analytical subset for period 2 (1997 to 2014), in the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.

**Legend:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

## **7 Article 2. Association between Bacillus Calmette-Guerin vaccination and type 1 diabetes in adolescence: a population-based birth cohort study in Quebec, Canada**

---

### **Vaccination au Bacille Calmette-Guérin et survenue de diabète de type 1 dans l'adolescence : une étude de cohorte de naissance basée sur la population québécoise**

**Auteurs :** Philippe Corsenac<sup>1</sup>, Marie-Élise Parent<sup>1</sup>, Andrea Benedetti<sup>2,3</sup>, Hugues Richard<sup>1</sup>, Simona Stäger<sup>1,4</sup>, Marie-Claude Rousseau<sup>1</sup>.

<sup>1</sup> Unité d'épidémiologie et biostatistiques, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique (INRS), Laval, Canada ;

<sup>2</sup> Respiratory Epidemiology and Clinical Research Unit, Research Institute of the McGill University Health Centre; and

<sup>3</sup> Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, 1020 Avenue des Pins, Montréal, QC H3A 1A2, Canada.

<sup>4</sup> Immunologie et virologie, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique (INRS), Laval, Canada ;

**Titre de la revue :** *Preventive Medicine*, publié le 17 novembre 2021.

DOI : 0.1016/j.ypmed.2021.106893

**Contribution des auteurs :** Philippe Corsenac : Conceptualisation, Méthodologie, Analyses statistiques et structuration des données, Rédaction en version originale. Marie-Élise Parent : Conceptualisation, Validation, Ressources, Révision & Corrections, Supervision. Andrea Benedetti : Conceptualisation, Validation, Révision & Corrections. Hugues Richard : Structuration des Données, Validation, Révision & Corrections. Simona Stäger : Révision et Corrections, Supervision. Marie-Claude Rousseau : Conceptualisation, Validation, Ressources, Révision & Corrections, Supervision, Administration de projet, Acquisition de financement.

## 7.1 Mise en contexte

Sept études ont précédemment été menées pour estimer l'association entre la vaccination au BCG et la survenue du diabète de type 1 dans l'enfance, l'adolescence et le jeune adulte (61-67). Toutes rapportaient une absence d'association mais présentaient des limites méthodologiques importantes. Seulement deux d'entre elles ont étudié l'âge à la vaccination comme exposition principale (61-63). Les différences d'association selon le sexe n'ont été prises en compte que dans trois d'entre elles (62, 63, 65). Enfin ces études se composaient d'échantillons de faible taille, d'un ajustement sur la confusion incomplet (61, 64-67), d'un suivi clinique trop court (66) et d'un statut vaccinal au BCG auto-rapporté dans quelques études (64, 65), dont la validité peut être remise en question (347).

L'article 2, intitulé : *Association between Bacillus Calmette-Guerin vaccination and type 1 diabetes in adolescence: a population-based birth cohort study in Quebec, Canada*, s'intéresse à estimer l'association entre la vaccination au BCG et la survenue de diabète de type 1, dans l'adolescence, entre 10 et 18 ans, en palliant les limites méthodologiques des précédentes études. Par rapport à la dernière étude de notre groupe en 2016 (63), utilisant également la fiabilité d'un registre provincial de vaccination au BCG, la taille de l'échantillon étudié a été quintuplée et les facteurs de confusion ont été précisés par l'élaboration et la documentation d'un diagramme acyclique dirigé. Les algorithmes utilisés pour identifier les cas de diabète paraissent identiques entre celui de Nakhla et al. (70) utilisé ici et celui de Guttmann et al. (315) appliqué dans l'étude antérieure (63). Pourtant, la définition de Nakhla est validée en population québécoise, et précise à 30 jours le délai entre deux visites médicales. Enfin, afin de ne pas introduire de cas prévalents dans la régression proportionnelle de Cox, faite pour modéliser l'incidence, une période de rodage (*run-in*) de 2 ans comparable au temps de notre définition a été appliquée. Cet article pointe vers la nécessité de caractériser au mieux les diabètes de type 1 (par exemple par des données pharmaceutiques sur l'insuline) et d'étudier cette association à un âge différent, ce qui est l'un des objectifs de l'article 3.

## 7.2 Abstract

The Bacillus Calmette-Guerin (BCG) vaccine could reduce the incidence of type 1 diabetes through non-specific immunomodulation. Previous epidemiological studies, presenting some limitations, report no association. We examined this association of early life BCG vaccination and age at vaccination with type 1 diabetes incidence in adolescence in large representative cohort in Quebec.

The cohort included 387,704 individuals born in Quebec between 1970 and 1974 whose BCG vaccination status was determined from a provincial registry. Individuals were followed up from 1985 to their 19<sup>th</sup> birthday (maximum to 1993) for their use of physician services. Individuals were defined as type 1 diabetes cases if they had  $\geq 4$  related physician claims over a 2-year period, with at least 30 days between two claims. Cox proportional hazards regression was used to estimate the association of BCG vaccination and age at vaccination with type 1 diabetes. Covariates were selected based on a directed acyclic graph. Interaction according to sex was evaluated.

A total of 178,133 (45.9%) individuals were vaccinated and 442 (0.11%) incident cases of type 1 diabetes were identified. The risk of type 1 diabetes was similar in vaccinated compared with unvaccinated individuals (adjusted hazard ratio = 1.06 [95% CI: 0.88–1.29]). There was no association with age at vaccination, and results did not differ by sex (Interaction,  $p = 0.60$ ).

Our results suggest that BCG vaccination does not prevent type 1 diabetes in adolescence.

**KEYWORDS:** Adolescence; Bacillus Calmette-Guérin vaccination; Population-based cohort study; Risk factor; Type 1 diabetes.

### HIGHLIGHTS:

- The BCG vaccine could reduce type 1 diabetes (T1D) onset through immunomodulation
- We addressed methodological limitations compared to previous studies
- We used a BCG registry and a new T1D definition validated in the source population
- BCG vaccination/age at vaccination are not associated with T1D onset in adolescence

### 7.3 Introduction

Type 1 diabetes is an autoimmune disease resulting from a paucity of immunoregulatory T cells and the induction of autoreactive T and B lymphocytes. Activation of autoreactive CD4<sup>+</sup> and CD8<sup>+</sup> T cells promotes the production of inflammatory cytokines in the islet microenvironment, which will ultimately cause pancreatic  $\beta$ -cells' death (383). Autoreactive follicular helper T and B lymphocytes (213) also participate to the secretion of autoantibodies. Environmental exposures considered as potential risk factors for autoimmune diseases (e.g., prior infections, lack of vitamin D, UVB radiation) have been studied in relation to type 1 diabetes onset (26, 231, 384, 385). To date, none has been convincingly implicated in the rising incidence of type 1 diabetes, even if the disease is now suspected to result from a complex interplay between environmental and individual factors (26, 231, 384, 385).

The Bacillus Calmette-Guerin (BCG) is a strain of *Mycobacterium bovis* used as a live attenuated vaccine in humans to stimulate immunity against the human strain of tuberculosis. Its non-specific or off-target effects on type 1 diabetes could be threefold. First, recent clinical trials among individuals with type 1 diabetes have shown that the BCG vaccine could improve glycolysis, and thus re-establish normal glycaemia (40-42). Second, BCG has been shown to correct the immune cell imbalances with tumor necrosis factor (TNF) (40). It could favor immunological tolerance (41) through enhancement of regulatory mechanisms (survival of regulatory T and B cells by TNF receptor 2) and counter inflammation of autoimmune cells (apoptosis by TNF receptor 1). Third, the BCG vaccine could "train" innate cells (33, 43) to circumscribe pro-inflammatory pathways of the first innate response, thereby counteracting autoimmune processes and decreasing the risk of type 1 diabetes (40, 296, 386).

No previous epidemiological studies conducted to assess the association between BCG vaccination and type 1 diabetes (61-67) have observed meaningful associations. However, all were likely under-powered and most suffered from methodological limitations including the lack of adjustment for potential confounders (61, 64-67), short follow-up (66), and self-reported BCG vaccination status (64, 65).

Furthermore, BCG vaccination could have different effects according to sex (46) and age at vaccination. Vaccination during the first year of life may have a greater impact (47, 48), but few studies have evaluated it (61-63).

To address these limitations, we assessed the association between BCG vaccination (or age at vaccination) and type 1 diabetes using a large birth cohort from the province of Quebec, Canada.

## **7.4 Methods**

### ***Study design and data sources***

The study used the Quebec Birth Cohort for Immunity and Health (QBCIH), previously described (303), which was established by probabilistic linkage of the birth registry and the Healthcare Registration file using surname, first name, sex, date of birth, and father's first name as record identifiers between datasets. Probabilistic linkage evaluates the similarity between selected identifier variables in two datasets. Tolerating some differences (misspelling, errors), it estimates a probability that two records belong to the same person. An 'overall match weight' is assigned to every pair of records based on agreement and data frequency (e.g., higher weight if agreement for a rare surname vs. a very common surname). Pairs of records are then rejected if the overall match weight is below a lower threshold, accepted if above an upper threshold, and treated manually if in-between (305). Linkage was done with the Generalized Record Linkage System software and the G-Link software version 2.4, both developed by Statistics Canada (304).

Probabilistic linkage was used to add data from the death registry (covering 1970-2014) and the Quebec BCG vaccination registry. Data from the physician claims were also imported using a unique identifier available in these datasets and in the Healthcare Registration File.

All physician claims have diagnoses coded using the International Classification of Disease (ICD) 9<sup>th</sup> revision (diabetes mellitus: 250.X).

### ***Study population***

Eligible subjects were born in Quebec between 1970 and 1974 after at least 32 weeks of gestation, documented in the birth registry, and covered by the public health system (Healthcare Registration File). The cohort initially included 400,611 individuals, representing 90.4% of eligible persons for whom the initial probabilistic linkage was successful (Figure 7.1, Flowchart).

Physician claims were available from January 1, 1983 to December 31, 2014. A run-in period of two years was applied to exclude prevalent cases identified in 1983 and 1984. Onset of type 1 diabetes could indeed conceivably be before 1983, and there was an accumulation of these cases

in 1983 and 1984. We thus followed individuals from January 1, 1985 to their 19<sup>th</sup> birthday (maximum to 1993) for their use of medical services. The upper age limit aimed at restricting to type 1 diabetes cases only.

### ***Outcome definition***

Different definitions have been used in the past for capturing type 1 diabetes cases from administrative health data in Canada, with good sensitivity (83-95.7%) and high specificity (99-100%) (70, 315). The superior ones were 1)  $\geq 4$  physician claims or 1 hospitalization over a 1-year period (sensitivity 91.2%, specificity 100%, PPV 93.5%) and 2)  $\geq 4$  physician claims over a 2-year period (sensitivity 90.4%, specificity 100%, PPV 93.2%) (70).

We selected the second definition because hospitalization data were only available from 1987, and thus were not used in the present study. We also applied the new Nakhla's recommendations by considering only physician claims separated by at least 30 days. Based on our definition and the upper age limit, the vast majority of incident DM cases were expected to be type 1 diabetes cases. The date of the first physician visit was considered as the date of diagnosis.

### ***Exposure definition***

An organized voluntary program of vaccination with BCG was held in the province of Quebec, Canada, from 1949 to the mid-1970s (279). The program targeted newborns and schoolchildren up to the 11<sup>th</sup> grade, and all persons at risk of tuberculosis (282). Preliminary intradermal tuberculin test was done for all except newborns, and vaccination was offered free of charge (282). Children whose parents were born outside of Quebec and those living in urban areas were less likely to be vaccinated (307), although it is not clear whether this reflects lower vaccine acceptance among the parents or among the health professionals tending to them. All vaccination certificates, filled by dedicated nursing staff, were sent to the centralized Quebec BCG vaccination registry, which is now computerized. The completeness and accuracy of the computerized registry were previously assessed (279, 282). Vaccinated and unvaccinated individuals were determined according to this registry. Vaccination status was first defined as a binary variable (yes / no) and then as a categorical variable according to age at BCG vaccination (not vaccinated / vaccinated < 1 year old / vaccinated  $\geq 1$  year old).

### *Statistical analysis*

Univariable and multivariable Cox proportional hazards regression analyses were performed with age as the time scale to assess the association between BCG vaccination and the onset of type 1 diabetes.

Data were left-truncated at 1985 and follow-up was right-censored on the day before their 19<sup>th</sup> birthday (maximum to 1993). Subjects with a date of diagnosis in 1983 or 1984 were considered as prevalent cases and excluded from the analyses (Figure 7.1). Follow-up ended at diagnosis of type 1 diabetes, at death, or on the day before their 19<sup>th</sup> birthday whichever came first. The proportionality assumption was ascertained by the Schoenfeld standardized residuals.

Since sex-differential non-specific effects of the BCG vaccine have already been observed, the potential differential effect of BCG vaccination by sex was assessed by introducing an interaction term (BCG vaccination status\*sex) in the univariable model (46).

Potential confounders were selected according to a directed acyclic graph (DAG). Table 7.A and figure 7.B in the supplementary file provide additional information and literature on which the DAG was based. Maternal and paternal birthplace were used as a proxy for ethnicity. Parents who were born in Quebec around the 1940s and 1950s were assumed to be mostly Caucasians from French or British ancestries, which represented 94% of the province's population (332). Before 1960, immigration in Quebec was limited and came mostly from Europe. Among the other ethnic groups were the First Nations, Inuit and Metis (~6%), Asians (3%) and Africans (~1%), as well as non-declared origin (4.6%) (332). Parents who were born outside of Quebec were considered as possibly having other ethnocultural origins. Also included as covariates were parents' age at childbirth (continuous), duration of gestation, birthweight for gestational age (using Kramer's Canadian-based reference) (306), having older siblings, and rural/urban area of residence in 1987 (the closest available time to BCG vaccination). There were 25,757 (6.6%) subjects with missing data on at least one of the variables included in the multivariable models. Multiple imputation by chained equations was thus applied (n=7 data sets) on confounders, using the distributions of other informative variables included in the dataset.

The functional forms of the continuous covariates, parents' age at childbirth and duration of gestation were studied as linear, categorical, square, log and restricted cubic spline variables to determine the best model fit with the Stanford University SAS macro (%SPECI).

All analyses were performed with SAS software version 9.4 and Rstudio.

### ***Analytical sample***

The study sample included 387,704 individuals, after exclusions and left-truncation on January 1<sup>st</sup> 1985 (Figure 7.1).

### ***Ethics approval and access to data***

The study was approved by the ethics committees of *Institut national de la recherche scientifique, Institut de la statistique du Québec and Régie de l'assurance maladie du Québec*, and access to data was granted by the *Commission d'accès à l'information du Québec*.

## **7.5 Results**

The male to female ratio was 1.06. Approximately two-thirds of the cohort members were born at an appropriate birthweight for gestational age, and lived in an urban area, while just over half had older siblings. Mothers were on average 27 years old at childbirth, whereas fathers were 30 years of age. The mean duration of gestation was 39.2 weeks [standard deviation (SD)= 1.7]. Most (82%) mothers of cohort members continued their pregnancy to term. Parents of the cohort members were mostly born in Quebec (85% of mothers and 82% of fathers) (Table 7.1).

Among subjects who received the BCG vaccine (n=178,133, 45.9% of the analytical sample), the vast majority were vaccinated under 1 year of age (87.8%) and received the vaccine only once (93.9%) (Table 7.1). A total of 442 cases were defined as incident type 1 diabetes cases, including 208 who were BCG vaccinated and 234 unvaccinated (Figure 7.1).

The mean age at entry in 1985, after a 2-year run-in period, was  $12.5 \pm 1.5$  years old (range: 10 to 14 years old).

Among various functional forms of duration of gestation and parents' age at childbirth, the linear representations were selected based on the Akaike Information Criterion (lowest value indicating best fit).

### ***Diabetes prevalence and incidence***

The overall prevalence of type 1 diabetes between 1985 and 1993 using our chosen algorithm was 0.114%, 95% CI: 0.113-0.115%. The incidence rate of type 1 diabetes between 1985 and 1993 was 17.6 cases per 100,000 person-years, 95% CI:16.0-19.2, within our study population.

### ***The proportionality assumption***

The multivariable regression models suggest that hazards for exposures and covariates were globally proportional over time (P= between 0.21 and 0.33 for exposures through the 7 imputed datasets).

### ***Association between BCG vaccination and type 1 diabetes***

There was no association between BCG vaccination and type 1 diabetes (Crude HR = 1.06, 95% CI: 0.88–1.27; adjHR= 1.06, 95% CI: 0.88–1.29; Table 7.2). Findings did not differ by sex (P-value for interaction term between sex and BCG vaccination status = 0.60). Similar results were observed when using age at BCG vaccination as the exposure (Table 7.2).

## **7.6 Discussion**

We observed no association between BCG vaccination and the onset of type 1 diabetes in a pediatric population aged 10 to 18 years. Age at vaccination was also not associated with risk of developing the disease. Further, there was no effect modification by sex.

These results are based on a population-based birth cohort using a registry-defined BCG vaccination status, and applying a validated definition for identifying type 1 diabetes cases from administrative health data in Quebec (70). The overall prevalence of type 1 diabetes within our cohort was 0.11%, 95% CI: 0.11-0.12%, between 1985 and 1993 (age 12-18). Nakhla et al. (70) found a prevalence of 0.15%, 95% CI 0.15–0.16%, in a pediatric population aged 0-15 years for 2002-2011, whereas the true prevalence of diabetes in their reference standard population was 0.17%. The time periods and age of subjects were different in the two studies, but the prevalence was relatively similar. The incidence rate in our study population was 17.6 cases per 100,000 person-years, 95% CI:16.0-19.2, between 1985 and 1993 (12-18 years of age). According to Legault et al. (68), the age- and sex- standardized incidence rate was around 15 per 100,000 (with a peak incidence in the 13 year-olds at 25 per 100,000), between 1989 and 2000 among individuals

10-18 years in Quebec. These comparisons tend to support the representativeness of our study population.

Previous epidemiological studies on the association between BCG vaccination and type 1 diabetes (61-67), presenting some limitations, generated the same conclusion as ours. Of seven studies, three were cohort (61, 63, 66) and four were case-control studies (62, 64, 65, 67). Overall, the associations ranged from 0.83 to 1.30, and none of the individual studies generated a risk estimate different from the null value.

In these studies, exposure assessment was based either on self-reported vaccination status (65), vaccination records (61, 66, 67), vaccination registry (62, 63), or a combination of both self-reported information and vaccination records (64). Blom et al. described that around 20% of families provided self-reported vaccination (67). However, they only used vaccinations records which were obtained from the local child health care centres and school health units in their analyses (67). By contrast, the EURODIAB study used around 15 to 20% of vaccination self-reports (from control and cases, respectively) as exposure information (64). Wadsworth et al. used the self-reported vaccination status from questionnaires (65). Self-reports have been shown to have poor validity (347) for determining vaccination status, and thus can lead to misclassification bias. Nevertheless, most previous studies used reliable sources such as vaccination records or registries for exposure assessment. Similarly to a previous study conducted by our team (63), the present study used a provincial BCG vaccination registry to determine exposure.

For outcome assessment, previous studies determined the onset of type 1 diabetes biologically through the development of islet autoimmunity (61), clinically through population-based registers of childhood onset diabetes (62, 64, 67) or pediatrician diagnosis (65), or both biologically and clinically (66). Although clinical data represent the gold standard for identifying the outcome, the development of islet autoimmunity does not indicate the real beginning of the disease in patients. Biological or clinical data were not available in our previous (63) and present studies. Nonetheless, the present definition used for identifying type 1 diabetes cases from physician claims was recently validated specifically in the Quebec population (70). Specificity (100%), sensitivity (90.4%), as well as the positive predictive value (93.2%) were very high, thereby limiting false-positives and capturing true cases of type 1 diabetes (70). Strictly speaking, the Nakhla definition (70) is the same as the previous Guttman definition (315) developed in and for the Ontarian population that

our group previously used (63):  $\geq 4$  physician claims within 2 years. However, Nakhla et al. introduced a stricter data cleaning procedure, considering only services that were 30 days apart. Both definitions reached nevertheless a high degree of reliability compared to clinical data.

The previous study with the greatest statistical power was EURODIAB (n=3,202) which included 900 cases and 2,302 controls (64). In EURODIAB, no association was reported between BCG vaccination and type 1 diabetes (OR =0.83; 95% CI=0.57-1.20) (64). Other studies had lower statistical power since the number of diabetes mellitus cases reached a maximum of 339 (61-63, 65-67). In an earlier study conducted by our group and focusing on BCG vaccination in the first year of life using the Guttman definition, no run-in period, and a follow-up until age 20, we identified 230 persons with diabetes mellitus in a study population of 78,492 subjects (63). The present study allowed for the identification of 442 cases, making it the second largest. Applying a diabetes mellitus definition recently validated in the Quebec population that considered physician claims at least 30 days apart, our results were nevertheless consistent with a null risk estimate and concordant with our previous study in which the definition counted claims on consecutive days (HR= 1.04, 95% CI 0.80, 1.37) (63).

The vast majority of vaccinated subjects (82.3%) in our study received a unique dose of BCG vaccine during the first year of life. Lasting effects of BCG vaccination on innate trained immunity and on Th1/Th17 immune responses have been demonstrated among BCG-vaccinated adults, up to 1 year after vaccination upon stimulation with non-mycobacterial pathogens (387). However, the off-target effects of a unique dose of BCG vaccine after 5, 10 or 15 years are not well known. If there was a short-term preventive effect among young children, our study would not have had the ability to detect it given the lack of data on incidence of type 1 diabetes before 1985. The long-lasting effect of the BCG vaccine on innate and adaptive immune memories against unrelated pathogens and autoimmunity has been partially addressed (40), but remains to be further investigated. Furthermore, the effect of the BCG vaccine on the heterologous Th1 (IFN- $\gamma$ ) and Th17 (IL-17 and IL-22) cytokine production could be interpreted as potentially leading to an increased risk of type 1 diabetes (299). There used to be a prevailing view that treatment or prevention of type 1 diabetes called for immunosuppression (40). It has now been shown that TNF stimulation, rather than immunosuppression, actually favors the immunological tolerance required for treatment of autoimmune diseases (40).

### ***Limitations***

Our study presents some limitations. The lack of administrative health data until 1983 and run-in period of 2 years (1983-1984) resulted in the follow-up extending from 12 to 18 years. In this age range, we cannot exclude that diabetes cases might have included a small proportion of type 2 diabetes and of maturity-onset diabetes of the young, both occurring in adolescence. In our follow-up period (1985-1993), type 2 diabetes cases were rare (63, 388). Type 1 diabetes occurs predominantly in childhood before age 19 and the incidence peaks at 10-14 years to stabilize after 20 years (12, 69). As such, early onset cases could not be identified, limiting type 1 diabetes cases (n=442) and our conclusions to incident type 1 diabetes from ages 12 to 18 inclusively. We indeed excluded 335 type 1 diabetes cases identified in between 1983 and 1984 for two reasons: First, prevalent and incident cases could not be distinguished from one another. Second, the true date of diagnosis for prevalent type 1 diabetes with onset between birth and 1983 was unknown.

### ***Strengths***

The present study is based on the QBCIH (303), but the cohort was expanded since our previous publication (63). This new version of QBCIH gathered individuals who were born in Quebec, not only in 1974, but between 1970 and 1974. The sample size was greatly increased and the study population is representative of the population born in the early 1970s in Quebec.

Most previous studies did not consider an exhaustive list of potential confounders (61, 62, 65-67). EURODIAB and our previous study (63, 64) adjusted for a somewhat different set of confounders than those used in the present study. The current models are not adjusted for three variables included in our previous work (sex, census-based family income, and presence of concomitant allergies) (63), but take into account two new variables (father's age at childbirth and duration of gestation). We indeed used a DAG (37) to formalize the selection of potential confounders. To date, except for the perinatal factors included in our DAG, no environmental or lifestyle factors have been implicated in the rising incidence of type 1 diabetes (26, 231, 384, 385). Nevertheless, a recent meta-analysis suggested that maternal smoking during pregnancy could be inversely related with T1D onset in children (243). Whether this association also exists with disease onset later in adolescence, the focus of the present investigation, is unknown. Moreover, information on maternal smoking was available for less than 0.4% of the 387,704 cohort members included in the present study, precluding a meaningful adjustment for this factor. Despite a formal process to build

our DAG, we observed no difference in estimates between crude and adjusted hazard ratios, suggesting that the factors available to us did not actually confound the associations.

Up to 90% of type 1 diabetes patients carry the MHC-II DR4-DQ8 or DR3-DQ2 haplotype or both, compared to 2% of the general population (25). Nevertheless, these MHC-class-II haplotypes are not confounders for the association between BCG vaccination and type 1 diabetes onset. Even if MHC-II DR4-DQ8 or DR3-DQ2 and ethnicity are intrinsically linked, MHC-II haplotypes cannot summarize all components of ethnicity. We thus adjusted for aspects of ethnicity other than MHC-class-II haplotypes by using the parents' birthplace as a crude proxy. Our final multivariable model thus considered the variability of onset of type 1 diabetes between Quebec and immigrant populations, as well as the differential likelihood of BCG vaccination between children with Quebec- and non-Quebec-born parents.

## **7.7 Conclusion**

Within a large birth cohort, we observed no association between either BCG vaccination or age at vaccination and the incidence of type 1 diabetes among children of both sexes aged 12 to 18 years. This suggests that BCG vaccination does not prevent type 1 diabetes onset in adolescence.

A potential protective effect of BCG on type 1 diabetes in early childhood warrants further investigation within other cohorts or populations with an available diabetes registry.

## **ACKNOWLEDGEMENTS**

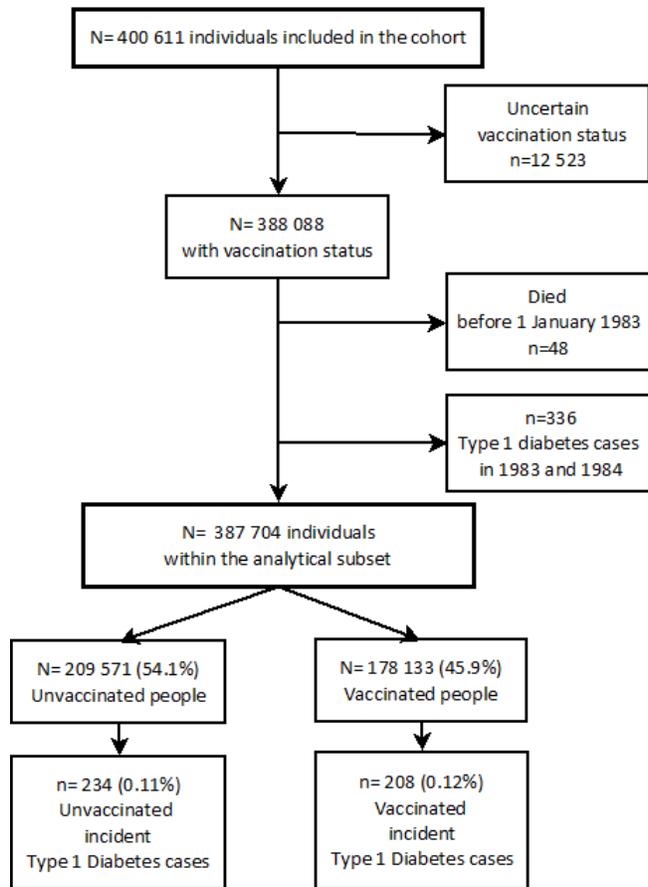
We gratefully acknowledge Isabelle Leroux and Jimmy Baulne from the *Institut de la statistique du Québec*, as well as François Blouin and Claude Verville from the *Régie de l'assurance maladie du Québec* for their contribution to various aspects of the establishment of the QBCIH. The *Institut de la statistique du Québec* was instrumental in realizing the linkages between administrative databases. We would like to express our sincere thanks to the personnel of the data access centre (*Centre d'accès aux données de recherche de l'Institut de la statistique du Québec, CADRISQ*) for their technical help and support. We also wish to acknowledge Dr. Tracie Barnett for her critical review of the manuscript.

Funding statement: Establishment of the QBCIH was supported by an infrastructure grant from the Canada Foundation for Innovation & the Quebec Ministry of Education, Leisure and Sports (#12532), research grants from the Canadian Institutes of Health Research (#MOP-97777, #MCH-97593, #MOP-142705, #PJT-159647), Fonds de recherche du Québec-Santé (FRQS, #16227), and Multiple Sclerosis Society of Canada (#2435). Drs. Benedetti and Parent are recipients of career awards from the FRQS.

**Declaration of conflicts of interest:** The authors have no conflict of interest to declare.

**Data availability:** The data that support the findings of this study are governed by strict data privacy considerations, require governmental bodies' approval, and so are not publicly available. For enquiries, please contact the last author.

**Contribution statement:** Philippe Corsenac: Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Original Draft, Visualization. Marie-Élise Parent: Conceptualization, Validation, Resources, Writing - Review & Editing, Supervision. Andrea Benedetti: Conceptualization, Validation, Writing - Review & Editing. Hugues Richard: Software, Data Curation, Validation, Writing - Review & Editing. Simona Stäger: Writing - Review & Editing, Supervision. Marie-Claude Rousseau: Conceptualization, Validation, Resources, Writing - Review & Editing, Supervision Project administration, Funding acquisition. All manuscript drafts were approved by all authors who are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.



**Figure 7.1: Flowchart of participants’ inclusions and exclusions in the analytical subset from the Quebec Birth Cohort on Immunity and Health (1970-2014), Quebec, Canada.**

**Legend for Figure 7.1:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 7.1: Description of the study population, overall (N= 387,704) and according to BCG vaccination status, Quebec Birth Cohort on Immunity and Health (1985-1993), Quebec, Canada.**

Variables	Overall study population		Vaccinated population		Unvaccinated population	
	N = 387 704 (100%)		N = 178 133 (45.9%)		N = 209 571 (54.1%)	
	N	(%) or Mean ± SD	N	(%) or Mean ± SD	N	(%) or Mean ± SD
<b>Sex</b>						
Male	199 380	51.4	91 203	51.2	108 177	51.6
Female	188 324	48.6	86 930	48.8	101 394	48.4
<b>Duration of gestation (weeks)</b>						
Mean ± SD	387 697	39.2 ± 1.7	178 132	39.2 ± 1.7	209 565	39.2 ± 1.8
Missing	7	-	1	-	6	-
<b>Duration of gestation (categorical)</b>						
Preterm birth	59 446	15.3	28 487	16.0	30 959	14.8
Full term birth	319 123	82.3	145 939	81.9	173 184	82.6
Post-term birth	9 128	2.4	3 706	2.1	5 422	2.6
Missing	7	0		0	6	0
<b>Birthweight for gestational age<sup>†</sup></b>						
Very small	60 476	15.6	25 512	14.3	34 964	16.7
Small	32 402	8.4	14 635	8.2	17 767	8.5
Appropriate	257 442	66.4	120 059	67.4	137 383	65.6
Large	37 293	9.6	17 896	10.0	19 397	9.3
Missing	91	0.0	31	0.0	60	0.0
<b>Number of older siblings</b>						
none	161 924	41.8	76 354	42.9	85 570	40.8
1 or more	214 428	55.3	99 914	56.1	114 514	54.6
Missing	11 352	2.9	1 865	1.0	9 487	4.5
<b>Area of residence 1987<sup>‡</sup></b>						
Rural	122 504	31.6	71 453	40.1	51 051	24.4
Urban	262 249	67.6	105 991	59.5	156 258	74.6
Missing	2 951	0.8	689	0.4	2 262	1.1
<b>Maternal age at childbirth (years)</b>						
Mean ± Std	378 620	26.6 ± 5.2	176 826	26.6 ± 5.1	201 794	26.7 ± 5.3
Missing	9 084	-	1 307	-	7 777	-
<b>Paternal age at childbirth (years)</b>						
Mean ± Std	370 672	29.5 ± 5.9	174 643	29.3 ± 5.8	196 029	29.7 ± 6.0
Missing	17 032	-	3 490	-	13 542	-
<b>Maternal birthplace</b>						
In Quebec	329 689	85.0	168 782	94.8	160 907	76.8

Outside	47 368	12.2	7 588	4.3	39 780	19.0
Missing	10 647	2.7	1 763	1.0	8 884	4.2
<b>Paternal birthplace</b>						
In Quebec	318 576	82.2	165288	92.8	153 288	73.1
Outside	50 348	13.0	8771	4.9	41 577	19.8
Missing	18 780	4.8	4 074	2.3	14 706	7.0
<b>Age at BCG vaccination</b>						
Unvaccinated	209 571	54.1	0	0	209 571	100.0
< 1 year old	156 342	40.3	156 342	87.8		0
≥ 1 year old	21 791	5.6	21 791	12.2		0
<b>Number of BCG vaccination</b>						
0	209 571	54.1	0	0	209 571	100.0
1	167 274	43.1	167 274	93.9		0
≥ 2	10 859	2.8	10 859	6.1		0

† For gestational ages between 21 and 44 weeks, defined as ‘very small’ if birth weight < 5th percentile, ‘small’ if ≥ 5th and < 10th percentile, ‘appropriate’ if ≥ 10th and < 90th percentile, and ‘large’ if > 90th percentile. Based on birth weights measured for each live birth occurring between 1994 and 1996 in all of the Canadian provinces except Ontario (Kramer et al., 2001).

‡ Determined using the second character of subjects’ residential postal code (0: rural, ≥1: urban).

**Legend for table 7.1:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 7.2: Crude and adjusted hazard ratios and 95% confidence intervals for the association between BCG vaccination, and age at vaccination, and the onset of incident type 1 diabetes, Quebec Birth Cohort on Immunity and Health (1985-1993), Quebec, Canada.**

	N subjects with type 1 diabetes (N=442)	Person-years (2,511,546)	Crude Hazard Ratio (95% CI ‡)	Adjusted Hazard Ratio † (95% CI ‡)
BCG § vaccination status				
Unvaccinated	234	1 363 643	1 (Ref)	1 (Ref)
Vaccinated	208	1 147 903	1.06 (0.88-1.27)	1.06 (0.88-1.29)
Age at BCG § vaccination				
Unvaccinated	234	1 363 642	1 (Ref)	1 (Ref)
< 1 year old	186	1 012 890	1.07 (0.88-1.30)	1.08 (0.88-1.32)
≥ 1 year old	22	135 014	0.95 (0.61-1.47)	0.96 (0.62-1.50)

† Models were adjusted for both maternal and paternal birthplace as a proxy of ethnicity, maternal and paternal age at childbirth (modeled as continuous variables), duration of gestation, birthweight for gestational age, having older siblings, and rural/urban area of residence in 1987. The adjusted hazard ratios and 95% confidence intervals for these variables are presented in Table 7.C of the supplementary file.

‡ CI, confidence interval.

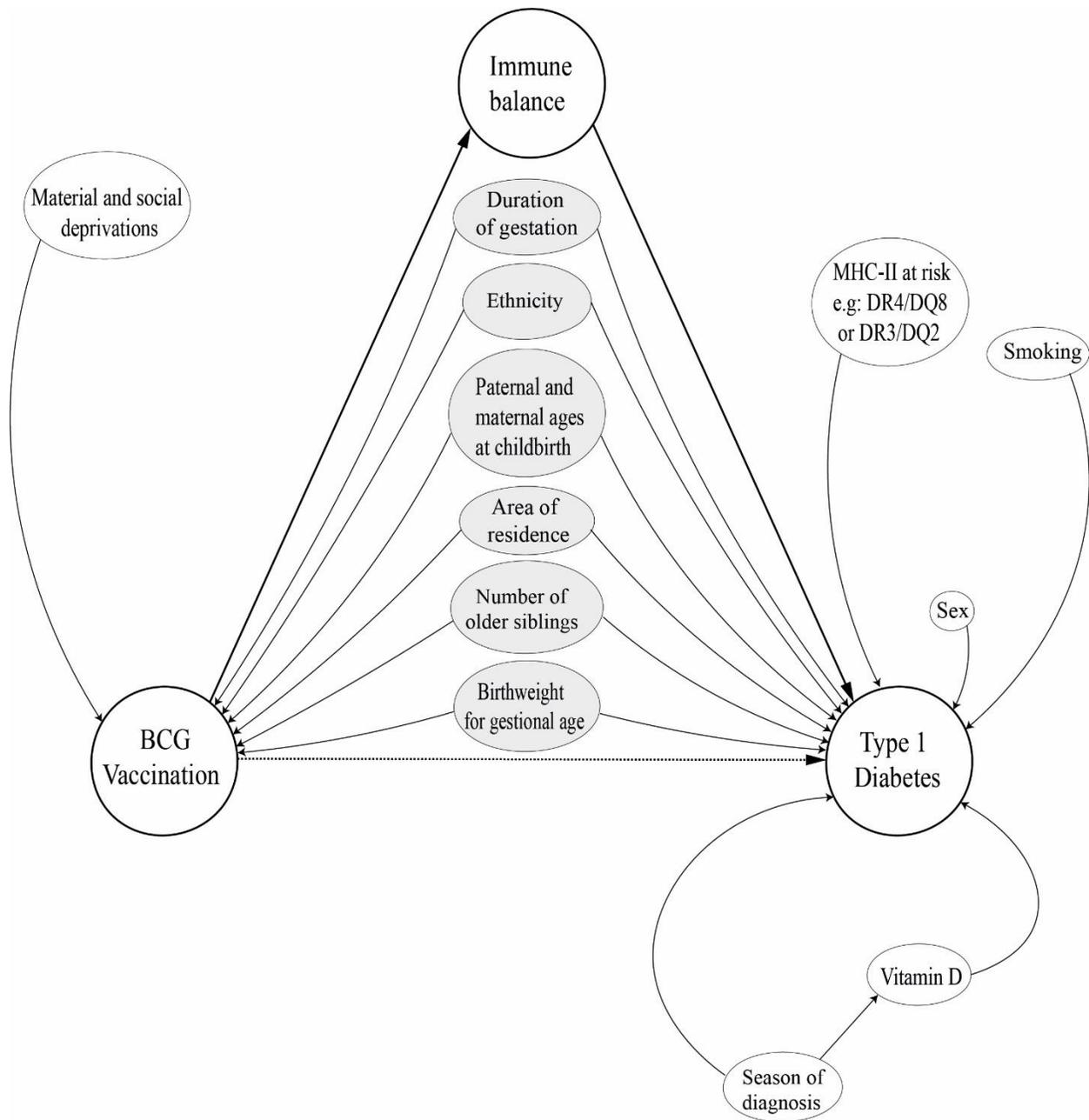
§ BCG, Bacillus Calmette–Guerin.

**Legend for table 7.2:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

## 7.8 Supplementary file

**Table 7.A : Factors considered in the directed acyclic graph of associations between BCG vaccine and the onset of type 1 diabetes, and supporting references**

Variables	Association with BCG	References	Association with type 1 diabetes	References
Season of diagnosis	No	-	Yes	(208, 389, 390)
Material and social deprivation	Yes	Income (307); education (368-370); socioeconomic status (369-372) (307)	No	-
Birth weight for gestational age	Yes	Post-term (307)	Yes	(235, 236)
Duration of gestation	Yes	(307)	Yes	Preterm birth (237, 238)
Number of older siblings	Yes	(307)	Yes	(267)
Area of residence	Yes	(307)	Yes	(269, 270)
Maternal age at childbirth	Yes	(307)	Yes	(241, 268)
Paternal age at childbirth	Yes	(307)	Yes	(241, 268)
Maternal smoking during pregnancy	No	-	Yes	(243)
Familial history of type 1 diabetes	No	-	Yes	(267)
Ethnicity	Yes	(307)	Yes	(391)
Sex	No	(282, 307)	Yes	(69, 207, 208)
Lack of Vitamin D	No	-	Yes	(389)
MHC-II at risk: DR4/DQ8 or DR3/DQ2	No	-	Yes	(25, 392)



**Figure 7.B:** Directed acyclic graph of the association between BCG vaccination and Type 1 diabetes.

**Legend for Figure 7.B:** Grey shading indicates confounding factors for the relationship between BCG vaccination and type 1 diabetes.

**Table 7.C:** Adjusted hazard ratios and 95% confidence intervals for confounders of the association between BCG vaccination and incident type 1 diabetes, Quebec Birth Cohort on Immunity and Health (1985-1993), Quebec, Canada.

	N subjects with type 1 diabetes <sup>†</sup> (N=442)	Person-years (2,511,546)	Adjusted HRs <sup>‡</sup> CIs	(95%
<b>BCG <sup>§</sup> vaccination status</b>				
Unvaccinated	234	1 363 643	1 (Ref)	
Vaccinated	208	1 147 903	1.06 (0.88-1.29)	
<b>Number of older siblings</b>				
none	183	1 098 963	1 (Ref)	
1 or more	159	1 412 583	0.94 (0.76-1.17)	
<b>Area of residence 1987</b>				
Rural	142	800 893	1 (Ref)	
Urban	300	1 710 653	0.98 (0.79-1.20)	
<b>Maternal birthplace</b>				
In Quebec	383	2 192 684	1 (Ref)	
Outside	59	318 862	1.10 (0.73-1.65)	
<b>Paternal birthplace</b>				
In Quebec	379	2 165 514	1 (Ref)	
Outside	63	346 032	0.97 (0.65-1.44)	
<b>Maternal age at childbirth</b>				
(in years)	442	2 511 546	1.04 (1.01-1.07)	
<b>Paternal age at childbirth</b>				
(in years)	442	2 511 546	1.00 (0.97-1.02)	
<b>Duration of gestation</b>				
(in weeks)	442	2 511 546	0.98 (0.93-1.04)	
<b>Birthweight for gestational age</b>				
Very small	60	387 611	0.88 (0.66-1.16)	
Small	31	209 425	0.84 (0.58-1.21)	
Appropriate	306	1 674 439	1 (Ref)	
Large	45	240 071	0.97 (0.69-1.36)	

<sup>†</sup> n were calculated on Imputation 1

<sup>‡</sup> Models were adjusted for all variables included in the table C

<sup>§</sup> BCG, Bacillus Calmette–Guerin.

Legend for table 7.C: These specific HRs were estimated with both BCG vaccination and age at vaccination as main exposure. Given the important similarities between the two sets of results, we only presented the HRs obtained with BCG vaccination as the main exposure.

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

## **8 Article 3. Early life Bacillus Calmette-Guerin vaccination and incidence of type 1, type 2, and latent autoimmune diabetes in adulthood**

---

### **Vaccination au Bacille Calmette-Guérin en bas âge et incidence du diabète de type 1, de type 2 et auto-immun latent à l'âge adulte**

**Auteurs :** Philippe Corsenac<sup>1</sup>, Marie-Élise Parent<sup>1</sup>, Hélène Mansaray<sup>2</sup>, Andrea Benedetti<sup>3,4</sup>, Hugues Richard<sup>1</sup>, Simona Stäger<sup>5</sup>, Marie-Claude Rousseau<sup>1</sup>

<sup>1</sup> Unité d'Épidémiologie et biostatistiques, Centre Armand-Frappier Santé biotechnologie, Institut national de la recherche scientifique, Laval, Canada;

<sup>2</sup> Institut de recherche clinique de Montréal, Montréal, Canada;

<sup>3</sup> Respiratory Epidemiology and Clinical Research Unit, Research Institute of the McGill University Health Centre, Montreal, Canada;

<sup>4</sup> Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, Montreal, Canada;

<sup>5</sup> Immunology et virologie, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique, Laval, Canada.

**Titre de la revue :** Diabetes & Metabolism ; publié le 1 Mars 2022.

DOI: 10.1016/j.diabet.2022.101337.

**Contribution des auteurs :**

Philippe Corsenac : Conceptualisation, Méthodologie, Analyses formelles, Structuration des données, Rédaction en version originale. Marie-Élise Parent : Conceptualisation, Validation, Ressources, Rédaction en révision & édition, Supervision. Hélène Mansaray : Conceptualisation, Méthodologie, Validation, Rédaction en relecture & édition. Andrea Benedetti : Conceptualisation, Validation, Rédaction en révision & édition. Hugues Richard : Data management, Structuration des Données, Validation, Rédaction en révision & édition. Simona Stäger : Rédaction en révision & édition, Supervision en immunologie. Marie-Claude Rousseau : Conceptualisation, Validation, Ressources, Rédaction en révision & édition, Supervision Administration de projet, Acquisition de financement. Toutes les ébauches du manuscrit ont été approuvées par tous les auteurs qui sont les garants de ce travail et, en tant que tels, ont eu un accès complet à toutes les données de l'étude et assument la responsabilité de l'intégrité des données et de l'exactitude de l'analyse des données.

## 8.1 Mise en contexte

A ce jour, aucune étude épidémiologique chez l'humain n'a été menée pour estimer la potentielle protection du vaccin au BCG à long terme sur les survenues de diabète de type 2 et des LADA, alors que Kühnreiter et al. ont mis en évidence chez la souris une protection contre l'hyperglycémie chronique, en dehors de toutes auto-immunités (41, 42). Ce mécanisme de glycolyse aérobie accompagne les modifications épigénétiques que provoque le vaccin au BCG (37, 41-43). Ce mécanisme pourrait également être protecteur vis-à-vis du diabète de type 1 en plus des mécanismes régulateurs contre l'auto-immunité. Les enseignements de l'article 2 nécessitent également d'identifier les cas de diabète de type 1 et d'étudier l'association d'intérêt à un âge adulte, différent de la période visée par toutes les précédentes études sur ce phénotype.

L'article 3, intitulé : *Early life Bacillus Calmette-Guerin vaccination and incidence of type 1, type 2, and latent autoimmune diabetes in adulthood* s'intéresse à estimer pour la première fois l'association entre la vaccination au BCG et les survenues de diabète de type 1, type 2 et LADA à un âge adulte, entre 22 et 44 ans. Il établit pour la première fois au Québec des prévalence et incidence de LADA et documente pour la première fois l'association entre la vaccination au BCG durant l'enfance et la survenue de diabètes de type 1, type 2 et LADA. Cet article met enfin en évidence la nécessité de futures analyses dans une population d'étude à un âge plus avancé étant donné un pic d'incidence du diabète de type 2 à 55 ans. Cet article souligne de plus la nécessité de distinguer tous les phénotypes de diabète, incluant la forme LADA, dans les études futures.

## 8.2 Abstract

**Aims.** - Bacillus Calmette-Guerin (BCG) vaccination limits blood sugar elevations and autoimmunity. Previous studies focused on type 1 diabetes among children, despite possible effects on other phenotypes later in life. We studied association between BCG vaccination and type 1, type 2 and latent autoimmune diabetes (LADA) in adulthood.

**Methods.** - A 1970-1974 birth cohort was linked with the BCG vaccination registry and administrative health data of Quebec. 396,118 people aged 22-44 years were followed-up for diabetes mellitus (DM) onset. Incident DM cases were subjects with  $\geq 1$  hospitalization or  $\geq 2$  physician claims related to DM over a 2-year period. Type 1 diabetes, type 2 diabetes, and LADA cases were individuals with  $\geq 1$  reimbursement of insulin, oral antidiabetic agent, or both. Cox proportional regressions were used to estimate hazard ratios (HR), adjusting for potential confounders.

**Results.** - Forty-four percent of subjects were BCG vaccinated, 88% of these before age 1. For type 1 diabetes, no association was found before 30 years old, but vaccinated subjects had a lower risk of this phenotype after age 30 (HR<sub>adj</sub>= 0.65, 95% CI: 0.44-0.95). BCG vaccination was associated with a lower risk of type 2 diabetes (HR<sub>adj</sub>=0.85, 95% CI: 0.79-0.92), whereas no association was observed for LADA (HR<sub>adj</sub>=1.30, 95% CI: 0.71-2.38). Results did not differ by sex.

**Conclusions.** - Early life BCG vaccination was associated with lower risks of both type 1 and type 2 diabetes from early to middle adulthood, but not of LADA. Future studies should explore these long-term associations, while distinguishing diabetes phenotypes.

**Keywords:** Adulthood; Bacillus Calmette-Guerin vaccination; Latent autoimmune diabetes in adults; Risk factor; Type 1 diabetes; Type 2 diabetes.

### 8.3 Introduction

Type 1 diabetes presents an acute autoimmune destruction of pancreatic beta cell islets, whereas type 2 diabetes results from insulin resistance and insufficient production (200). Both main phenotypes of diabetes mellitus (DM)<sup>1</sup> occur at all ages, although type 1 diabetes predominates before age 30 and type 2 diabetes at older ages (11). Latent autoimmune diabetes in adults (LADA) generates a slow progressive destruction of pancreatic beta cell islets, finally leading to an inability to produce insulin as in type 1 diabetes (6, 200). However, LADA occurs later than type 1 diabetes, typically between 20 and 45 years of age, although this range remains debated (6).

The worldwide prevalence of DM is 6%, of which type 1 diabetes represents 10% (13) and type 2 diabetes 90% of all cases (204). LADA is often misclassified as either type 1 or type 2 diabetes. It may represent 2 to 12% of all adult DM cases (6, 200), but cohorts enabling reliable estimates of its incidence are rare.

The Bacillus Calmette-Guerin (BCG) vaccine could, through non-specific effects, limit blood sugar elevations and autoimmunity (40-42). For instance, some authors explain protection against autoimmunity by a better tumor necrosis factor (TNF- $\alpha$ ) production (40, 298) which can signal through two distinct TNF receptors, 1 and 2 (TNFR1 and 2). TNFR1, more likely carried by autoimmune lymphocytes, leads to their apoptosis, whereas TNFR2 carried by regulatory T lymphocytes could favor their survival (40, 298). Trained immunity (37, 39) of innate cells could also improve regulatory responses by feedback of elevated levels of pro-inflammatory cytokines production. Such regulatory mechanism may counteract chronic inflammation produced by both obesity and/or autoimmunity, allowing protection against all phenotypes of diabetes. Furthermore, BCG vaccination may have different effects according to sex (46) and age at vaccination. Vaccination during the first year of life may lead to a better heterologous innate stimulation and proinflammatory cytokines/chemokines production (47, 48).

---

<sup>1</sup> **Abbreviations:**

BCG, Bacillus Calmette-Guerin; DM, Diabetes Mellitus; LADA Latent Autoimmune Diabetes In Adults  
IDS, the Immunology Diabetes Society; JDS, the Japanese Diabetes Society; OAD, oral antidiabetic agents  
QBCIH, Quebec Birth Cohort on Immunity and Health

All previous epidemiological studies assessing the association between BCG vaccination and DM focused on type 1 diabetes in childhood or adolescence (61-67). These studies suggested no association (61-67), but most suffered from methodological limitations (61, 64-67). No study has examined this association in adult populations or with other forms of DM like type 2 diabetes or LADA.

We first studied the association between BCG vaccination, or age at vaccination, and both type 1 and type 2 diabetes in adults. Then, we estimated these associations after applying a classification of DM cases that took LADA into account.

## **8.4 Material and Methods**

### **Study design and data sources**

The study used an expanded version of the Quebec Birth Cohort on Immunity and Health (QBCIH) (303), which was established by probabilistic linkage of the birth registry and of the Healthcare Registration file using surname, first name, sex, date of birth, and father's first name as common identifiers.

Probabilistic linkage was used to add data from the death and Quebec BCG vaccination registries, whereas administrative health data were merged using a unique identifier.

### **Study population**

Eligible subjects were born in Quebec between 1970 and 1974 after  $\geq 32$  weeks of gestation, documented in the birth registry and covered by the public health system. The cohort initially included 400,611 individuals, representing 90.4% of eligible persons for whom the initial probabilistic linkage was successful.

Physician claims and drug reimbursements were available from January 1<sup>st</sup> 1983 and hospitalizations from April 1<sup>st</sup> 1987, until December 31<sup>st</sup> 2014. Only persons on social assistance had public medication insurance in the period 1983-1996 (5% of the cohort), and from 1997 coverage was extended to any person who did not have medication insurance through employment. This allowed for a more comprehensive follow-up of medication from 1997 to 2014, when cohort members were between 22 and 44 years of

age. We expected in this age range a higher incidence of LADA and type 2 diabetes, and a lower though still sizeable incidence of type 1 diabetes. We followed individuals who had not been previously identified as having DM, from January 1<sup>st</sup> 1997 to December 31<sup>st</sup> 2014 for their use of medical services and drug reimbursements for DM.

Analyses were conducted in three analytical samples (Figure 8.1). The eligible study sample included 396,118 individuals (98.9% of initial cohort members), after exclusions and left-truncation on January 1<sup>st</sup> 1997. Sample 2 included subjects who had drug insurance coverage for any length of time during the follow-up ( $N_2=260,440$ , 65.7% of eligible sample). Sample 3 included individuals who had long-term drug insurance coverage (i.e., were covered for  $\geq 90\%$  of the follow-up) and those who had reimbursements for DM medication in their first period of coverage ( $N_3=160,622$ , 40.5% of eligible sample). Associations between BCG vaccination, or age at vaccination, and type 1 and type 2 diabetes were first estimated within sample 2 and then re-estimated within sample 3, in which LADA cases were also identified and analysed. Only 0.23% and 0.37% of cohort members in samples 2 and 3 had postal codes missing over the entire follow-up, potentially suggesting that they had been lost to follow-up.

### **Outcome definitions**

Physician claims were coded using the International Classification of Disease (ICD) 9<sup>th</sup> revision (4-digit, 250.X). For hospitalizations, ICD-9 diagnostic codes (most at the 5-digit level, 250.XX) were available until 2006, then ICD-10 codes (E10-14) were used from April 2006 to December 31<sup>st</sup> 2014 (319). First, we used the Canadian Chronic Disease Surveillance System (CCDSS) DM definition, validated in adults (316, 317), with good sensitivity (88.9–89.8%) and high specificity (97.5–97.7%) (318). Incident DM cases were those in the eligible sample with  $\geq 1$  hospitalization or  $\geq 2$  physician claims for DM within two years, with  $\geq 1$  day between two services (319). To exclude gestational diabetes, evidence of DM in women aged 10-54 years was excluded if it occurred 120 days before or 180 days after hospital records or physician visits containing any pregnancy-related or obstetrical code (ICD-9/ICD-10: 641-676, V27/O1, O21-95, O98, O99, Z37) (319). In our data, the codes for normal delivery and outcome of delivery (ICD-9/ICD-10: 650, V27/O80, Z37) were not systematically present. However, using data from physician

claims as opposed to only those from hospitalizations allowed us to identify any pregnancy with complications.

Second, we identified type 1 and type 2 diabetes based on reimbursements for insulin and/or OAD over the follow-up period within sample 2. Type 1 diabetes is immediately treated by insulin (200). Type 2 diabetes is initially treated with oral antidiabetic agents (OAD) (200), but some patients require insulin therapy following beta cell failure, which occurs on average 8 years after diagnosis (198). Type 1 diabetes was thus defined as persons using only insulin and type 2 diabetes as those who used only OAD or both insulin and OAD (as insulin-requiring type 2 diabetes).

The date of DM diagnosis was the date of the second physician claim or the date of hospital discharge (whichever came first) (319, 393), except for cases found only through medication. For them, the date of first insulin or of OAD reimbursement (whichever came first) was set as the date of diagnosis. Tables 8.A1 and 8.A2 (see supplementary materials associated with this article on line) show the list of insulins and OAD used in our cohort.

Third, we identified type 1, type 2 diabetes, and LADA cases in sample 3. Unlike type 1 diabetes, LADA patients do not need insulin therapy for 3-6 months to a year after diagnosis (6, 200) and they present insulin resistance like type 2 diabetes (7). Thus, they are sometimes initially treated with OAD but will soon need insulin (6, 200). LADA cases were defined as those receiving either insulin, or insulin and OAD, based on the delay elapsed between DM diagnosis and uptake of the first insulin therapy or between the first OAD and uptake of the first insulin therapy. In 2005, the Immunology of Diabetes Society (IDS) established that LADA cases should be older than 35 years and require insulin 6 months or more after diagnosis (6, 320), although the lower age limit remains debated (6). More recently, the Japan Diabetes Society (JDS), based on comparison with acute type 1 diabetes, proposed that age of onset should not be restricted to > 35 years and that the diagnosis of LADA should be made after a 3-month period in which no insulin is required (6). We used the aforementioned two delays in insulin requirement to identify two series of LADA cases: 1) between 6 months and 1 year (IDS definition), and 2) between 3 months and 1 year (JDS definition). Given that delays are still debated (201), the maximum delay was chosen as one year based on our team's clinical observations. In keeping with the most

recent evidence, we did not apply age restrictions. Delays were calculated from the date of DM diagnosis until the first reimbursement for insulin. For cases identified only through medication, gestational diabetes was also excluded as described previously and the delay was calculated between the first reimbursement for OAD and the first reimbursement for insulin (319).

### **Exposure definition**

An organized voluntary program of vaccination with BCG was held in the province of Quebec from 1949 to the mid-1970s (279). The program targeted newborns and schoolchildren. Vaccination certificates were sent to the centralized Quebec BCG vaccination registry. It is now computerized and was found to be comprehensive and accurate (279). For a small proportion of cohort members, the probabilistic linkage generated a low, but non-null, probability of linkage to the registry. These persons were considered as having an undetermined vaccination status. Vaccination status was first defined as a binary variable (yes / no) and then as a categorical variable according to age at BCG vaccination (not vaccinated / vaccinated < 1 year / vaccinated  $\geq$  1 year).

### **Statistical analysis**

Cox proportional hazards regression analyses were performed with age as the time scale to assess the association between BCG vaccination and the onset of type 1, type 2 diabetes, and LADA.

Follow-up began in 1997 and ended at the date of DM diagnosis, at death, or on December 31<sup>st</sup> 2014, whichever came first. The proportionality assumption was verified by inspecting the Schoenfeld standardized residuals.

Since the non-specific effects of BCG vaccination would vary by sex, potential effect modification by sex was assessed by introducing an interaction term in the multivariable model (46).

Potential confounders were selected according to directed acyclic graphs (DAGs). Tables 8.A3 and 8.A4 (see supplementary materials associated with this article on line) show the literature on which the DAGs were based (Figure 8.A1; see supplementary materials

associated with this article on line). Type 2 diabetes shares the same risk factors as LADA (27, 209).

There were 37,834 (9.6%) subjects with missing data, including 12,382 (3.1%) individuals with undetermined vaccination status. Multiple imputation by chained equations was applied (n=5 datasets), using the distributions of other informative variables included in the dataset.

The functional forms of the continuous covariates, parents' age at childbirth, duration of gestation and birthweight were studied as linear, categorical, squared, log transformed and restricted cubic spline variables to determine the best model fit with the Stanford University SAS macro (%SPECI).

Because of differences in the socioeconomic characteristics between the eligible study sample and the two analytical drug insurance samples (N<sub>2</sub> and N<sub>3</sub>), we calculated weights through stabilized inverse probability weighting procedures using propensity scores (338). These scores were estimated with a logistic regression model for each sample (N<sub>2</sub> and N<sub>3</sub>) in which the dependent variable was a binary variable indicating whether each individual belonged in that sample. For both models, the independent variables included: year of birth, sex, birthweight, number of older siblings, maternal age at childbirth, maternal birthplace, area of residence in 1997, material and social deprivation in 1997, and income in 1996. We then examined whether weighting balanced measured covariates between those covered and not covered by drug insurance (Figures 8.A2 to 8.A19; see supplementary materials associated with this article on line).

All analyses were performed with SAS software version 9.4 and with R version 4.1.0.

### **Ethics approval and access to data**

The ethics committees of Institut national de la recherche scientifique, Institut de la statistique du Québec and Régie de l'assurance maladie du Québec, approved the study and access to data was granted by the Commission d'accès à l'information du Québec.

## 8.5 Results

The eligible study population (N=396,118) is presented in Table 8.1. The female to male ratio was 0.96. Mothers were on average 27 years old at childbirth, whereas fathers were 30 years of age. The mean age in 1997 was 24.5 years old (standard deviation=1.5; range 22-27). Among subjects who received the BCG vaccine (n=176,209, 44.5%), most were vaccinated before one year of age (87.8%) and received the vaccine only once (93.9%). Individuals who had drug insurance (N<sub>2</sub>=260,440), were covered for a median duration of 5.9 years.

Among various functional forms of duration of gestation, parents' age at childbirth, and birthweight, linear representations were selected based on the Akaike Information Criterion (lowest value indicating best fit).

### ***Association between BCG vaccination and type 1 or type 2 diabetes based on drug reimbursements – Sample 2 (subjects with drug insurance)***

In this sample, 5,582 DM cases could be categorised into type 1 or type 2 diabetes (Figure 8.2). The proportionality assumption was met for all regression models. We observed no association between BCG vaccination (or age at vaccination) and type 1 diabetes (HR<sub>adj</sub>=0.94, 95% CI: 0.77-1.15), whereas vaccinated subjects had a lower risk of type 2 diabetes overall and for those vaccinated before 1 year of age (HR<sub>adj</sub>=0.90, 95% CI: 0.85-0.95), compared with unvaccinated subjects after adjusting for relevant confounders (Table 8.2).

### ***Association between BCG vaccination and type 1, type 2 diabetes, or LADA – Sample 3 (subjects with long-term drug insurance)***

The overall prevalence of DM between 1997 and 2014 (N<sub>3</sub>=160,622), based on the CCDSS definition and on medication data was 3.80%, 95% CI: 3.70-3.89%. The incidence rate of DM over this period was 213.9 cases per 100,000 person-years (95% CI: 208.5-219.2). The prevalence and incidence by phenotype of DM according to the IDS and JDS definitions are summarized in Table 8.3. The IDS definition detected 65% less cases of LADA than the JDS definition for a period prevalence of 0.03% (95% CI: 0.02-0.04) versus 0.05% (95% CI: 0.04-0.06). The female to male ratios of LADA were 2.1 and 1.5 with IDS and

JDS definitions, respectively. Of note, 86% and 84% of LADA detected by the IDS and JDS definitions were classified as type 2 diabetes in our analyses on sample 2. The mean ages at diagnosis for the different phenotypes are summarized in Table 8.A5 (see supplementary materials associated with this article on line)

The associations between BCG vaccination and the three DM phenotypes are presented in Table 8.4. The multivariable regression models for type 1 diabetes were stratified at age 30 years (cut-off point at 11,100 days, chosen graphically) and some adjustment variables had to be omitted (specified in Table 8.4 footnotes) to meet the proportionality assumption. Using the IDS definition, we observed decreased risks of type 1 diabetes after 30 years old among vaccinated individuals (HRadj= 0.65, 95% CI: 0.44-0.95), but not before age 30 (1.06, 95%CI: 0.74-1.52). The protective association between BCG vaccination and type 1 diabetes was observed in subjects vaccinated before age 1 (HRadj= 0.66, 95% CI: 0.45-0.98). The association among subjects vaccinated at or after 1 year of age, as compared with unvaccinated ones, was imprecise due to small numbers (HRadj= 0.56, 95% CI: 0.22-1.42). Protective associations were observed for type 2 diabetes among vaccinated compared with unvaccinated subjects (HRadj= 0.85, 95% CI: 0.79-0.92), independently of age at vaccination. An elevated point estimate was observed for BCG vaccination and LADA, although inconclusive due to imprecision (HRadj= 1.30, 95% CI: 0.71-2.38). Similar tendencies were observed with all clinical types of DM using the JDS definition. Findings did not differ by sex. *P*-values for the interaction terms between sex and BCG vaccination status ranged globally between 0.15 and 0.99 through the different associations and the 5 imputed datasets.

## **8.6 Discussion**

This study is the first to document protective associations between BCG vaccination and type 1 and type 2 diabetes onset in adults. By contrast, the small number of identified LADA precluded formulating conclusions for this phenotype.

Our findings are derived from a population-based birth cohort using BCG vaccination status ascertained from a registry and applying the CCDSS validated definition (316-319)

for firstly identifying overall DM cases from administrative health data. The second definition was developed by our group to identify type 1 and type 2 diabetes, and relied on reimbursement for diabetes drug prescriptions. Applying the IDS and JDS definitions for LADA, based on time before requiring insulin within a sub-sample of subjects with long-term drug insurance coverage, we developed a third definition merging information from definition 2 and time before requiring insulin. Albeit not yet validated, the second and third definitions represent original contributions, the latter allowing us to identify LADA cases by re-classifying some type 1 and type 2 diabetes cases.

Between 1997 and 2014, the period prevalence and incidence of DM were 3.80% (95% CI: 3.70%-3.89%) and 213.9 cases (95% CI: 208.5-219.2) per 100,000 person-years respectively, among individuals aged 22-44 years. In Canada in 2016, the prevalence of DM among persons aged 20-34 years and 35-49 years was estimated at 1.2% and 4.8% respectively, and the incidence at 140 and 550 per 100,000 (394)]. Our findings thus meet the expected order of magnitude.

A meta-analysis by Mobasser *et al.* (13), including some studies conducted in Canada and covering our study period, reported an overall incidence of type 1 diabetes around 20 per 100,000 person-years in the general population. We estimated the incidence in our study at 8.7 cases (95% CI: 7.6-9.7) per 100,000 person-years, in a subgroup of population aged 22 to 44 years old, which is within expectations given that this age group does not correspond to the peak of incidence. In 2017, Khan *et al.* estimated the global prevalence and incidence of type 2 diabetes between 2% and 7%, and between 250/100,000 and 400/100,000, for age groups 20-24 and 40-44 years, respectively (204). In our cohort, the prevalence and incidence of type 2 diabetes were lower than these estimates at 1.87% (95% CI: 1.81-1.94) and 105.6 cases (95% CI: 101.8-109.3) per 100,000 person-years, respectively. However, almost as many cases (n=2,766) identified with the CCDSS definition remained of unspecified phenotype in the subgroup with long-term drug insurance coverage (N<sub>3</sub>), since they did not have medication reimbursements. The prevalence and incidence of unspecified DM were thus estimated at 1.72% (95% CI: 1.66-1.79) and 97.0 cases (95% CI: 93.4-100.6) per 100,000 person-years (N<sub>3</sub>=160,622). Age at diagnosis was similar between individuals with type 2 diabetes and those with unspecified

DM (Table 8.A5; see supplementary materials associated with this article on line). At least 45-50% of type 2 diabetes patients are not adherent to their treatment (395, 396). Therefore, the unspecified DM cases, living neither with OAD nor insulin, could be classified as potential early or non-compliant type 2 diabetes cases. Combining the type 2 diabetes and unspecified DM cases would result in type 2 diabetes prevalence and incidence comparable to those reported by Khan *et al* (204). The novelty of our work also resided in estimating the prevalence of LADA (0.05%, 95% CI: 0.04%-0.06%). This represented 1,3% of all DM cases, in the lower range of previous estimates: 2 to 12% of all adult DM cases (6, 200).

We found no association between BCG vaccination and type 1 diabetes before age 30. This finding is in line with our previous analyses conducted in children and adolescents (63, 397). For the first time, we observed a protective association between BCG vaccination and the onset of type 1 diabetes ( $\geq 30$  years) and type 2 diabetes among individuals aged 22-44 years. This can be considered an “old” population for type 1 diabetes and a “young” one for type 2 diabetes, since the peaks of incidence are situated at 14 and 55 years old, respectively (69, 394). Both associations are novel, and will require further investigation. Late type 1 diabetes could present a different physiopathogenesis than acute type 1 diabetes occurring at a younger age. The results on an association between BCG vaccination and type 2 diabetes need to be reproduced in an older population that would encompass the age at which type 2 diabetes incidence peaks. As a sensitivity analysis, we estimated the association between BCG vaccination and type 2 diabetes, regrouping the type 2 diabetes cases with medications and the unspecified DM cases ( $n=2,766$ ), treating the latter as potential type 2 diabetes cases (Table 8.A6; see supplementary materials associated with this article on line). We observed a protective effect of BCG, although not quite as pronounced as when considering only the type 2 diabetes cases with medications. While the JDS definition allowed for the detection of more LADA cases than the IDS definition, a certain number of LADA cases may have remained unidentified among type 1 and type 2 diabetes cases. The protective association between BCG vaccination and type 1 diabetes ( $\geq 30$  years) and type 2 diabetes could thus be stronger than observed. Within the first analysis ( $N_2=260,440$ ), we indeed found no association between BCG vaccination and the onset of type 1 diabetes, and the association between BCG vaccination and type 2 diabetes

was weaker than what we observed within sample 3 after identifying and analyzing the LADA cases separately. About 85% of LADA detected were classified as type 2 diabetes and 15% as type 1 diabetes in our analyses on sample 2. Given the limited number of LADA cases identified within the cohort (n=49 and 75, using IDS' and JDS' definitions), risk estimates were unstable and results must be interpreted cautiously.

Non-specific effects of BCG vaccination have been demonstrated in humans within clinical trials (40-42). The BCG vaccine improved glycolysis by switching oxidative phosphorylation to aerobic glycolysis, and re-established normal glycaemia in individuals with acute autoimmune diabetes (40-42). This was also observed in mice with and without hyperglycemia, and in the absence of autoimmunity (41, 42). BCG could thus reduce blood sugar elevations in mice without autoimmunity and be protective against type 2 diabetes (41), a mechanism that is equally relevant for type 1 diabetes and LADA. Due to their autoimmune origin, type 1 diabetes and LADA could also be prevented by two other immunological mechanisms. Firstly, the BCG vaccine could favor immunological tolerance through enhancement of regulatory mechanisms (37, 41). Secondly, it leads to epigenetic changes in innate cells (37) increasing their pro-inflammatory pathways and counteracting autoimmune processes by feedback through regulatory mechanisms (40, 386). Such regulatory mechanisms may also limit obesity-related chronic inflammation and protect against insulin resistance in type 2 diabetes and LADA. Moreover, long-lasting effects of the BCG vaccine, even with single neonatal injection have been recently re-evaluated to 50 years (37, 274). Results from the present study concur with this experimental evidence.

### **Limitations**

The identification of the three main phenotypes of DM relied on information about medication reimbursements from public drug insurance, covering the most socio-economically deprived people. However, the majority (65.7%, N<sub>2</sub>=260,440) of cohort members had some drug insurance coverage during the follow-up period, with a median cumulative duration of 5.9 years. To ensure that there was sufficient insurance coverage to capture diabetes medication use, we further restricted the analysed subset to those cohort members who had long-term coverage ( $\geq 90\%$  of the follow-up period) or had their first

drug reimbursements within their first period of coverage (40.5%,  $N_3=160,622$ ). Subjects identified as having  $\geq 1$  diabetes medication over the follow-up had a median cumulative duration of drug insurance of 12.9 years. They were thus covered by the public drug insurance for a longer period than others who did not take medication for diabetes. To mitigate this issue, we used inverse probability weighting and achieved a balance of sociodemographic characteristics in the subsets of those who had drug insurance (samples 2 and 3) versus the eligible study sample.

We may have misclassified a small number of very specific cases like inaugural type 2 diabetes with important hyperglycemia or diabetic ketoacidosis needing insulin, or obese people with type 1 diabetes taking OAD to limit insulin resistance.

The IDS and JDS definitions were useful to circumscribe LADA and to refine type 1 and type 2 diabetes cases. However, more information, for instance on BMI and glutamic acid decarboxylase autoantibodies would have improved our algorithm (6, 200). Furthermore, the clinical label of type 1 and 2 diabetes could not be used to identify phenotypes because the 5-digit ICD-9 codes and the ICD-10 codes, which would have conveyed this information were only available within hospitalization records, for a small proportion of the diabetes cases (11.5%; 1,139/9,874) (Figure 8.2). Using IDS and JDS definitions to filter LADA from type 1 and 2 diabetes without other information was thus a limitation of our method, highlighted by the small number of LADA cases identified. Our findings on LADA must be interpreted very carefully. Similarly for individuals receiving BCG after age 1, the small numbers identified in sample 2 and 3 preclude all inference.

### **Strengths**

The present study is based on a vast and highly representative population-based birth cohort (303). The highlighted protective effects of early life BCG vaccination against type 1 and type 2 diabetes in adulthood are compatible with long-lasting effects of the vaccine (37). The study benefited from a reliable assessment of exposure to BCG through a provincial registry. Moreover, we could ascertain DM outcomes using a validated definition and two novel ones based on medications. We were thus able to study BCG vaccination with the three main phenotypes of DM in adults for the first time. This is important, given the

different pathogenesis of autoimmune and type 2 diabetes, since these differences would be expected to result in distinct determinants.

## 8.7 Conclusions

For the first time, we observed protective associations between BCG vaccination in early life, or age at vaccination, and the incidence of type 1 and type 2 diabetes in adults aged 22 to 44 years. Results for LADA, based on few exposed subjects, remain inconclusive. These findings need to be reproduced in clinical cohorts, population-based studies or using diabetes registries which will allow a valid classification of clinical phenotypes.

## ACKNOWLEDGEMENTS

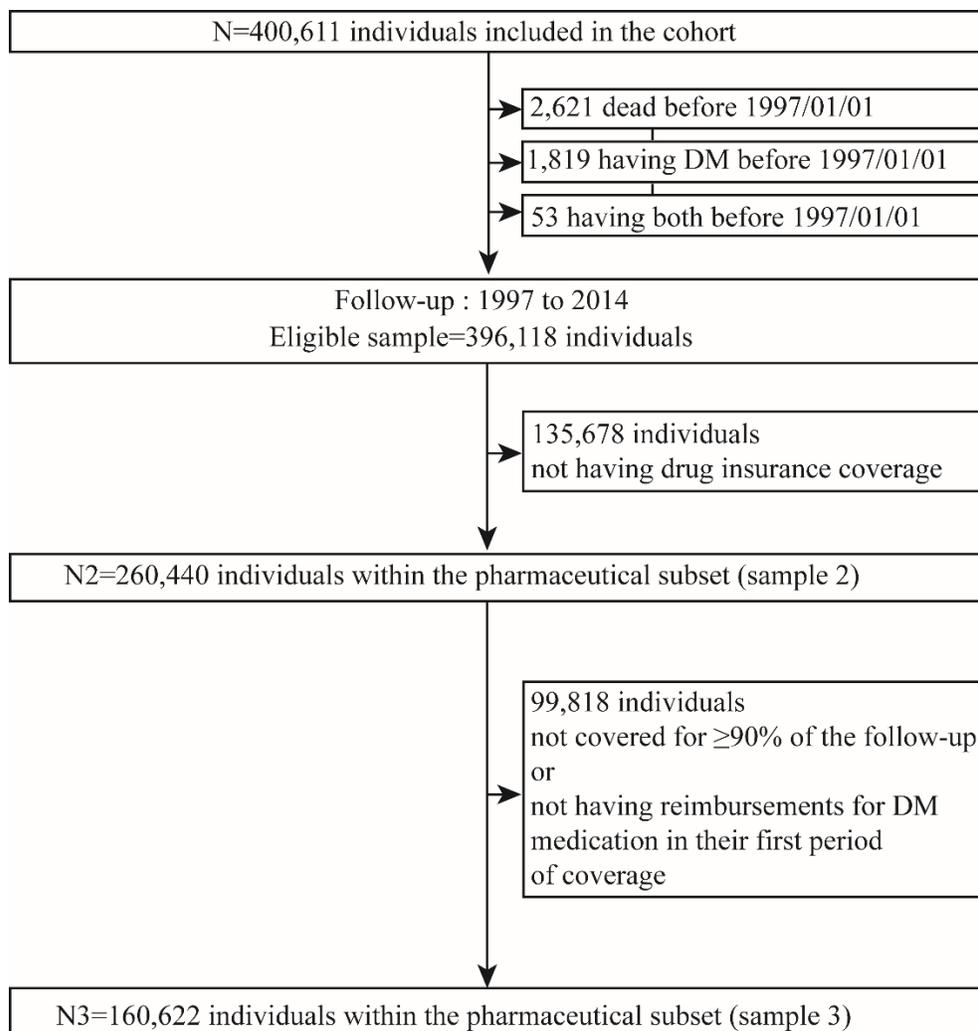
We gratefully acknowledge Isabelle Leroux and Jimmy Baulne from the *Institut de la statistique du Québec*, as well as François Blouin and Claude Verville from the *Régie de l'assurance maladie du Québec* for their contribution to various aspects of the establishment of the QBCIH. The *Institut de la statistique du Québec* was instrumental in realizing the linkages between administrative databases. We would like to express our sincere thanks to the personnel of the data access centre (*Centre d'accès aux données de recherche de l'Institut de la statistique du Québec, CADRISQ*) for their technical help and support.

**Funding statement:** Establishment of the QBCIH was supported by an infrastructure grant from the Canada Foundation for Innovation & the Quebec Ministry of Education, Leisure and Sports (#12532), research grants from the Canadian Institutes of Health Research (#MOP-97777, #MCH-97593, #MOP-142705, #PJT-159647), Fonds de recherche du Québec-Santé (FRQS, #16227), and Multiple Sclerosis Society of Canada (#2435). Drs. Benedetti and Parent are recipients of career awards from the FRQS.

**Declaration of conflicts of interest:** The authors have no conflict of interest to declare.

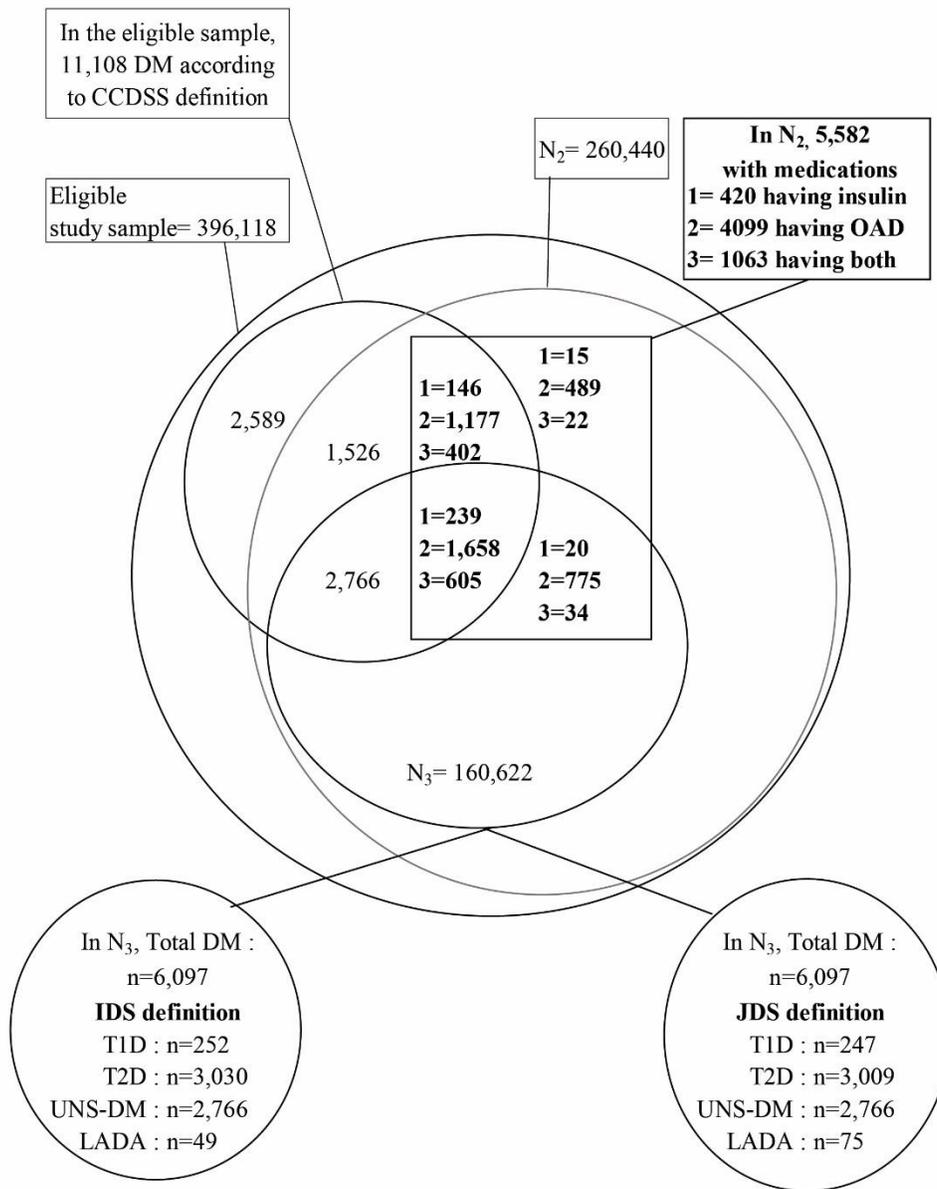
**Data availability:** The data that support the findings of this study are governed by strict data privacy considerations, require governmental bodies' approval, and so are not publicly available. For enquiries, please contact the last author.

**Contribution statement:** Philippe Corsenac: Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Original Draft, Visualization. Marie-Élise Parent: Conceptualization, Validation, Resources, Writing - Review & Editing, Supervision. H  l  ne Mansaray: Conceptualization, Methodology, Validation, Writing - Review & Editing. Andrea Benedetti: Conceptualization, Validation, Writing - Review & Editing. Hugues Richard: Software, Data Curation, Validation, Writing - Review & Editing. Simona St  ger: Writing - Review & Editing, Supervision. Marie-Claude Rousseau: Conceptualization, Validation, Resources, Writing - Review & Editing, Supervision Project administration, Funding acquisition. All manuscript drafts were approved by all authors who are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.



**Figure 8.1: Flowchart of participants’ inclusions and exclusions in the three analytical subsets from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.**

**Legend for Figure 8.1:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.



**Figure 8.2: Diabetes mellitus (DM) definitions among participants included in the analytical subsets from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.**

**Legend for Figure 8.2:** OAD, Oral antidiabetic agents; CCDSS, the Canadian Chronic Disease Surveillance System; DM, Diabetes mellitus; IDS, Immunology of Diabetes Society definition; JDS, Japanese Diabetes Society definition; LADA, Latent Autoimmune

Diabetes in Adults; T1D, Type 1 Diabetes; T2D, Type 2 Diabetes; UNS, unspecified= Unknown clinical form.

The eligible study sample represented 98.8% of initial cohort members of QBCIH (N=396,118), after exclusions and left-truncation on January 1<sup>st</sup> 1997. The sample 2 represented 65.7% of the eligible study sample and included subjects who had drug insurance coverage for any length of time during the follow-up (N<sub>2</sub>=260,440). The sample 3 included individuals who had long-term drug insurance coverage (i.e., were covered for  $\geq 90\%$  of the follow-up) and those who had reimbursements for DM medication in their first period of coverage (N<sub>3</sub>=160,622, 40.5% of the eligible study sample).

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 8.1: Description of the study population, overall (N<sub>1</sub>=396,118) and according to BCG vaccination status, Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.**

Variables	Overall study population		Vaccinated population		Unvaccinated population		Subjects with undetermined vaccination status	
	N = 396 118 (100%)		N = 176 209 (44.5%)		N =207 527 (52.4%)		N =12 382 (3.1%)	
	N	(%) or Mean ± SD	N	(%) or Mean ± SD	N	(%) or Mean ± SD	N	(%) or Mean ± SD
<b>Sex</b>								
Male	202 327	51.1	89 633	50.9	106 430	51.3	6 264	50.6
Female	193 791	48.9	86 576	49.1	101 097	48.7	6 118	49.4
<b>Duration of gestation (weeks)</b>								
Mean ± SD	396 111	39.2 ± 1.7	176 208	39.2 ± 1.7	207 521	39.2 ± 1.7	12 382	39.2 ± 1.7
Missing	7	-	1	-	6	-	0	-
<b>Birthweight (in grams)</b>								
Mean ± SD	396 085	3240 ± 508	176 201	3259.4 ± 490	207 503	3223 ± 523	12 381	3258.8 ± 487
Missing	33	-	8	-	24	-	1	-
<b>Birthweight for gestational age <sup>a</sup></b>								
Very small	61 610	15.6	25 206	14.3	34 615	16.7	1 789	14.4
Small	33 124	8.4	14 504	8.2	17 593	8.5	1 027	8.3
Normal	263 170	66.4	118 781	67.4	136 056	65.6	8 333	67.3
Large	38 122	9.6	17 687	10.0	19 206	9.3	1 229	9.9
Missing	92	0	31	0	57	0	4	0
<b>Number of older siblings</b>								
None	165 902	41.9	75 573	42.9	84 799	40.9	5 530	44.7
1 or more	218 660	55.2	98 790	56.1	113 357	54.6	6 513	52.6

Missing	11 556	2.9	1 846	1.0	9 371	4.5	339	2.7
<b>Area of residence 1987<sup>b</sup></b>								
Rural	124 947	31.5	70 548	40.0	50 405	24.3	3 994	32.3
Urban	268 133	67.7	104 975	59.6	154 865	74.6	8 293	67.0
Missing	3 038	0.8	686	0.4	2 257	1.1	95	0.8
<b>Maternal age at childbirth (years)</b>								
Mean ± Std	386 833	26.6 ± 5.2	174 919	26.5 ± 5.1	199 855	26.7 ± 5.3	12 059	26.3 ± 5.4
Missing	9 285	-	1 290	-	7 672	-	323	-
<b>Paternal age at childbirth (years)</b>								
Mean ± Std	378 037	29.5 ± 5.9	172 759	29.3 ± 5.8	194 177	29.7 ± 6.0	11 101	29.3 ± 6.0
Missing	18 081	-	3 450	-	13 350	-	1 281	-
<b>Maternal birthplace</b>								
In Quebec	337 506	85.2	166 943	94.7	159 224	76.7	11 339	91.6
Outside	47 732	12.0	7 524	4.3	39 530	19.0	678	5.5
Missing	10 880	2.7	1 742	1.0	8 773	4.2	365	2.9
<b>Paternal birthplace</b>								
In Quebec	325 620	82.2	163485	92.8	151 726	73.1	10 409	84.1
Outside	50 644	12.8	8696	4.9	41 296	19.9	652	5.3
Missing	19 854	5.0	4 028	2.3	14 505	7.0	1 321	10.7
<b>Material deprivation 1987</b>								
Very privileged	65 869	16.6	20 444	11.6	43 595	21.0	1 830	14.8
Privileged	76 612	19.3	32 554	18.5	41 545	20.0	2 513	20.3
Moderately privileged	77 148	19.5	36 392	20.7	38 301	18.5	2 455	19.8
Deprived	76 522	19.3	37 126	21.1	37 058	17.9	2 338	18.9
Very deprived	79 474	20.1	40 843	23.2	36 005	17.3	2 626	21.2
Missing	20 493	5.2	8 850	5.0	11 023	5.3	620	5.0
<b>Social deprivation 1987</b>								

Very privileged	83 442	21.1	40 281	22.9	40 752	19.6	2 409	19.5
Privileged	76 370	19.3	37 171	21.1	36 878	17.8	2 321	18.7
Moderately privileged	81 217	20.5	38 551	21.9	40 185	19.4	2 481	20.0
Deprived	72 255	18.2	29 106	16.5	40 753	19.6	2 396	19.4
Very deprived	62 341	15.7	22 250	12.6	37 936	18.3	2 155	17.4
Missing	20 493	5.2	8 850	5.0	11 023	5.3	620	5.0
<b>Age at BCG vaccination (years)</b>								
Unvaccinated	207 527	52.4	0	0.0	207 527	100	0	0
<1	154 667	39.0	154 667	87.8	0	0	0	0
≥1	21 542	5.4	21 542	12.2	0	0	0	0
Uncertain vaccination	12 382	3.1	0	0	0	0	12 382	100
<b>Number of BCG vaccination</b>								
0	207 527	52.4	0	0.0	207 527	100	0	0
1	165 493	41.8	165 493	93.9	0	0	0	0
≥2	10 716	2.7	10 716	6.1	0	0	0	0
Missing	12 382	3.1	0	0	0	0	12 382	100

BCG, Bacille Calmette-Guerin;

<sup>a</sup> For gestational ages between 21 and 44 weeks, defined as ‘very small’ if birth weight <5th percentile, ‘small’ if ≥5th and <10th percentile, ‘Normal’ if ≥10th and <90th percentile, and ‘large’ if >90th percentile. Based on birth weights measured for each live birth occurring between 1994 and 1996 in all of the Canadian provinces except Ontario (Kramer *et al.*, 2001).

<sup>b</sup> Determined using the second character of subjects’ residential postal code (0: rural, ≥1: urban).

Legend for Table 8.1: Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 8.2: Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident type 1 and 2 diabetes mellitus, as defined by our pharmaceutical definitions among subjects with drug insurance (N<sub>2</sub>=260,440), Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.**

Type of diabetes	N subjects with DM (N=5,582) <sup>a</sup>	Person-years (N=4 625 083)	Crude Hazard Ratio (95% CI)	Adjusted Hazard Ratio (95% CI)
Type 1	420			
<b>BCG vaccination status</b>				
Unvaccinated	230	2 464 309	1 (Ref)	1 (Ref)
Vaccinated	190	2 160 774	0.91 (0.75-1.11)	0.94 (0.77-1.15) <sup>b</sup>
<b>Age at BCG vaccination (years)</b>				
Unvaccinated	230	2 464 309	1 (Ref)	1 (Ref)
<1	169	1 898 389	0.93 (0.76-1.14)	0.96 (0.78-1.18) <sup>b</sup>
≥1	21	262 385	0.80 (0.50-1.27)	0.80 (0.50-1.28) <sup>b</sup>
Type 2	5 162			
<b>BCG vaccination status</b>				
Unvaccinated	2 893	2 464 309	1 (Ref)	1 (Ref)
Vaccinated	2 269	2 160 774	0.90 (0.85-0.95)	0.90 (0.85-0.95) <sup>c</sup>
<b>Age at BCG vaccination (years)</b>				
Unvaccinated	2 893	2 464 309	1 (Ref)	1 (Ref)
<1	1 979	1 898 389	0.90 (0.85-0.95)	0.90 (0.84-0.95) <sup>c</sup>
≥1	290	262 385	0.92 (0.82-1.05)	0.91 (0.80-1.03) <sup>c</sup>

DM, diabetes mellitus; CI, confidence interval; BCG, Bacille Calmette-Guerin;

<sup>a</sup> 4,292 subjects with unspecified DM were excluded from these analyses

<sup>b</sup> Models were adjusted for duration of gestation, maternal and paternal ages at childbirth, having older siblings, area of residence in 1987, birthweight for gestational age, maternal and paternal birthplaces.

<sup>c</sup> Models were adjusted for birthweight, material and social deprivation in 1987, maternal and paternal birthplaces.

Legend for Table 8.2: Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 8.3: Non-weighted prevalence and incidence of phenotypes of diabetes mellitus (DM), as defined by the Immunology of Diabetes Society and the Japanese Diabetes Society, among subjects with long-term drug insurance (N<sub>3</sub>=160,622), Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.**

Type of diabetes	N subjects with DM (N=6,097)	Person-years	Prevalence %	Incidence per 100,000	N subjects with DM (N=6,097)	Person-years	Prevalence %	Incidence per 100,000
			(95% CI)	(95% CI)			(95% CI)	(95% CI)
Immunology of Diabetes Society definition <sup>a</sup>					Japanese Diabetes Society definition			
T1D	252	2 850 602	0.16 (0.14-0.18)	8.8 (7.7-9.9)	247	2 850 602	0.15 (0.14-0.17)	8.7 (7.6-9.7)
T2D	3 030		1.89 (1.82-1.95)	106.3 (102.5-110.1)	3 009		1.87 (1.81-1.94)	105.6 (101.8-109.3)
Unspecified <sup>b</sup>	2 766		1.72 (1.66-1.79)	97.0 (93.4-100.6)	2 766		1.72 (1.66-1.79)	97.0 (93.4-100.6)
LADA	49		0.03 (0.02-0.04)	1.7 (1.2-2.2)	75		0.05 (0.04-0.06)	2.6 (2.0-3.2)
Total DM	6 097		3.80 (3.70-3.89)	213.9 (208.5-219.2)	6 097		3.80 (3.70-3.89)	213.9 (208.5-219.2)

CI, confidence interval; DM, diabetes mellitus; LADA, latent autoimmune diabetes in adults; T1D, type 1 diabetes; T2D, type 2 diabetes

<sup>a</sup> Without age restriction

<sup>b</sup> Unknown clinical form

Legend for Table 8.3: Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 8.4: Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident type 1, type 2 diabetes mellitus (DM) and LADA, as defined by the Immunology of Diabetes Society and the Japanese Diabetes Society, among subjects with long-term drug insurance (N<sub>3</sub>=160,622), Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.**

Type of diabetes	N subjects with DM		Crude Hazard Ratio (95% CI)	Adjusted Hazard Ratio (95% CI)	N subjects with DM		Crude Hazard Ratio (95% CI)	Adjusted Hazard Ratio (95% CI)
	(N=3,331) <sup>a</sup>				(N=3,331) <sup>a</sup>			
	Immunology of Diabetes Society definition <sup>b</sup>				Japanese Diabetes Society definition			
<b>T1D ≤30 years</b>	127				127			
	<b>BCG vaccination status</b>							
Unvaccinated	65	1 526 830	1 (Ref)	1 (Ref)	65	1 526 830	1 (Ref)	1 (Ref)
Vaccinated	62	1 323 991	1.08 (0.76-1.53)	1.06 (0.74-1.52) <sup>c</sup>	62	1 323 991	1.08 (0.76-1.53)	1.06 (0.74-1.52) <sup>c</sup>
	<b>Age at BCG vaccination (years)</b>							
Unvaccinated	65	1 527 591	1 (Ref)	1 (Ref)	65	1 527 591	1 (Ref)	1 (Ref)
<1	54	1 163 286	1.06 (0.73-1.52)	1.04 (0.71-1.51) <sup>c</sup>	54	1 163 286	1.06 (0.73-1.52)	1.04 (0.71-1.51) <sup>c</sup>
≥1	8	159 725	1.25 (0.60-2.60)	1.21 (0.58-2.53) <sup>c</sup>	8	159 725	1.25 (0.60-2.60)	1.21 (0.58-2.53) <sup>c</sup>
<b>T1D &gt;30 years</b>	125				120			
	<b>BCG vaccination status</b>							
Unvaccinated	78	1 526 830	1 (Ref)	1 (Ref)	73	1 526 830	1 (Ref)	1 (Ref)
Vaccinated	47	1 323 991	0.66 (0.46-0.96)	0.65 (0.44-0.95) <sup>d</sup>	47	1 323 991	0.71 (0.49-1.03)	0.69 (0.47-1.01) <sup>d</sup>
	<b>Age at BCG vaccination (years)</b>							
Unvaccinated	78	1 527 591	1 (Ref)	1 (Ref)	73	1 527 591	1 (Ref)	1 (Ref)
<1	41	1 163 286	0.67 (0.46-0.99)	0.66 (0.45-0.98) <sup>d</sup>	41	1 163 286	0.72 (0.49-1.06)	0.70 (0.47-1.05) <sup>d</sup>

	≥1	6	159 725	0.58 (0.23-1.47)	0.56 (0.22-1.42) <sup>d</sup>	6	159 725	0.62 (0.24-1.58)	0.59 (0.23-1.51) <sup>d</sup>
<b>T2D</b>		3030				3009			
	<b>BCG vaccination status</b>								
	Unvaccinated	1 744	1 527 591	1 (Ref)	1 (Ref)	1 737	1 527 591	1 (Ref)	1 (Ref)
	Vaccinated	1 286	1 323 011	0.85 (0.79-0.91)	0.85 (0.79-0.92) <sup>e</sup>	1 272	1 323 011	0.84 (0.78-0.91)	0.84 (0.78-0.91) <sup>e</sup>
	<b>Age at BCG vaccination (years)</b>								
	Unvaccinated	1 744	1 527 591	1 (Ref)	1 (Ref)	1 737	1 527 591	1 (Ref)	1 (Ref)
	<1	1 128	1 163 286	0.85 (0.79-0.92)	0.85 (0.78-0.92) <sup>e</sup>	1 118	1 163 286	0.85 (0.78-0.91)	0.85 (0.78-0.92) <sup>e</sup>
	≥1	158	159 725	0.84 (0.71-1.00)	0.83 (0.70-0.98) <sup>e</sup>	154	159 725	0.83 (0.70-0.98)	0.81 (0.68-0.96) <sup>e</sup>
<b>LADA</b>		49				75			
	<b>BCG vaccination status</b>								
	Unvaccinated	25	1 527 591	1 (Ref)	1 (Ref)	37	1 527 591	1 (Ref)	1 (Ref)
	Vaccinated	24	1 323 011	1.11 (0.62-1.96)	1.30 (0.71-2.38) <sup>e</sup>	38	1 323 011	1.17 (0.74-1.86)	1.40 (0.86-2.29) <sup>e</sup>
	<b>Age at BCG vaccination (years)</b>								
	Unvaccinated	25	1 527 591	1 (Ref)	1 (Ref)	37	1 527 591	1 (Ref)	1 (Ref)
	<1	24 <sup>f</sup>	1 163 286	1.11 (0.62-2.01)	1.33 (0.71-2.48) <sup>b</sup>	31	1 163 286	1.10 (0.68-1.78)	1.33 (0.80-2.22) <sup>e</sup>
	≥1		159 725	1.05 (0.30-3.75)	1.14 (0.32-4.08) <sup>b</sup>	7	159 725	1.70 (0.74-3.92)	1.85 (0.79-4.32) <sup>e</sup>

BCG, Bacille Calmette-Guerin; CI, confidence interval; DM, diabetes mellitus; LADA, latent autoimmune diabetes in adults.; T1D, type 1 diabetes; T2D, type 2 diabetes

<sup>a</sup> 2,766 subjects with unspecified DM were excluded from these analyses

<sup>b</sup> Without age restriction

<sup>c</sup> Models were adjusted for duration of gestation, maternal and paternal ages at childbirth, having older siblings, area of residence in 1987, birthweight for gestational age, maternal and paternal birthplaces.

<sup>d</sup> Models were adjusted like (c) without area of residence in 1987, because the specific and global HR were not proportional with this variable.

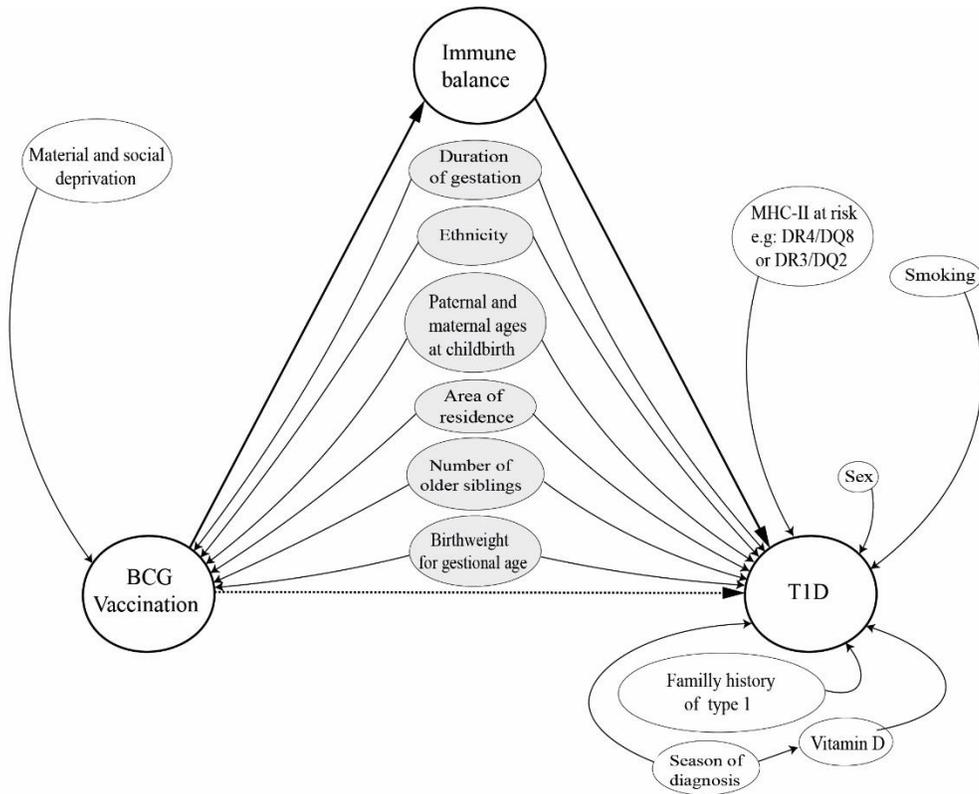
<sup>e</sup> Models were adjusted for birthweight, material and social deprivation in 1987, maternal and paternal birthplaces.

<sup>f</sup> One of the two categories had a total number of subjects  $\leq 5$ . Both categories were grouped to respect the disclosure standards of ISQ.

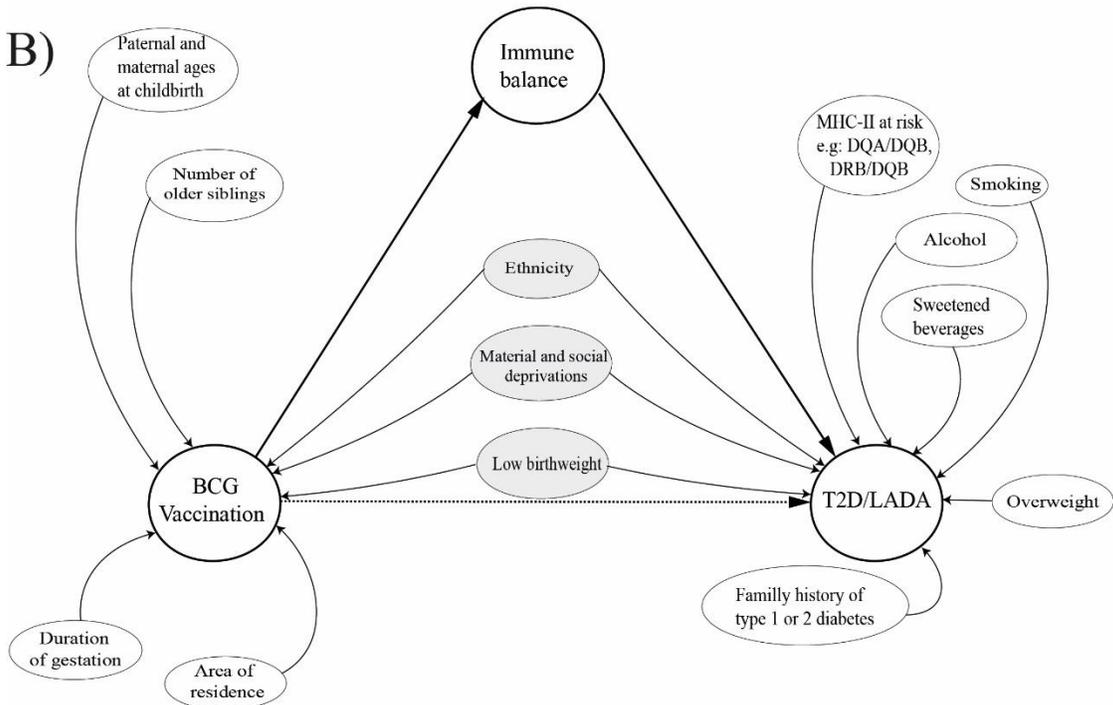
Legend for Table 8.4: Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

## 8.8 Supplementary File

A)



B)



**Figure 8.A1:** Directed acyclic graphs of the association between BCG vaccination and diabetes: A) type 1 diabetes, B) type 2 diabetes and LADA.

**Legend for Figure 8.A1:** Grey shading indicates confounding factors for the relationship between BCG vaccination and type 1 diabetes or type 2 diabetes and LADA.

**Table 8.A1:** Insulin therapies used in our definitions for identifying insulin agents users, authorized in Quebec between 1997 and 2014, and included in the public drug insurance database.

Type of insulin	Description	Codes of International Nonproprietary Name
<b>Insulins from animal sources</b>		
	Cristalline zinc (pork)	18296, 47004
	Cristalline zinc (beef)	39523, 43735
	Cristalline zinc (beef and pork)	39185, 46536
	Zinc protamin (pork)	18322, 39497, 39497
	Zinc protamin (beef)	18309, 39484
	Zinc protamin (beef and pork)	39146
	Isophan (pork)	18335
	Isophan (beef)	18348, 39458
	Isophan (beef and pork)	39133, 46537
	Long-acting (beef and pork)	39120, 46538
	Semi long-acting (beef and pork)	39159
	Ultra long-acting (beef and pork)	39172
	Long-acting (pork)	41655
	Cristalline zinc (pork)/ Isophane (pork)	43033
<b>Rapid- or long-acting insulin analogs*</b>		
	Globin zinc	04823
	Sulfated insulin	04888
	Isophan semi-synthetic analogue insulin	44151
	Isophan bio-synthetic analogue insulin	44164
	Long-acting semi-synthetic analogue insulin	44476
	Cristalline zinc bio-synthetic analogue insulin	44489
	Cristalline zinc semi-synthetic analogue insulin	44502
	Ultra long-acting semi-synthetic analogue insulin	44996
	Long-acting bio-synthetic analogue insulin	45405
	Long-acting bio-synthetic analogue insulin	45415
	Ultra long-acting bio-synthetic analogue insulin	45483
	Isophan and Cristalline zinc bio-synthetic analogue insulin	45511
	Cristalline zinc and Isophane bio-synthetic analogue insulin	45531
	Cristalline zinc and Isophane semi-synthetic analogue insulin	45534
	Lispro insulin	46322
	Isophane analogue insulin/ injectable analogue insulin	46592
	Isophan analogue insulin	46602
	Injectable analogue insulin	46603
	Lispro/Isophane analogues insulins	46607

\* Aspart, Lispro, Glargine, Detemir, Glulisine were not present in the database between 1997 and 2014

**Legend for table 8.A1:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 8.A2:** Oral anti-diabetic agents (OAD) used in our definitions for identifying OAD users, authorized in Quebec between 1997 and 2014, and included in the public drug insurance database.

Therapeutic class	International Non-proprietary Name	Codes of International Nonproprietary Name
<i>Alpha-glucosidase Inhibitors</i>		
	Acarbose	46300; 47151
<i>Biguanides</i>		
	Metformin	05824
<i>Meglitinides</i>		
	Repaglinide	47357; 46568
<i>Thiazolidinediones</i>		
	Pioglitazone	46678
	Rosiglitazone	46642; 47371
<i>Sulfonylureas</i>		
	Gliclazide	47329
	Glyburide	04264
	Tolbutamide	09672

**Legend for table 8.A2:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 8.A3** : Factors considered in the directed acyclic graph of an association between BCG vaccination and the onset of type 1 diabetes, and supporting references

Variable	Causal association with BCG	References	Causal association with type 1 diabetes	References
Season of diagnosis	No	-	Yes	(208, 389, 390)
Material and social deprivation	Yes	Income (307); education (368-370); socioeconomic status (369-372)	No	-
Birth weight for gestational age	Yes	(307)	Yes	(235, 236)
Duration of gestation	Yes	Post-term (307)	Yes	Preterm birth (237, 238)
Number of older siblings	Yes	(307)	Yes	(267)
Area of residence	Yes	(307)	Yes	(269, 398)
Maternal age at childbirth	Yes	(307)	Yes	(241, 268)
Paternal age at childbirth	Yes	(307)	Yes	(241, 268)
Passive smoking	No	-	Yes	(399)
Familial history of type 1 diabetes	No	-	Yes	(267)
Ethnicity	Yes	(307)	Yes	(391)
Sex	No	(282, 307)	Yes	(69, 207, 208)
Lack of Vitamin D	No	-	Yes	(389)
MHC-II at risk: DR4/DQ8 or DR3/DQ2	No	-	Yes	(392, 400)

**Legend for table 8.A3:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 8.A4** : Factors considered in the directed acyclic graph of an association between BCG vaccination and the onset of type 2 diabetes (T2D) and latent autoimmune diabetes in adults (LADA), and supporting references

Variable	Causal association with BCG	References	Causal association with T2D/LADA	References
Material and social deprivation	Yes	Income (307); education (368-370); socioeconomic status (369-372)	Yes	Education (27, 401); socioeconomic status (27, 401)
Low Birth weight	Yes	Birth weight for gestational age (307)	Yes	(27)
Ethnicity	Yes	(307)	Yes	(402)
Alcohol	No	-	Yes	(260)
Overweight	No	-	Yes	(403)
Consumption of sweetened beverages	No	-	Yes	(27)
Smoking	No	-	Yes	(27)
Familial history of type 1 and type 2 of diabetes	No	-	Yes	(215)
Sex	No	(282, 307)	No	-
MHC-II lowers or increase risk of LADA: (e.g: DR2, DR3, DR4, DQA/DQB, DRB/DQB)	No	-	Yes	(6, 404)

**Legend for table 8.A4:** Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

**Table 8.A5:** Age at diagnosis for the different phenotypes of diabetes mellitus (DM), as defined by the Immunology of Diabetes Society and the Japanese Diabetes Society, among subjects with long-term drug insurance (N<sub>3</sub>= 160,286), Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

Type of diabetes	Immunology of Diabetes Society definition <sup>a</sup>			Japanese Diabetes Society definition		
	n	Mean age (years)	Standard deviation	n	Mean age (years)	Standard deviation
<b>T1D</b>	252	31.1	5.4	247	31	5.3
<b>T1D≤30 ans</b>	127	26.7	2.0	127	26.7	2.0
<b>T1D&gt;30 ans</b>	125	35.6	3.7	120	35.5	3.7
<b>T2D</b>	3 030	36.0	4.9	3 009	36.0	4.9
<b>Unspecified<sup>b</sup></b>	2 766	36.7	4.6	2 766	36.7	4.6
<b>LADA</b>	49	33.4	5.3	75	33.7	5.2

<sup>a</sup> Without age restriction

<sup>b</sup> Unknown clinical form

**Table 8.A6:** Hazard ratios and 95% confidence intervals for the association between BCG vaccination and incident type 2 diabetes mellitus (DM), unspecified DM and both subgroups combined, among subjects with long-term drug insurance (N<sub>3</sub>=160,622), Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

Type of diabetes	N subjects with DM (N=6097)	Person-years (=2 850 602)	Crude Hazard Ratio (95% CI)	Adjusted Hazard Ratio <sup>b</sup> (95% CI)	N subjects with DM (N=6,097)	Person-years (=2 850 602)	Crude Hazard Ratio (95% CI)	Adjusted Hazard Ratio <sup>b</sup> (95% CI)
<b>Immunology of Diabetes Society definition<sup>a</sup></b>					<b>Japanese Diabetes Society definition</b>			
T2D	3 030				3 009			
<b>BCG vaccination status</b>								
Unvaccinated	1 744	1 527 591	1 (Ref)	1 (Ref)	1 737	1 527 591	1 (Ref)	1 (Ref)
Vaccinated	1 286	1 323 011	0.85 (0.79-0.91)	0.85 (0.79-0.92)	1 272	1 323 011	0.84 (0.78-0.91)	0.84 (0.78-0.91)
<b>Age at BCG vaccination (years)</b>								
Unvaccinated	1 744	1 527 591	1 (Ref)	1 (Ref)	1 737	1 527 591	1 (Ref)	1 (Ref)
<1	1 128	1 163 286	0.85 (0.79-0.92)	0.85 (0.78-0.92)	1 118	1 163 286	0.85 (0.78-0.91)	0.85 (0.78-0.92)
≥1	158	159 725	0.84 (0.71-1.00)	0.83 (0.70-0.98)	154	159 725	0.83 (0.70-0.98)	0.81 (0.68-0.96)
Unspecified <sup>c</sup> DM	2 766				2 766			
<b>BCG vaccination status</b>								
Unvaccinated	1 475	1 527 591	1 (Ref)	1 (Ref)	1 475	1 527 591	1 (Ref)	1 (Ref)
Vaccinated	1 291	1 323 011	1.00 (0.93-1.08)	0.98 (0.91-1.06)	1 291	1 323 011	1.00 (0.93-1.08)	0.98 (0.91-1.06)
<b>Age at BCG vaccination (years)</b>								
Unvaccinated	1 475	1 527 591	1 (Ref)	1 (Ref)	1 475	1 527 591	1 (Ref)	1 (Ref)
<1	1 111	1 163 286	0.99 (0.93-1.05)	0.97 (0.90-1.06)	1 111	1 163 286	0.99 (0.93-1.05)	0.97 (0.90-1.06)
≥1	180	159 725	1.09 (0.93-1.28)	1.05 (0.89-1.24)	180	159 725	1.09 (0.93-1.28)	1.05 (0.89-1.24)
Both subgroups combined	5 796				5 775			
<b>BCG vaccination status</b>								
Unvaccinated	3 219	1 527 591	1 (Ref)	1 (Ref)	3 212	1 527 591	1 (Ref)	1 (Ref)

Vaccinated	2 577	1 323 011	0.92 (0.87-0.97)	0.91 (0.86-0.96)	2 563	1 323 011	0.92 (0.87-0.97)	0.91 (0.86-0.96)
<b>Age at BCG vaccination (years)</b>								
Unvaccinated	3 219	1 527 591	1 (Ref)	1 (Ref)	3 212	1 527 591	1 (Ref)	1 (Ref)
<1	2 239	1 163 286	0.91 (0.87-0.97)	0.91 (0.86-0.96)	2 229	1 163 286	0.91 (0.86-0.96)	0.90 (0.85-0.96)
≥1	338	159 725	0.96 (0.85-1.08)	0.93 (0.83-1.05)	334	159 725	0.95 (0.84-1.07)	0.92 (0.82-1.04)

BCG, Bacille Calmette-Guerin; CI, confidence interval; DM, diabetes mellitus.

<sup>a</sup> Without age restriction

<sup>b</sup> Models were adjusted for birthweight, material and social deprivation in 1987, maternal and paternal birthplaces.

<sup>c</sup> Unknown phenotype

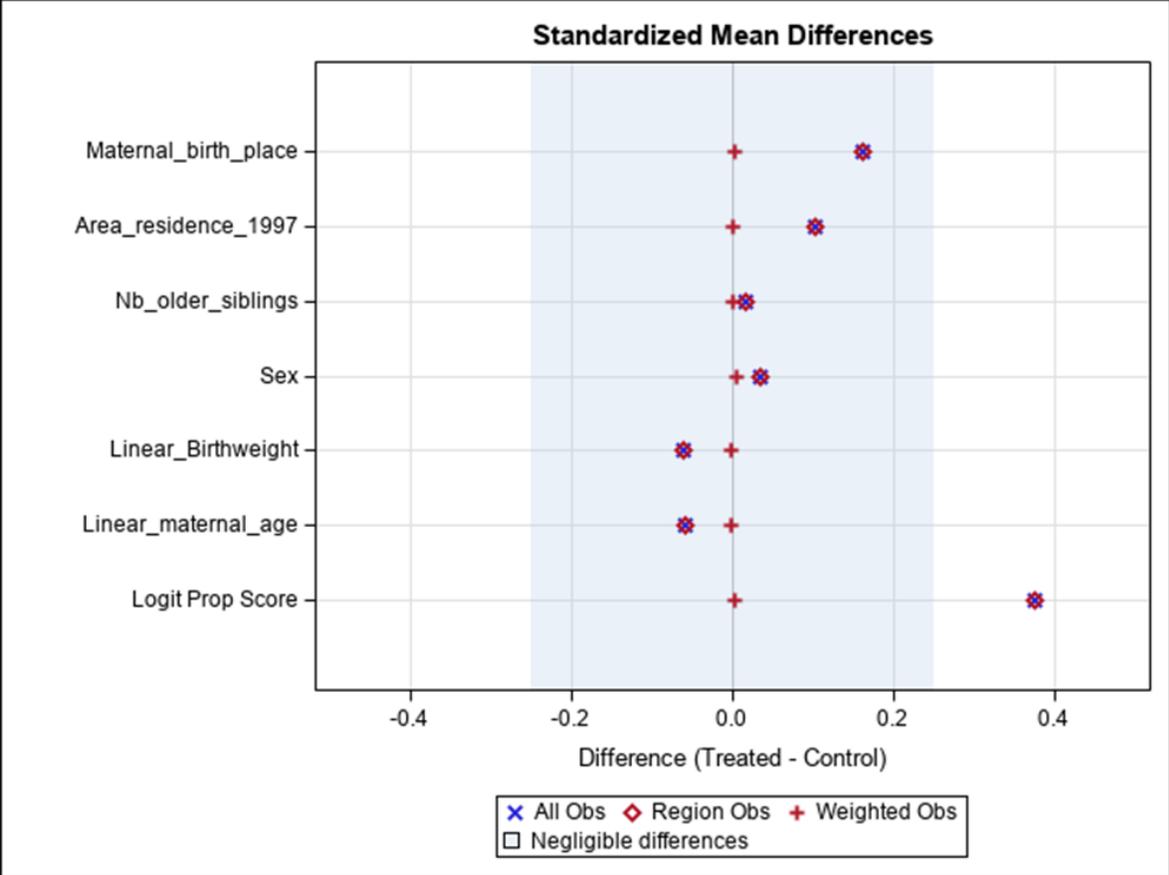


Figure 8.A2: Standardized mean difference of continuous and binary variables used in the propensity score to calculate weights between the analytical sample 2 and the eligible sample, from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

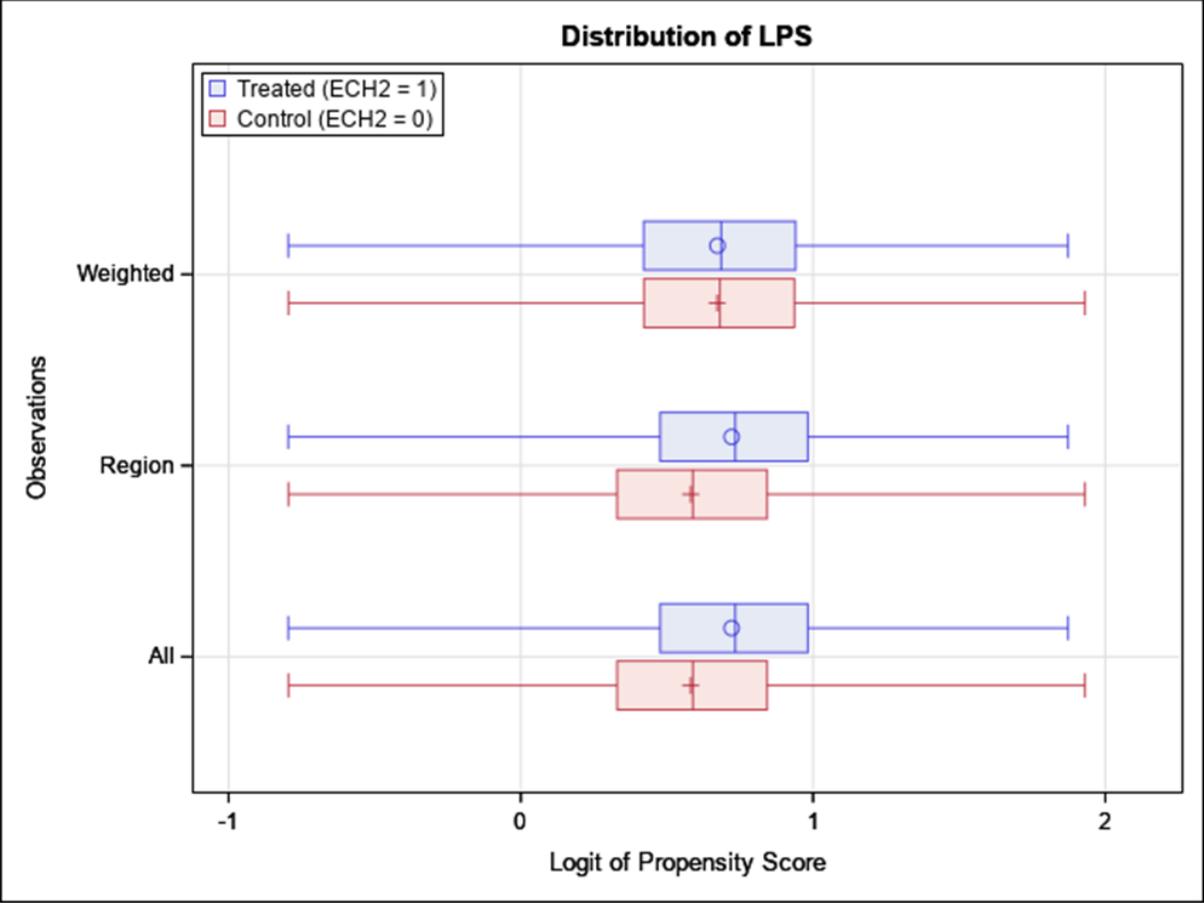


Figure 8.A3: Distribution of propensity score logit calculated between the analytical samples 2 and the eligible sample, from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

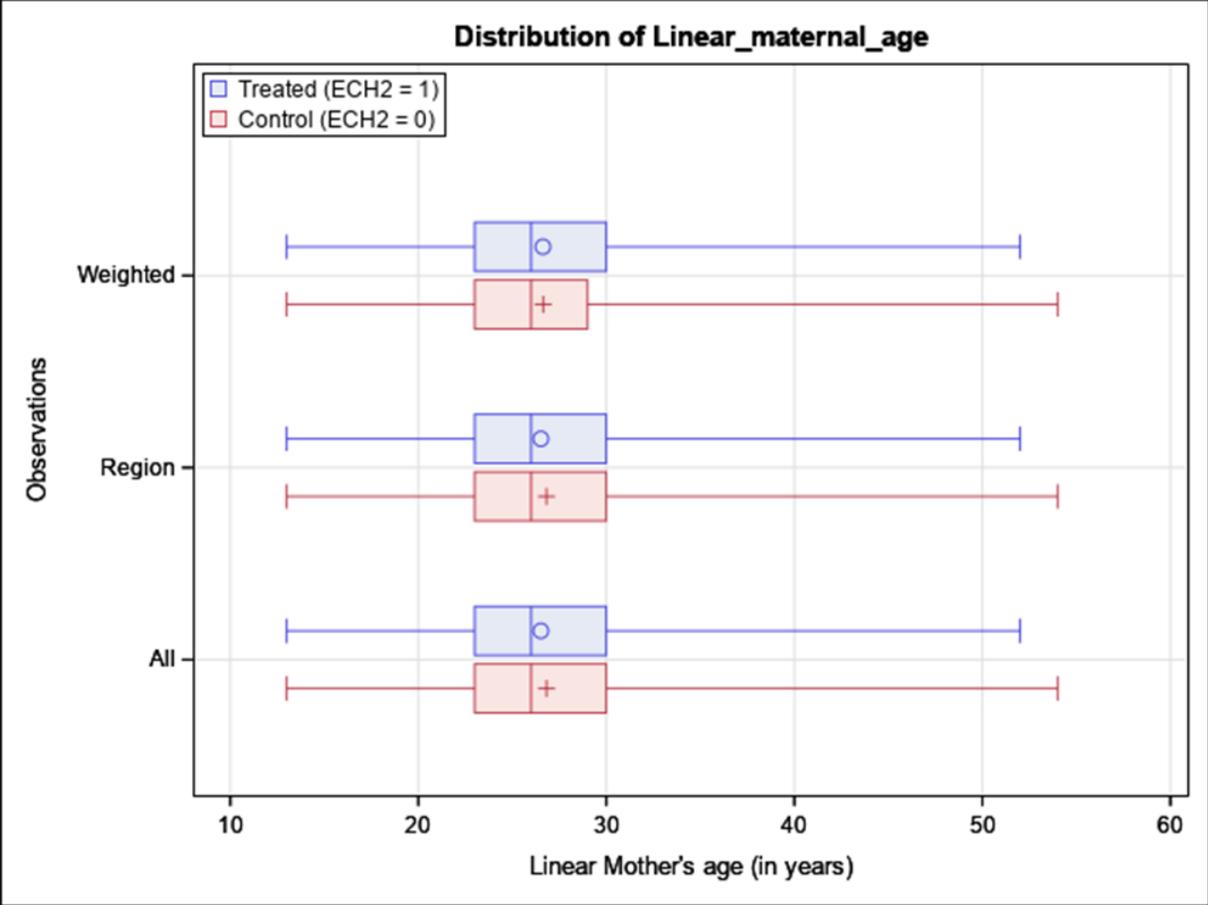


Figure 8.A4: Boxplot of continuous maternal age before (region) and after (weighted) applying weights between the analytical samples 2 and 1 from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

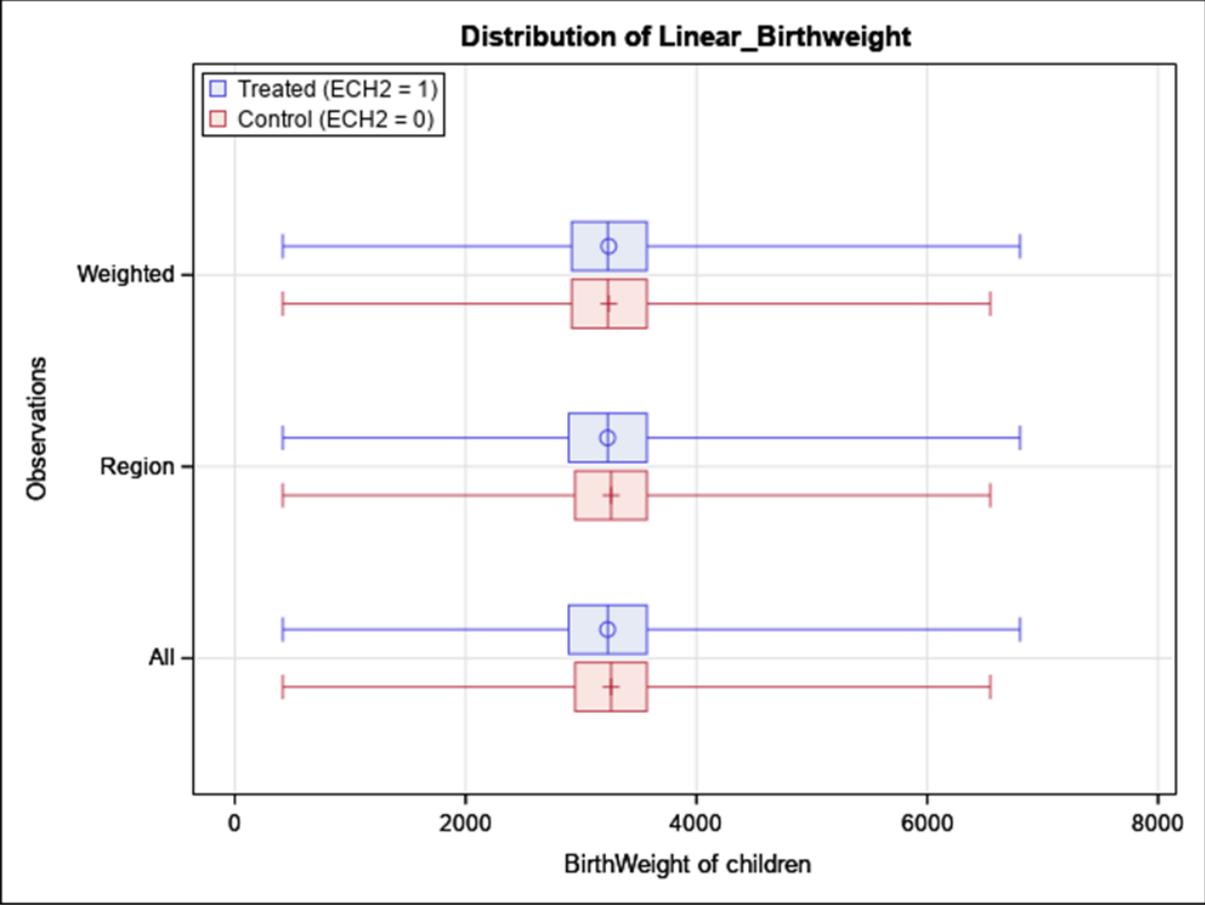


Figure 8.A5: Boxplot of continuous birthweight before (region) and after (weighted) applying weights between the analytical sample 2 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

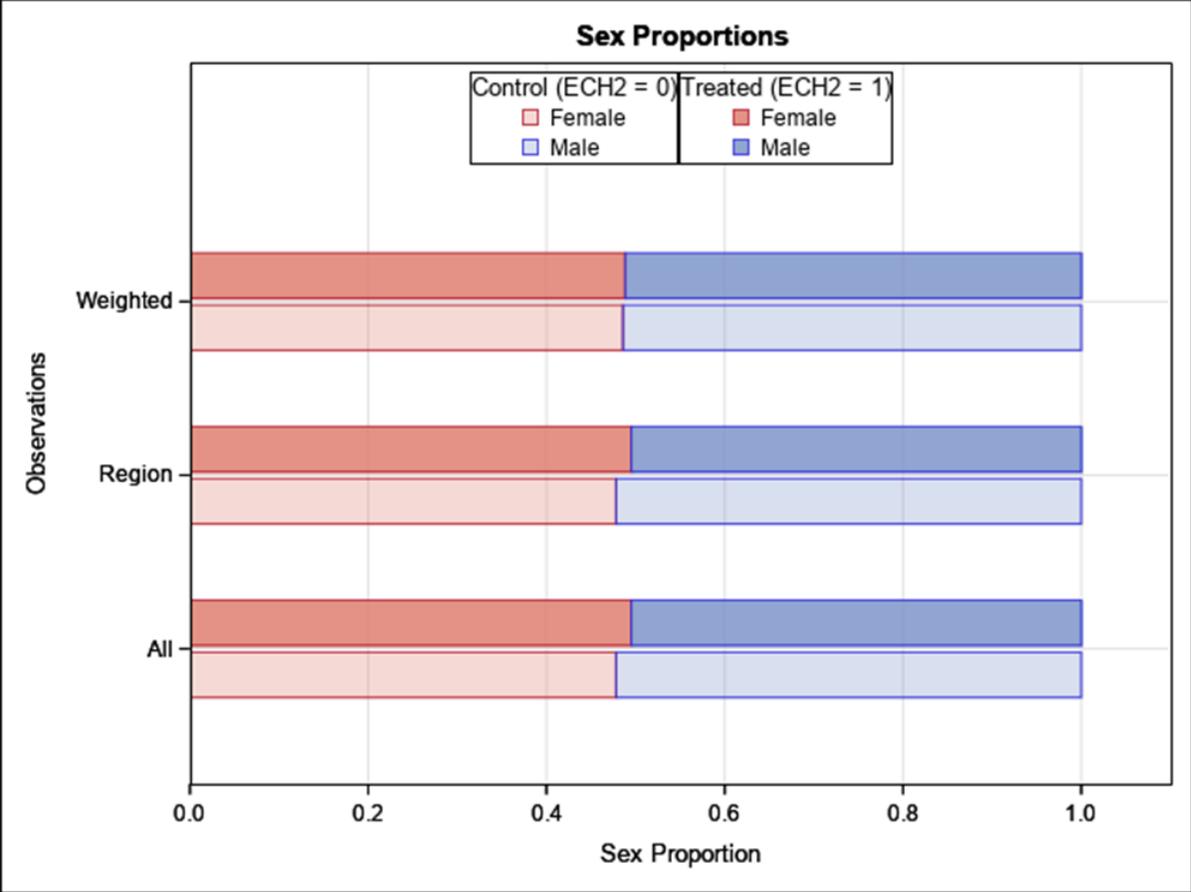


Figure 8.A6: Chart of sex proportions before (region) and after (weighted) applying weights between the analytical samples 2 and the eligible sample, from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

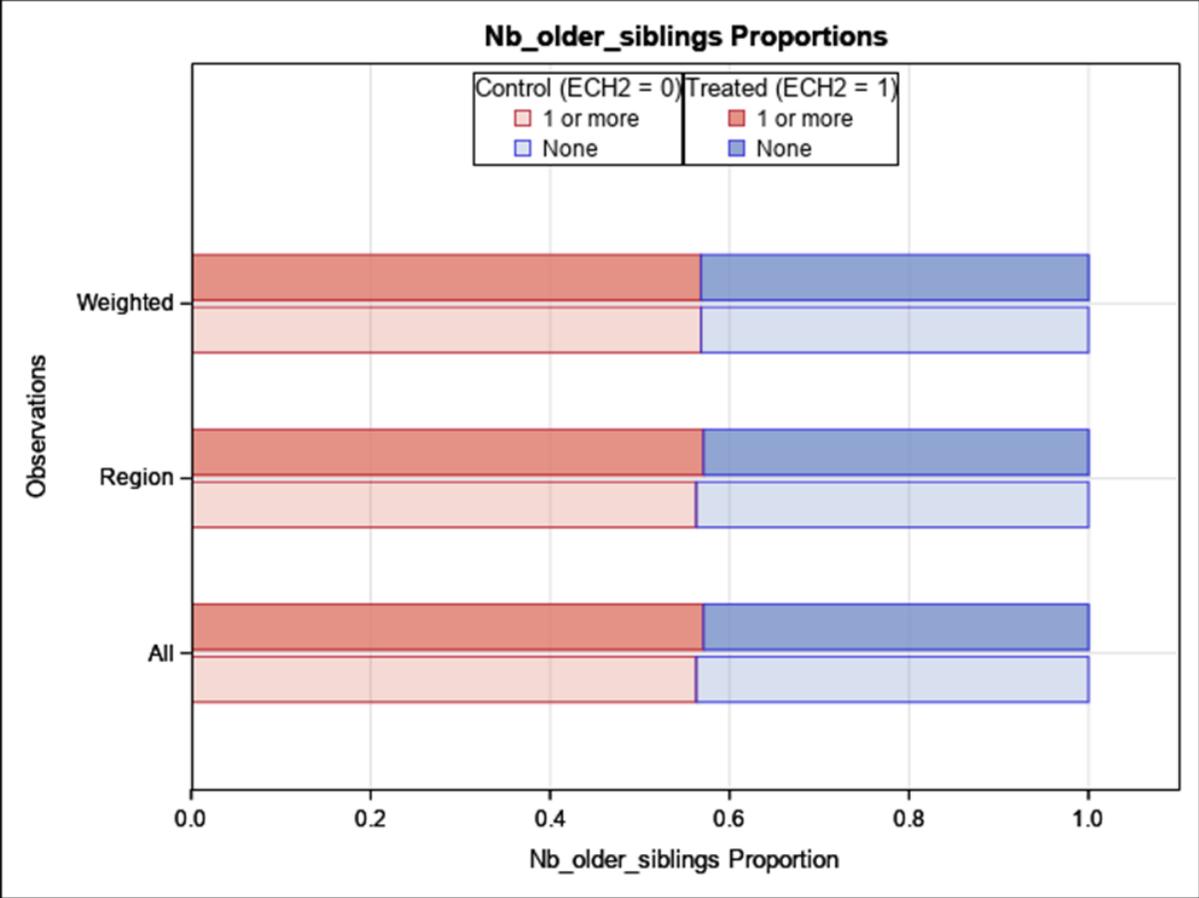


Figure 8.A7: Chart of number of older siblings proportions before (region) and after (weighted) applying weights between the analytical sample 2 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

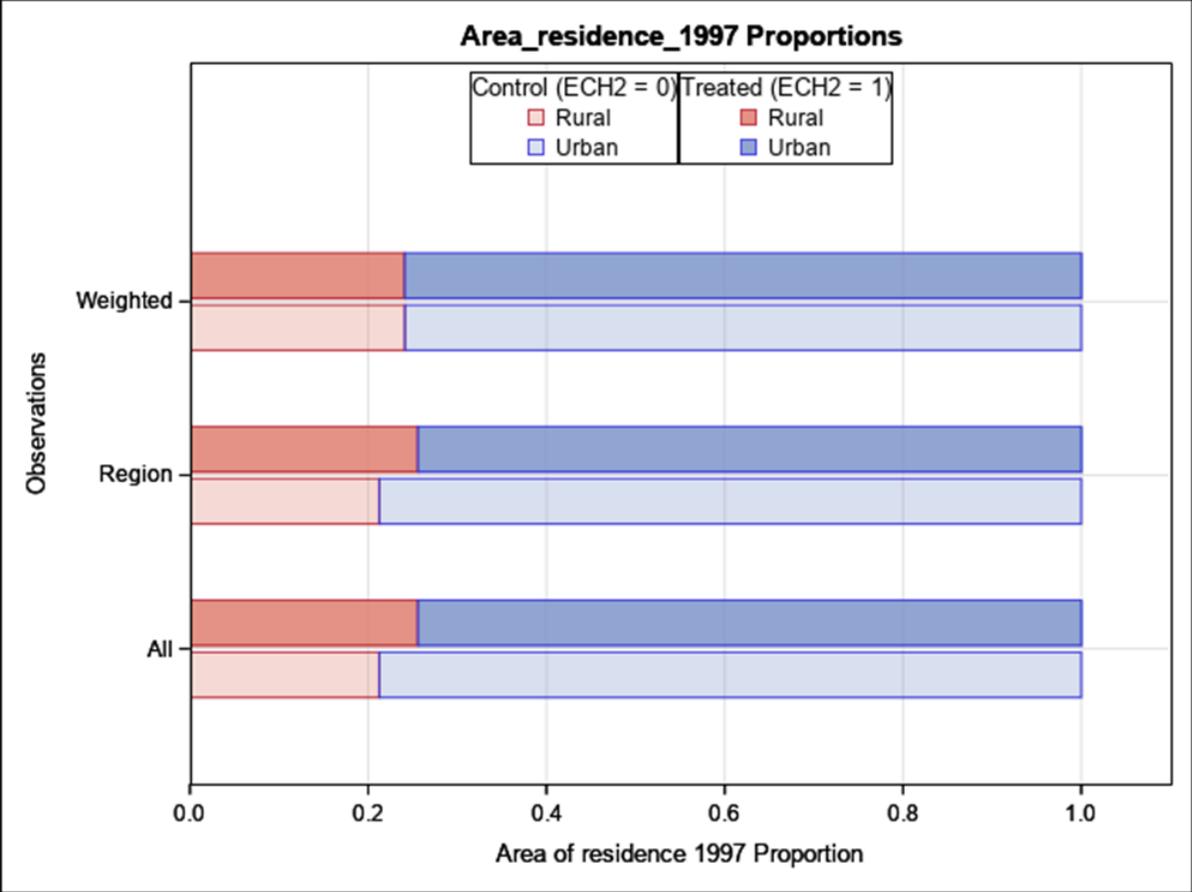


Figure 8.A8: Chart of area of residence in 1997 proportions before (region) and after (weighted) applying weights between the analytical sample 2 and the eligible sample, from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

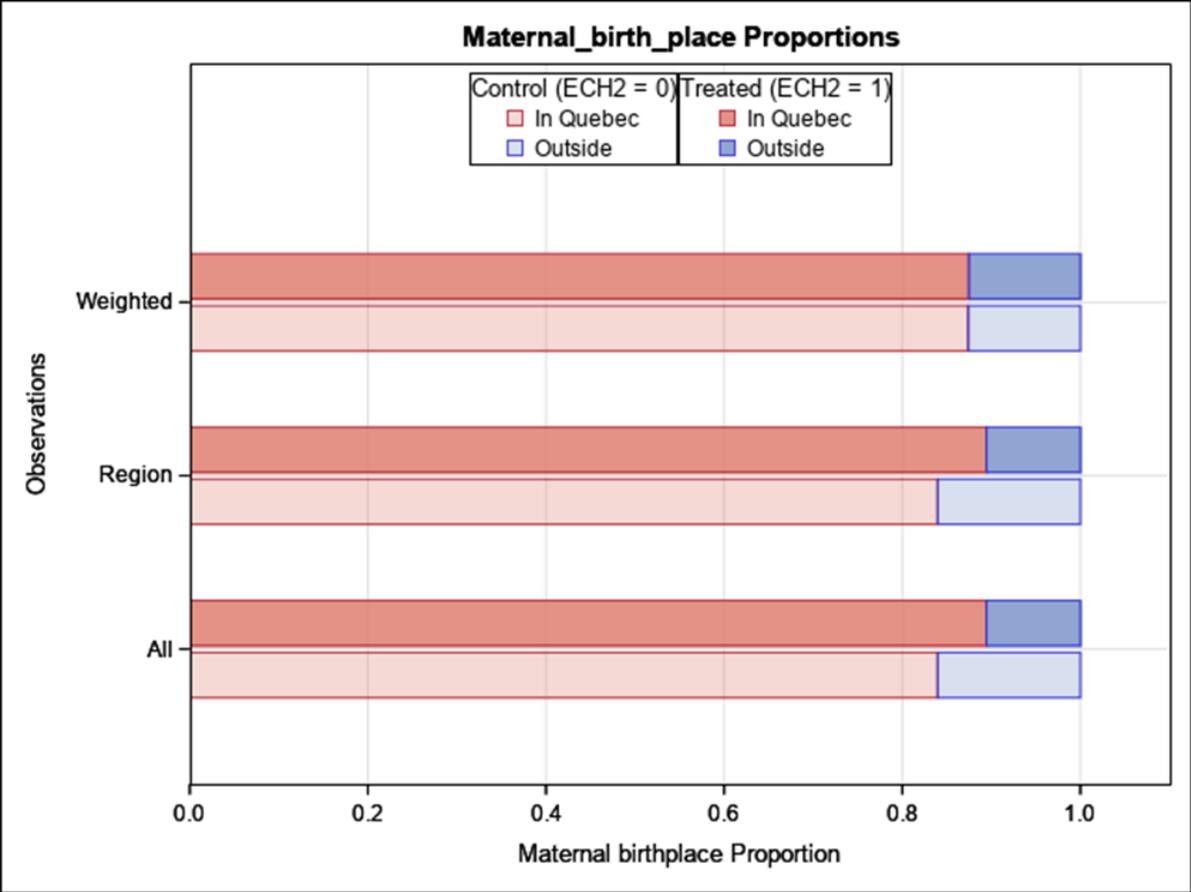


Figure 8.A9: Chart of maternal birthplace proportions before (region) and after (weighted) applying weights between the analytical samples 2 and the eligible sample, from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

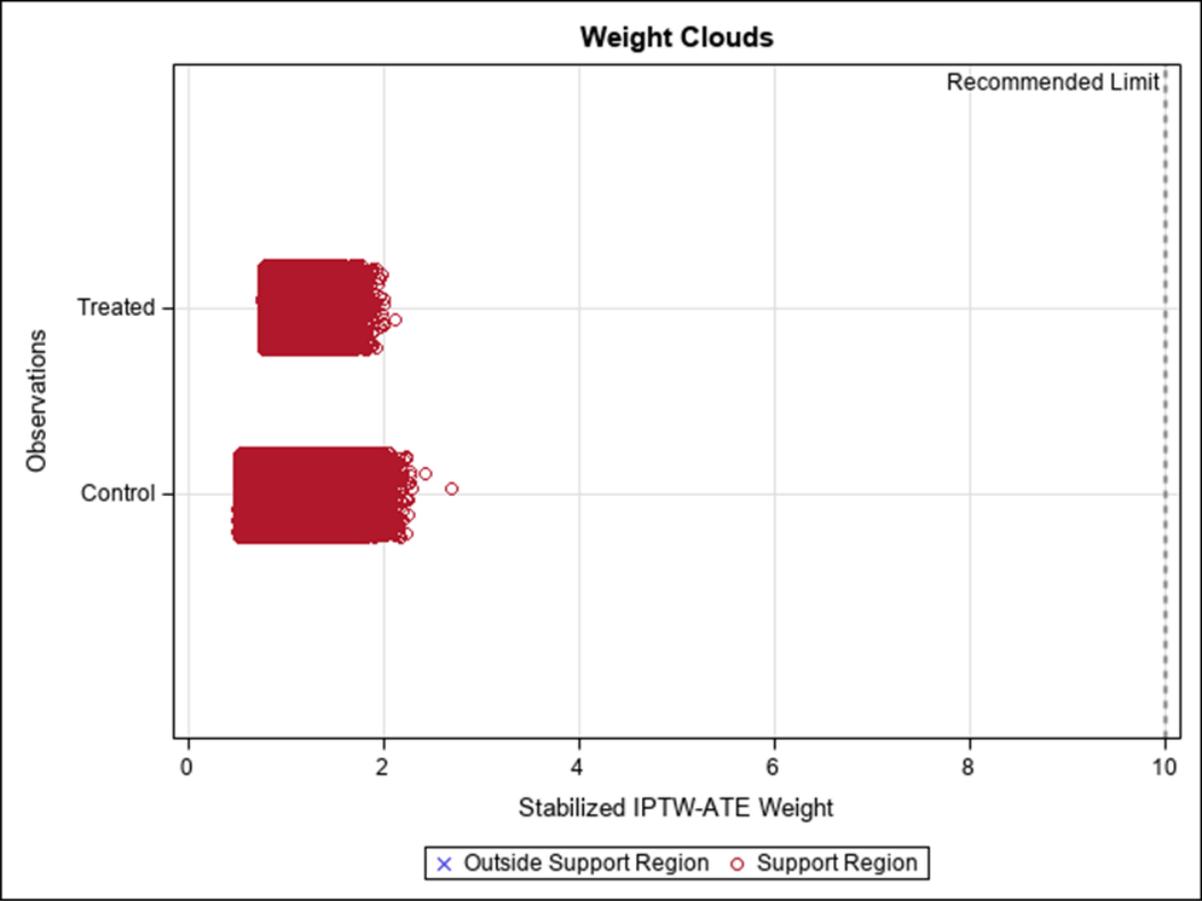


Figure 8.A10: Weight cloud of stabilized inverse probability weightings within the drug insurance coverage (treated) and non drug insurance coverage (control) subsets, from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

Legend for all figures: Ordinal categorical variables (material and social deprivation 1997 as well as income in 1996) requiring dummy variables for presentation in charts were dropped to facilitate the presentation of deviances and weights.

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

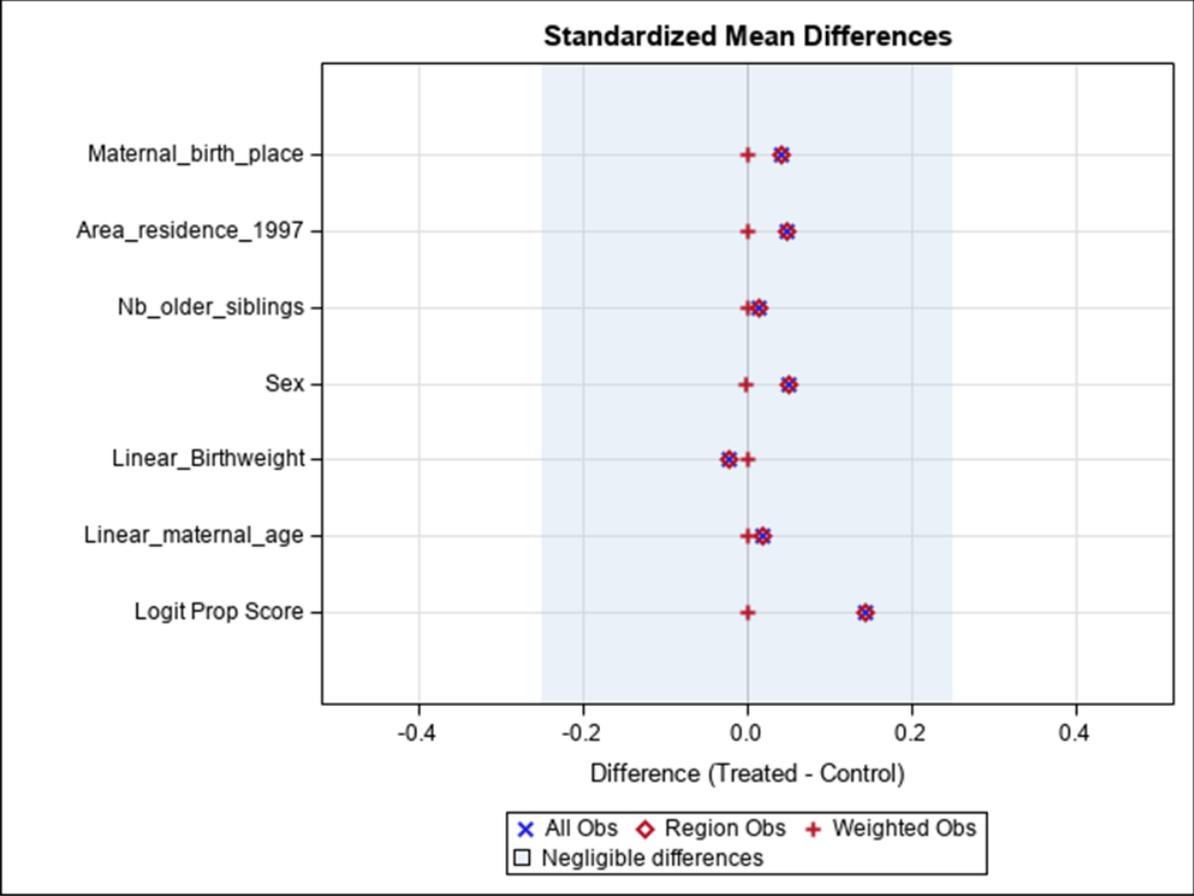


Figure 8.A11: Standardized mean difference of continuous and binary variables used in the propensity score to calculate weights between the analytical sample 3 and the eligible sample, from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

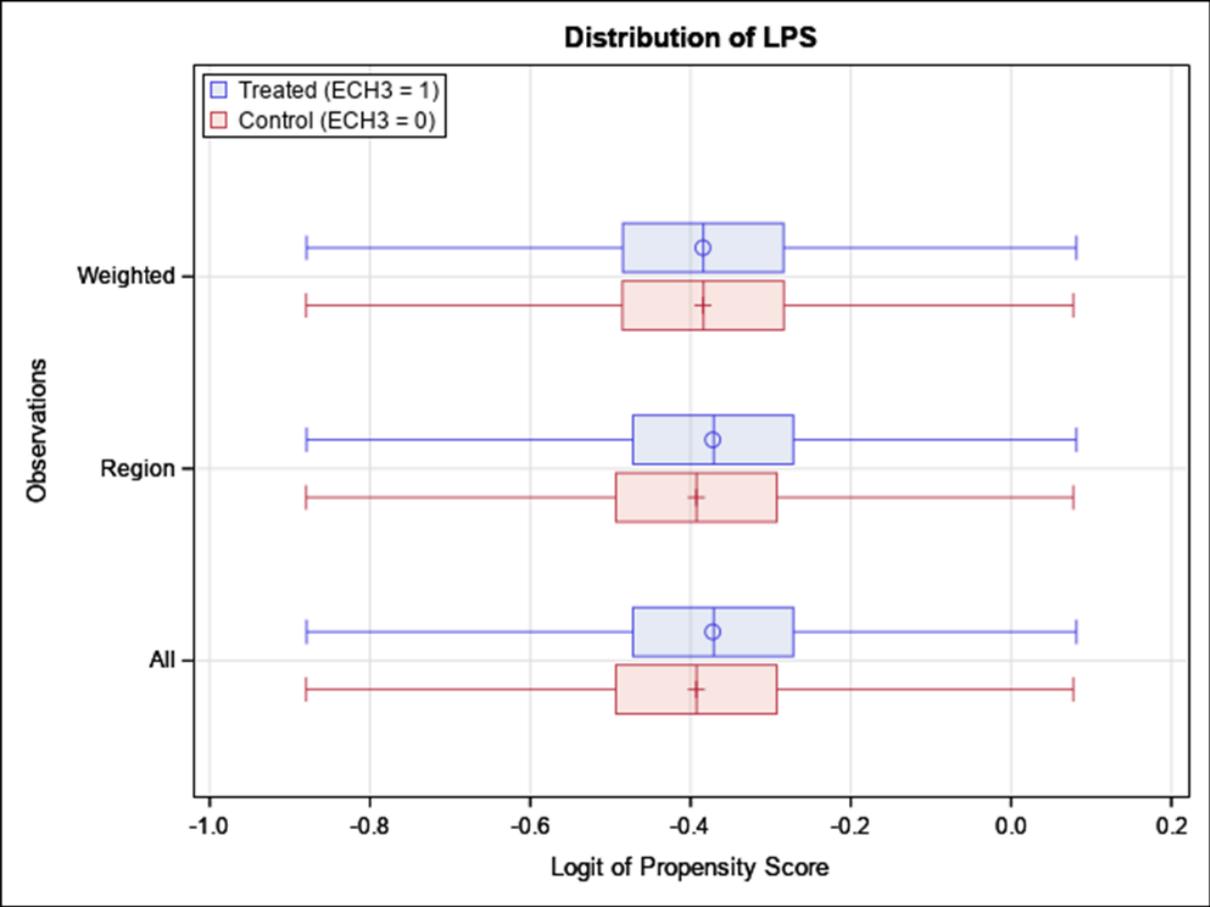


Figure 8.A12: Distribution of propensity score logit calculated between the analytical sample 3 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

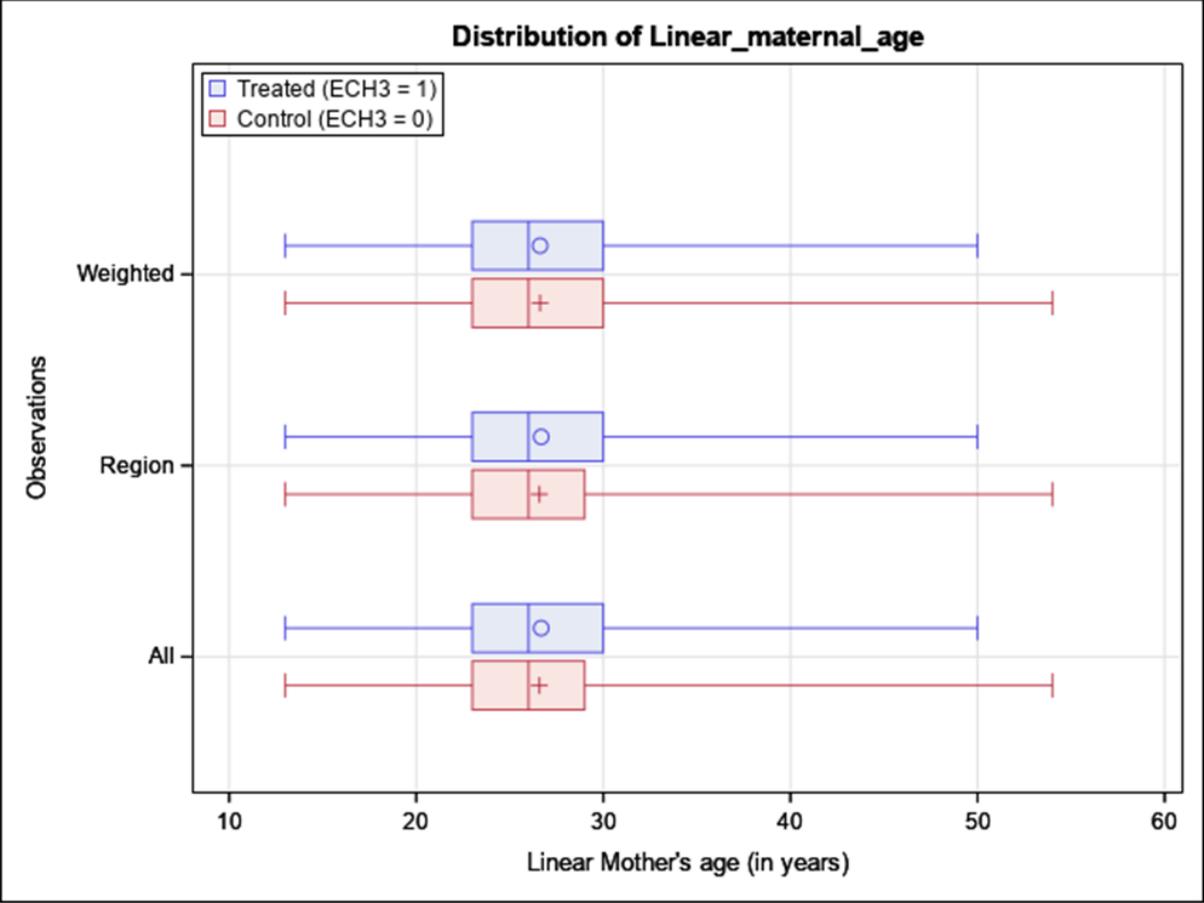


Figure 8.A13: Boxplot of continuous maternal age before (region) and after (weighted) applying weights between the analytical sample 3 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

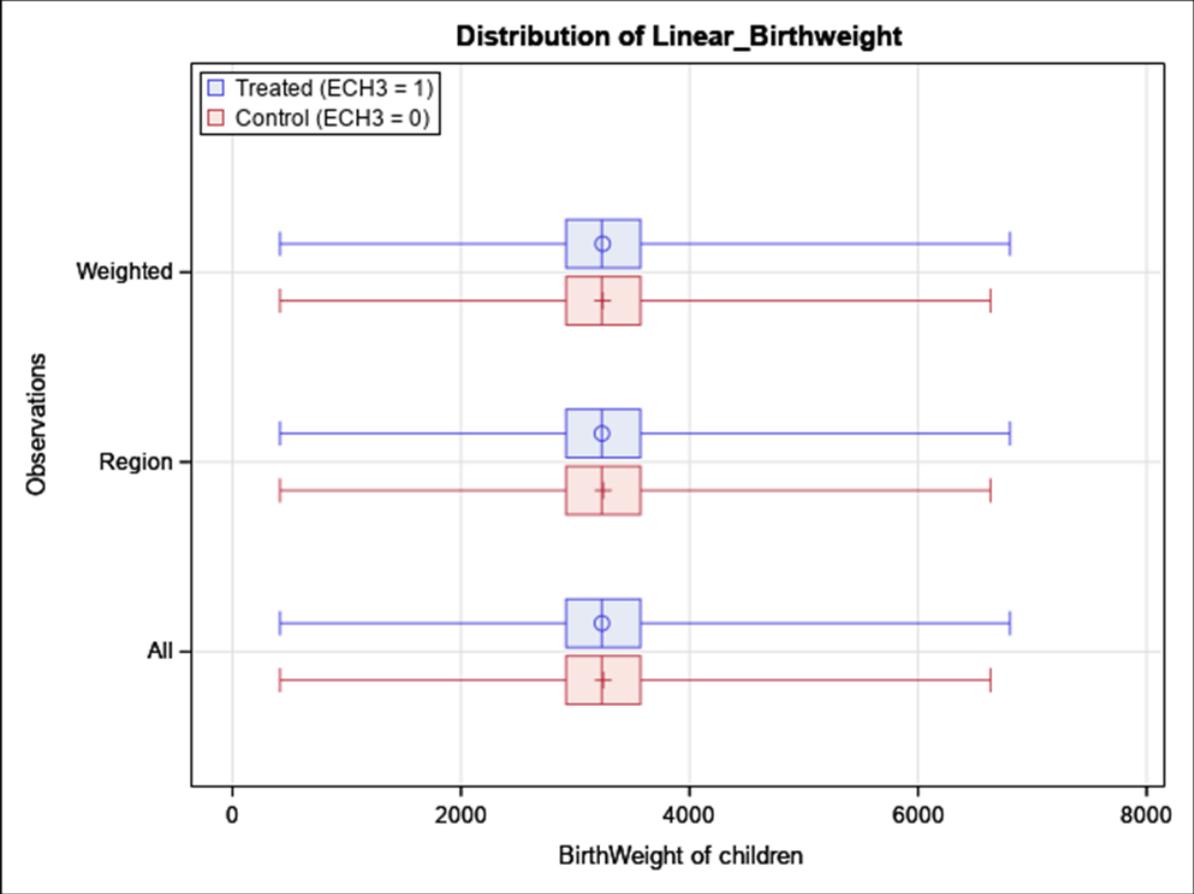


Figure 8.A14: Boxplot of continuous birthweight before (region) and after (weighted) applying weights between the analytical sample 3 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

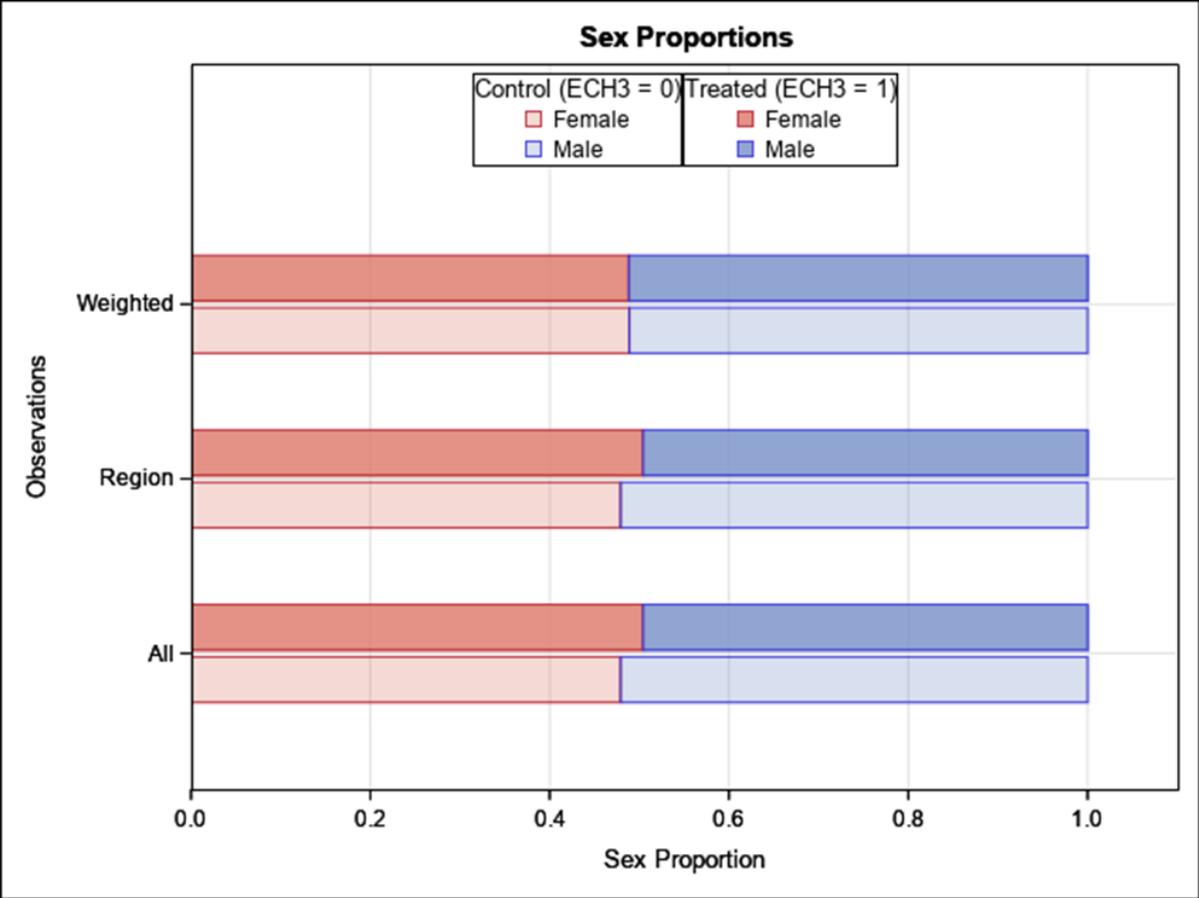


Figure 8.A15: Chart of sex proportions before (region) and after (weighted) applying weights between the analytical sample 3 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

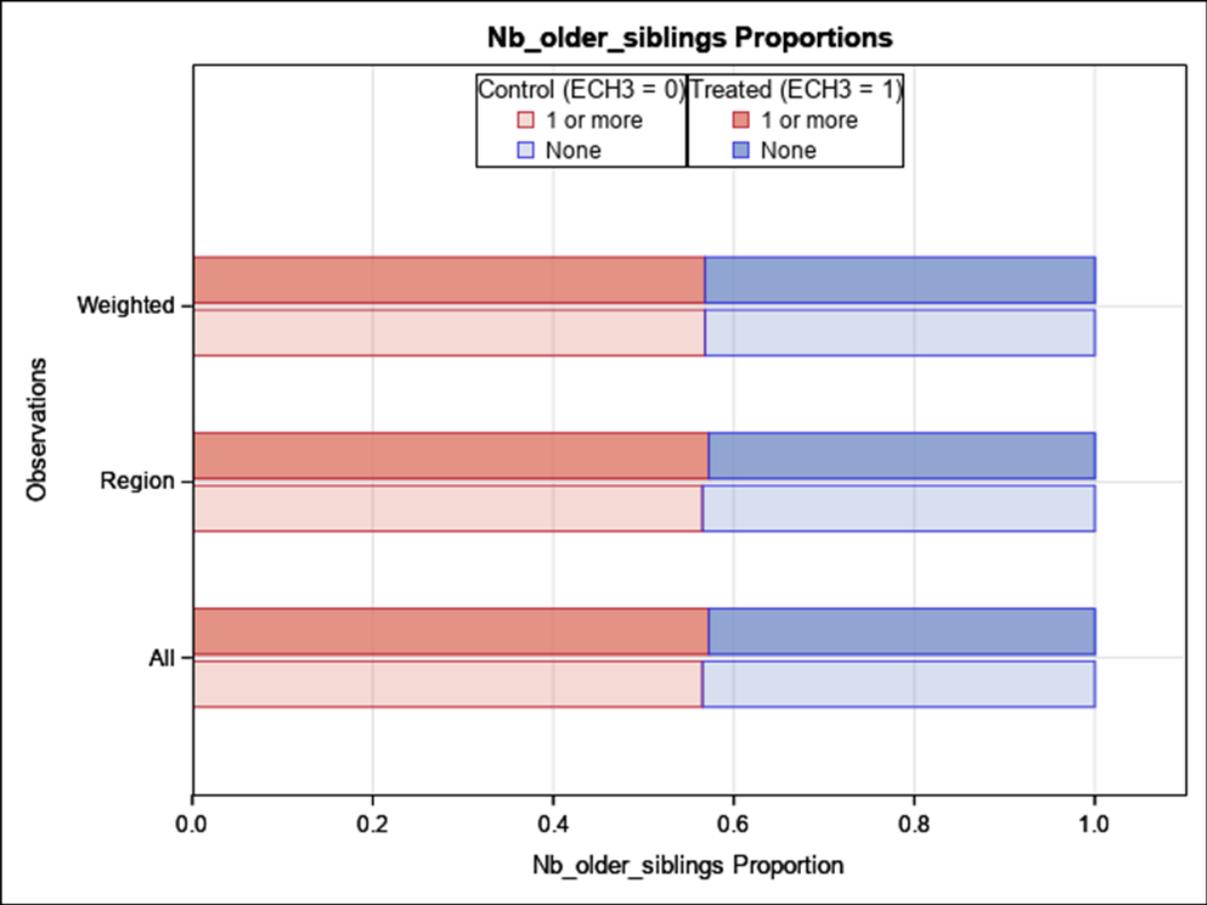


Figure 8.A16: Chart of number of older siblings proportions before (region) and after (weighted) applying weights between the analytical sample 3 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

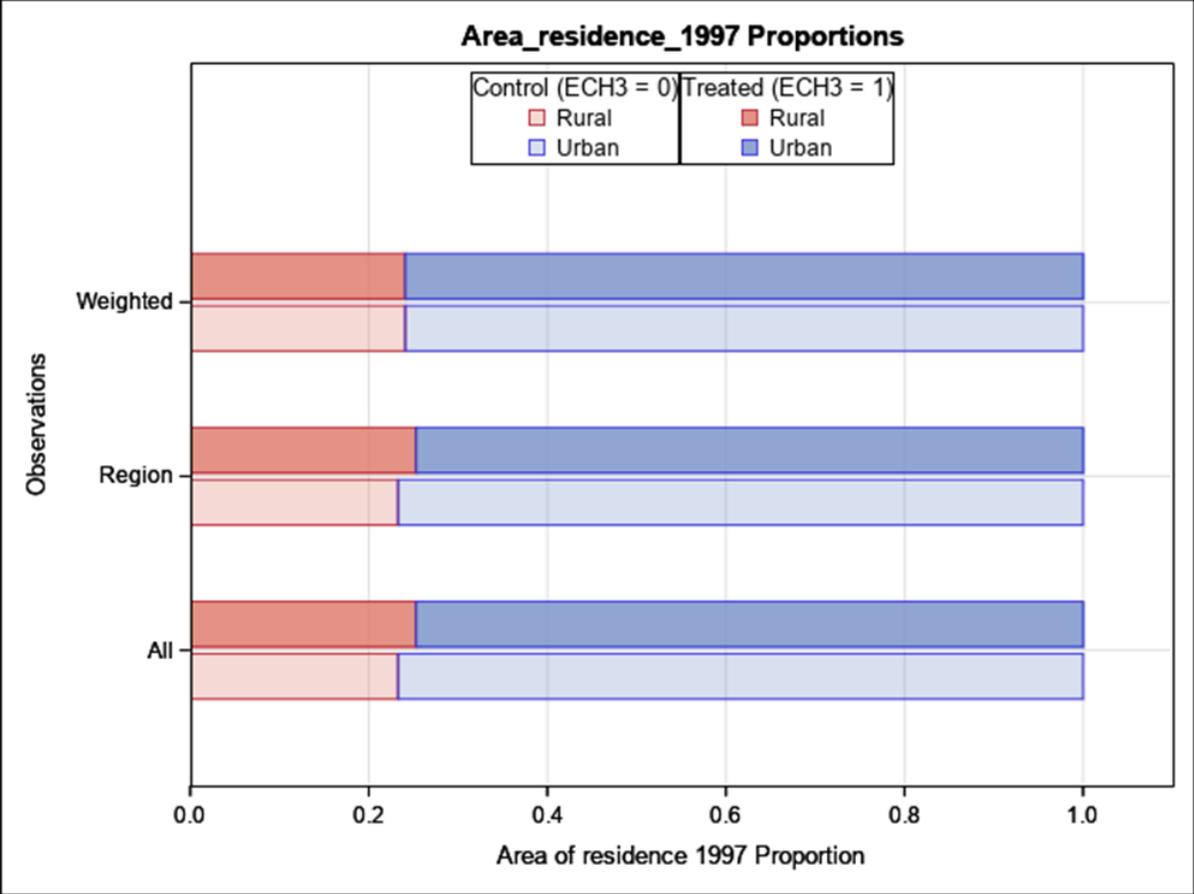


Figure 8.A17: Chart of area of residence in 1997 proportions before (region) and after (weighted) applying weights between the analytical sample 3 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

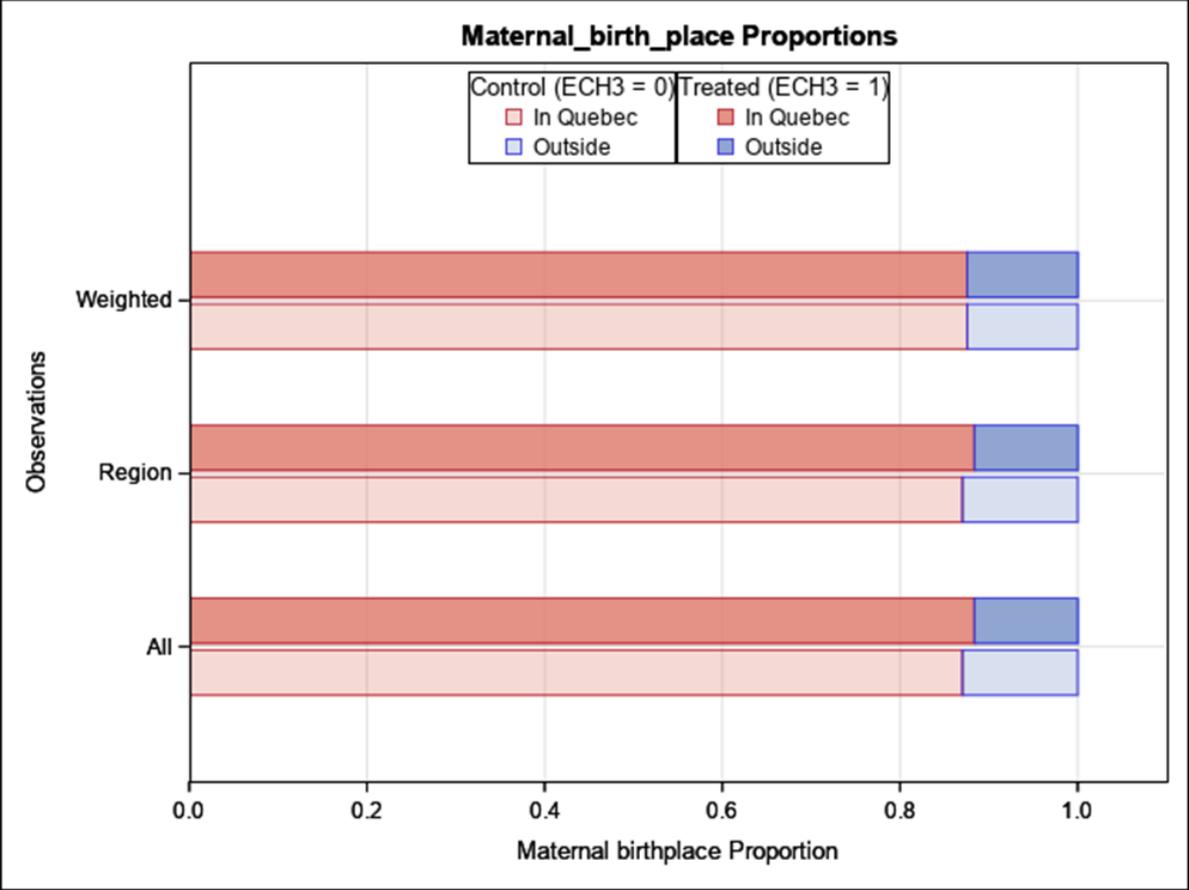


Figure 8.A18: Chart of maternal birthplace proportions before (region) and after (weighted) applying weights between the analytical sample 3 and the eligible sample from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

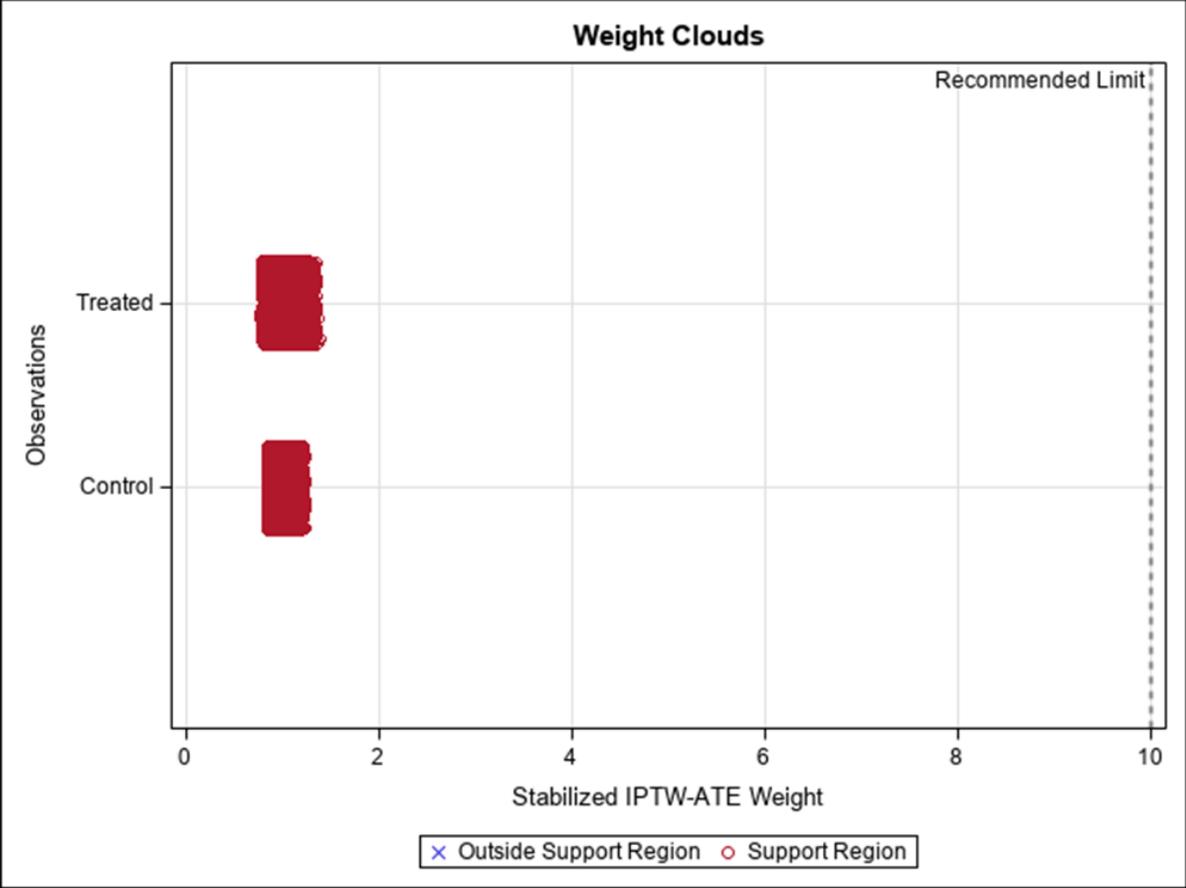


Figure 8.A19: Weight cloud of stabilized inverse probability weightings within the drug insurance coverage (treated, sample 3) and non drug insurance coverage (control) subsets from the Quebec Birth Cohort on Immunity and Health (1997-2014), Quebec, Canada.

Legend for all Figures: Ordinal categorical variables (material and social deprivation 1997 as well as income in 1996) requiring the use of dummy variables for presentation in charts were dropped to facilitate the presentation of deviances and weights.

Compilation based on data from the ©Government of Quebec, Institut de la statistique du Québec, 2017. Institut de la statistique du Québec is not responsible for compilations or interpretation of results.

## 9 Discussion générale

---

Les trois articles du présent doctorat présentent des associations entre la vaccination au BCG, ou l'âge à la vaccination, et les survenues de sclérose en plaques ou de différents phénotypes de diabète afin d'évaluer si cette vaccination est protectrice. Tous trois apportent de nouvelles connaissances épidémiologiques sur ces sujets.

### 9.1 Retour sur les résultats principaux

Dans le premier article, lorsqu'une première association est estimée sur l'ensemble du suivi, en incluant toutes les formes cliniques de sclérose en plaques, nous observons une association délétère entre la vaccination au BCG et la survenue de cette maladie. Pourtant, des analyses plus approfondies, divisées en deux périodes de suivi ( $\leq 21$  et 22-44 ans), et selon les formes cliniques, récurrente-rémittente avérée et non spécifiée, ont révélé plusieurs informations importantes. Tout d'abord, nous avons mis en lumière une absence d'association avec la forme récurrente-rémittente, et cela quel que soit l'âge de survenue sur l'ensemble du suivi. Avant 21 ans, peu importe le phénotype clinique, la survenue de sclérose en plaques n'était pas associée à la vaccination au BCG. Seule la survenue de sclérose en plaques, entre 22 et 44 ans, et dont le phénotype clinique était inconnu, étaient associées à cette vaccination. Cette catégorie pouvait regrouper des formes progressives primaire et secondaire, mais aussi des cas de forme récurrente-rémittente d'apparition plus tardive, non identifiés par notre caractérisation phénotypique.

Les articles 2 et 3 décrivent des résultats novateurs portant sur la vaccination au BCG et la survenue de diabète. Elles sont complémentaires et portent notamment sur le diabète de type 1 et sur de nouvelles associations inverses pour les autres phénotypes cliniques de diabète, à ce jour non rapportées dans la littérature scientifique. Dans l'article 2, nous avons pu tout d'abord estimer l'association entre la vaccination au BCG et la survenue du diabète de type 1 dans l'adolescence grâce à une approche méthodologique tentant de pallier au mieux les limites des précédentes études. Nous n'avons pas observé d'association, conformément aux précédentes études. Alors que ces dernières se concentraient toutes sur ce seul phénotype dans l'enfance et l'adolescence, périodes où l'incidence est la plus forte pour le diabète de type 1, nous nous sommes intéressés à la survenue de diabète à un âge plus avancé, entre 22 et 44 ans. Ceci nous a permis d'établir une caractérisation phénotypique novatrice sur trois formes cliniques de diabète dans la troisième

contribution. Dans cette période de suivi, nous avons à nouveau observé une absence d'association entre la vaccination au BCG et le diabète de type 1 survenant avant l'âge de 30 ans. Ceci est concordant avec nos premiers résultats portant sur la période de l'adolescence. Nous avons également observé, pour la première fois, des associations inverses avec la survenue de diabète de type 1 tardive (>30 ans, HR ajusté= 0,65, IC 95% : 0,44-0,95) et de type 2 précoce (HR ajusté=0,85, IC 95% : 0,79-0,92). Enfin, nous avons décrit pour la première fois les prévalence et incidence du diabète LADA, au Québec, dans notre période d'étude. Nous avons observé avec cette forme clinique une association délétère non-significative avec la vaccination au BCG (HR ajusté=1,30, IC 95%: 0,71-2,38), sur laquelle il est impossible de statuer avec certitude compte tenu des faibles effectifs identifiés.

Étant donné que les effets non-spécifiques de la vaccination au BCG varient selon le sexe (46), systématiquement, nos contributions se sont intéressées à évaluer la potentielle modification d'effet du sexe sur nos mesures d'association, ce que les précédentes études n'ont pas fait.

## **9.2 Validité interne**

### **9.2.1 Qualité de l'information sur la vaccination au BCG**

La qualité des informations sur la vaccination au BCG contenue dans le registre numérisé du Québec a été évaluée préalablement aux trois articles composant cette thèse. Toutes les associations étaient basées sur un statut vaccinal et un âge à la vaccination issus de ce registre dont le faible nombre d'informations manquantes (concordance à 99,6 %) et la fiabilité ont déjà été démontrées (279). Les moyennes annuelles de vaccinations, notamment dans les 5 années considérées de la cohorte, étaient élevés en bas âge (jusqu'à 46% pour les enfants vaccinés <1 an) (282). Ce registre recherchait l'exhaustivité des vaccinations réalisées, néanmoins proposées sur la base du volontariat (282). Cette spécificité, nous garantissait un nombre important de vaccinés, mais aussi de non-vaccinés, tous aussi important dans l'établissement de nos associations. Lorsque les poids de correspondance globaux étaient inférieurs au seuil choisi (et rejetés) entre les paires d'enregistrements du couplage probabiliste du registre et de la cohorte (305), ces expositions vaccinales étaient dites « incertaines » (3,1%, en faible nombre). Ces statuts d'exposition ont été éliminés puis remplacés par une information issue d'une imputation par équations chaînées basée

sur l'ensemble des informations disponibles des membres de la cohorte (prévention interne d'un biais de classification). Néanmoins, tout ceci contribue à une connaissance quasi parfaite des vaccinés mais pas des non-vaccinés. Le statut non-vacciné est en effet le résultat d'une non-présence de l'individu dans le registre, et de l'appariement à la cohorte. Or, certains individus auraient pu se faire vacciner en dehors du programme de vaccination du Québec déployé en 1949-1976, (279, 280) ou leurs informations auraient pu ne pas être transmises au registre central. Ce biais d'information conduirait à une sous-estimation du nombre de vaccinés et une surestimation du nombre de non-vaccinés. Cependant, cette erreur de classification ne serait vraisemblablement pas différentielle selon la survenue des maladies d'intérêt (405). En effet, si le fait de vacciner un enfant est influencé par des différences sociales, économiques et démographiques (282, 307), la connaissance à priori de diabète ou de sclérose en plaques dans la famille ne prédestine en rien les parents à s'opposer à la vaccination au BCG. Dans les dernières décennies, la vaccination, parfois polémique, pouvait potentiellement obscurcir leurs jugements. Néanmoins, à ce jour, selon une méta-analyse récente, aucune vaccination ne peut être incriminée dans la survenue des maladies auto-immunes (406) (et encore moins pour une maladie métabolique comme le diabète de type 2). Ce biais de classification non-différentiel aurait tendance à mener à une sous-estimation de l'effet (405).

### **9.2.2 Identification des cas incidents pour les maladies d'intérêt**

Ces trois articles utilisaient toutes en premier lieu des définitions validées pour la détection des maladies d'intérêt à partir de données médico-administratives (70, 308-310). Afin de détailler la caractérisation phénotypique de nos maladies, ces articles circonscrivaient des périodes d'étude selon l'âge de survenue et/ou utilisaient les données pharmaceutiques avec des définitions novatrices.

Les définitions de Marrie (308-310), Nakhla (70) et du Système de surveillance canadien des maladies chroniques (pour le diabète) (318, 319) ont ainsi tour à tour été utilisées pour identifier les cas incidents de sclérose en plaques pédiatrique ou adulte, et de diabète à l'adolescence ou à l'âge adulte, sans caractérisation phénotypique. Chacune de ces définitions présentait de très bonnes sensibilité et spécificité pour identifier les cas incidents des maladies d'intérêt dans les bases médico-administratives. La définition de Marrie pour la sclérose en plaques a ainsi une

sensibilité de 96% (IC 95% : 92,3-98,3) et une spécificité de 99,5% (IC 95%: 97,3-100,0) (308-310); celle de Nakhla pour le diabète de type 1, une sensibilité de 90,4% et une spécificité de 100% (70); et celle du Système de surveillance canadien des maladies chroniques une sensibilité comprise entre 88,9–89,8% et une spécificité entre 97,5–97,7% (318).

La caractérisation phénotypique des maladies d'intérêt s'est faite à partir de l'information pharmaceutique, sauf pour l'article 2, portant sur le diabète de type 1 dans l'adolescence. Dans celle-ci, les cas de diabètes de type 1 ont été circonscrits à l'aide d'une fin du suivi à 18 ans. Dans la période d'étude, entre 1985 et 1993, les cas de diabète de type 2 chez les enfants et les adolescents étaient encore rares (63, 388).

Pour l'article 1, l'utilisation des données pharmaceutiques a permis de préciser les phénotypes cliniques mais aussi de potentiellement augmenter la sensibilité de la définition médico-administrative validée de Marrie. Nous avons en effet identifié 576 cas de sclérose en plaques récurrente-rémittente avérée par l'utilisation de médicaments qui, dans notre période d'étude, ne pouvaient être utilisés que par ce type de patients (spécificité de l'utilisation des DMT). Parmi ces 576 cas de sclérose en plaques récurrente-rémittente, 48 (8,3%) étaient des cas supplémentaires, uniquement identifiés par l'utilisation des médicaments mais pas par la définition de Marrie. Les 528 autres cas ont donc été doublement identifiés par une définition précédemment validée (Marrie) et notre nouvelle définition pharmaceutique. Bien que l'identification des cas de sclérose en plaques récurrente-rémittente à l'aide des données pharmaceutiques n'ait pas été validée en soi, ces deux points témoignent d'un accroissement potentiel de la sensibilité si ces 48 cas sont réellement de vrais cas positifs. L'ajout des données pharmaceutiques par rapport à la seule définition médico-administrative de Marrie, augmenterait ainsi la détection de vrais positifs. La très grande spécificité de l'utilisation des DMT par les patients atteints de sclérose en plaques récurrente-rémittente justifie leur utilisation pour l'identification de ce phénotype. Néanmoins, leur utilisation n'a été détectée que dans l'échantillon pharmaceutique (70,4% des membres de la cohorte). Nous ne pouvons exclure que de potentiels cas supplémentaires de sclérose en plaques récurrente-rémittente auraient pu être détectés dans la partie complémentaire de l'échantillon éligible (non couvert par l'assurance pharmaceutique). Néanmoins pour pallier l'absence d'information dans ce sous-groupe, nous avons circonscrit nos analyses à l'échantillon pharmaceutique puis redressé le biais de sélection de cette analyse par une pondération qui corrigeait les potentielles différences entre ces deux parties de l'échantillon éligible.

Pour les diabètes à l'âge adulte (article 3), nous avons encore raffiné l'utilisation des données pharmaceutiques. Dans l'échantillon couvert par l'assurance pharmaceutique ( $N_2=260\ 440$ ), nous identifions en premier lieu 9 874 cas de diabètes incidents, dont 4 292 cas identifiés par la seule définition CCDSS, 1 355 par la seule utilisation de médicaments spécifiques au diabète et 4 227 par les deux pour un total de 5 582 cas ayant une caractérisation phénotypique de diabète de type 1, de type 2 ou de LADA. La définition CCDSS détectait donc 8 519 cas (86,3%) et notre nouvelle définition pharmaceutique 1 355 cas supplémentaires (13,7%). La sensibilité de cette définition étant de 88,9–89,8%, la proportion de cas supplémentaires semble relativement plausible. Comme pour le précédent article sur la sclérose en plaques, nous ne disposions pas de données cliniques pour valider que ces cas supplémentaires étaient bien de vrais positifs. Néanmoins, là encore la très grande spécificité de l'utilisation de l'insuline et des antidiabétiques oraux (ADO) justifie leur utilisation. Le fait de ne détecter aucune insuline (spécifique d'un diabète auto-immun) ou médicaments ADO chez des individus uniquement identifiés par la définition CCDSS et couverts par l'assurance médicaments, nous a également permis d'émettre l'hypothèse que 2 766 cas (autres que les 5 582), pouvaient potentiellement être des diabètes de type 2 non insulino-requérants, dont le traitement vise la modification des habitudes de vie, ou non observants de leur ADO. En effet, au moins 45 à 50 % des patients diabétiques de type 2 n'adhèrent pas à leur traitement (395, 396), et ces 2 766 cas sans caractérisation phénotypique avaient de plus des âges de diagnostic similaires aux personnes identifiées comme ayant un diabète de type 2. Là encore l'utilisation des données pharmaceutiques a potentiellement permis d'augmenter la sensibilité de la première définition médico-administrative validée du CCDSS.

### **9.2.3 Confusion**

Systématiquement, la confusion a été traitée en établissant des graphiques acycliques dirigés (337) basés sur des analyses rigoureuses de la littérature scientifique pour chacune des maladies et phénotypes d'intérêt et des expositions principales. Aucune autre étude précédente n'a été aussi attentive à la formalisation de la structure de cette confusion potentielle. Or, un mauvais ajustement sur la confusion peut entraîner de la confusion résiduelle ou un sur-ajustement dans les études observationnelles, ce qui peut être tout aussi important qu'un biais de classification sur l'exposition ou l'évènement d'intérêt (405). Aucune variable établie comme facteur de confusion dans les trois

contributions n'était indisponible dans la cohorte étudiée, une limite souvent soulevée pour des études utilisant des données administratives. Néanmoins, il est impossible d'exclure des erreurs de mesure sur les facteurs de confusion utilisés. De plus, la non-proportionnalité des HR a parfois obligé l'exclusion de certains facteurs de confusion de nos modèles. La confusion résiduelle reste donc encore possible.

#### **9.2.4 Pertes au suivi**

Lors de la constitution de la Cohorte de naissance québécoise sur l'immunité et la santé, des variables liées aux interruptions de la couverture RAMQ ont été demandées mais la transmission de celles-ci a été refusée pour des raisons légales et de confidentialité. Afin d'étudier le potentiel biais de sélection engendré par des perdus de vue durant le suivi, nous avons mené des analyses de sensibilité sur les mesures d'association de l'article 1 qui utilisait l'ensemble du suivi de 1983 à 2014. Le renseignement annuel des codes postaux de résidence a ainsi été utilisé pour déduire le statut de maintien ou de perte au suivi, des valeurs manquantes sur plusieurs années étant considérées comme des potentielles pertes au suivi. Ces pertes au suivi étaient centralisées majoritairement sur la période 1997-2014. Trois analyses de sensibilité ont donc été réalisées sur cette unique période, consistant à reproduire les mesures d'association en excluant les individus ayant, soit un statut de potentielle perte au suivi total par le manque total des codes postaux de résidence à partir d'une certaine date, soit par des manques parcellaires de ces mêmes codes postaux durant le suivi de la période considérée. La première analyse n'excluait que 10 701 individus, soit 2,7% de la cohorte étudiée, présentant des codes postaux totalement manquants de 1998 à 2014. Néanmoins, sur la période 1997-2014, 23 512 personnes (5,9% de la cohorte) présentaient des codes postaux manquants sur une ou des périodes, pouvant être continues ou discontinues durant  $\geq 10$  ans et 38 905 (9,8% de la cohorte) pour une durée totale  $\geq 5$  ans. Au maximum, le nombre total de personnes exclues représentait donc 9,8% de notre large échantillon (N=397 666). Ce nombre de sujets exclus n'a eu que peu d'influence sur les personnes-années de suivi entre 1997 à 2014. La répartition des sujets analysés selon le statut vacciné/non-vacciné était très légèrement impactée, de façon non différentielle selon la survenue de sclérose en plaques puisque l'estimation des mesures d'association dans cette période, prenant en compte ces trois exclusions successives, ne présentait aucune variation avec les premières estimations effectuées.

Ces analyses suggèrent donc que les potentielles pertes au suivi n'ont pas engendré de biais sélection.

### **9.2.5 Impact de l'imputation des données**

Dans nos trois articles, 3,1% des individus avaient un statut vaccinal incertain. Ces sujets ont été exclus dans l'article 2. Pour les articles 1 et 3, ces valeurs manquantes ont été imputées. De plus, dans les trois articles, les valeurs manquantes pour les co-variables ont été imputées. Les proportions de sujets ayant au moins une valeur manquante ont varié selon les articles : 17,8% pour l'article 1, 6,6% pour l'article 2 et 9,6% pour l'article 3. Pour éviter d'exclure les sujets ayant des données manquantes et risquer d'introduire un biais sélection, l'imputation multiple par équation chaînées a été utilisée. Néanmoins, les différentes analyses sans imputation puis avec imputation montraient que les estimations étaient quasi-identiques.

## **9.3 Validité externe**

### **9.3.1 Représentativité de la cohorte**

Tous nos résultats ont été établis grâce à une large cohorte de naissance basée sur la population québécoise, appariée aux données médico-administratives et pharmaceutiques de ces individus et d'un statut et d'un âge de vaccination au BCG établi à partir d'un registre à la vaccination.

Afin d'être représentatif d'une population donnée, un échantillon a besoin de ne pas être une sélection biaisée de cette dernière. La sélection aléatoire des individus selon un plan de sondage prédéterminé prenant en compte la structure de la population d'origine, garantit habituellement cette représentativité. Néanmoins dans le cadre de nos études, la Cohorte de naissance québécoise sur l'immunité et la santé et le registre de vaccination au BCG visent tous deux l'exhaustivité de la population née entre 1970 et 1974. Notre cohorte est composée de 90,4% des individus nés dans ces 5 années de naissance après au moins 32 semaines de grossesse. Seuls les très grands prématurés (avant 32 semaines) ont été exclus des 5 années de naissances considérées en raison de leurs morbidités multiples et mortalité importante (302), ainsi que de leur immaturité immunologique. Ils représentaient donc une population spécifique, différente de la population

générale cible de nos études sur laquelle nous souhaitions généraliser nos résultats. La cohorte, de laquelle sont exclus 9.6% des personnes éligibles parce que le jumelage entre le registre des naissances et le FIPA n'a pas été réussi, est en donc hautement représentative de la population source.

Pour chacun de nos articles, les prévalence et incidence obtenues dans notre cohorte à partir de définitions médico-administratives validées et précisées par notre caractérisation phénotypique sur les données pharmaceutiques, ont été comparées à des estimations précédemment publiées (par phénotype) afin d'en évaluer la comparabilité et représentativité à notre population cible. Malgré des limites inhérentes, incluant des populations, des catégories d'âge, ou des périodes calendrier pouvant parfois différer, ces comparaisons ont suggéré que la cohorte était représentative et produisait des estimés de prévalence et d'incidence concordant avec les valeurs attendues.

Finalement, il est toutefois pertinent de mentionner que les taux d'incidence de sclérose en plaques récurrente-rémittente (article 1) estimés dans les deux périodes ne sont pas directement comparables, pour les raisons suivantes : 1) il s'agit d'une cohorte de naissance (donc qui avance en âge au fil du temps) et 2) l'utilisation des thérapies s'est accrue avec l'arrivée de nouveaux traitements.

#### **9.4 Études antérieures**

Pour la sclérose en plaques, six études cas-témoins (54-59), regroupées dans une méta-analyse de Farez et al. (60), établissaient jusqu'à présent une absence d'association, OR=0,96, IC 95% : 0,69–1,34 (60). Ces études n'étaient pas exsangues de limites méthodologiques qui pouvaient faire douter de leurs estimations qui oscillaient toutes entre 0,70 et 1,30 avec des intervalles de confiance non-significatifs. L'étude la plus large regroupait par exemple 140 cas de sclérose en plaques (58) et aucune ne distinguait les phénotypes cliniques.

Pour le diabète de type 1 dans l'enfance et l'adolescence, ce sont sept études qui présentaient toutes autant de limites méthodologiques (61-67). Parmi ces études, trois étaient des études de cohorte (61, 63, 66) et quatre, des études cas-témoins (62, 64, 65, 67). Leurs estimations d'association oscillaient entre 0,83 et 1,30 et là encore aucune association significative n'était mise en lumière.

Pour le diabète de type 1 à un âge plus avancé, il n'y avait jusqu'à ce jour aucune estimation d'association rapportée, tout comme pour le diabète de type 2 et pour le diabète LADA.

Les précédentes études, suggérant l'absence d'une association, pouvaient utiliser des variables d'exposition à la vaccination BCG basées sur une information auto-rapportée. C'était ainsi le cas pour cinq des six études portant sur la sclérose en plaques (55-59), et de deux sur sept portant sur le diabète de type 1 (64, 65). Néanmoins, dans l'étude Eurodiab, il s'agissait d'une combinaison entre des données auto-rapportées dans 15 et 20% des cas et des témoins ou une information fiable basée sur les carnets de santé des enfants (64). Or, il a été précédemment démontré que le statut vaccinal auto-rapporté présentait une faible validité qui pouvait engendrer des biais de classification/information (347).

## **9.5 Limites de nos trois études**

Pour la première fois, l'article 1 établit une mesure d'association entre la vaccination au BCG et le phénotype clinique le plus courant de sclérose en plaques, la forme récurrente-rémittente (85% des cas initiaux de la maladie). Néanmoins, le sur-risque mis en évidence dans le groupe de sclérose en plaques dont le phénotype était non identifié, soulève plusieurs questions. Les estimations de risque ont toutes été redressées par pondération selon une probabilité inverse, car établies dans une sous-population de l'échantillon éligible, dans laquelle la caractérisation phénotypique était possible. Ce redressement a résulté en une bonne correction du biais de sélection, tel que démontré par l'étude des poids. Malgré cela, nous ne pouvons exclure que de probables cas de sclérose en plaques récurrente-rémittente aient été présents dans l'échantillon non couvert par l'assurance médicaments. Il pourrait donc persister un biais d'information sur le groupe uniforme de sclérose en plaques récurrente-rémittente. Ces cas non détectés par la seconde définition (pharmaceutique) pourraient en partie être inclus dans le groupe dont le phénotype clinique est inconnu (progressive primaire, secondaire ou récurrente-rémittente) car la première définition (de Marrie) a été appliquée sur l'échantillon éligible et sa sensibilité est de 96%. Ce biais de classement sur le groupe de sclérose en plaques récurrente-rémittente pourrait être différentiel ou non différentiel selon l'exposition. Néanmoins, un biais différentiel est peu probable dans une cohorte dans laquelle l'information a été colligée de façon prospective (405). Si ce biais de classement était non différentiel, l'association entre la vaccination au BCG et la sclérose en plaques

récurrente-rémittente tendrait vers la valeur nulle. Pour la sclérose en plaques dont le phénotype clinique est inconnu, le biais de classement pourrait provenir du manque de sensibilité de la définition de Marrie. Cependant, l'impact de ce biais serait vraisemblablement faible (sensibilité=96%) et il serait également non différentiel. De la confusion résiduelle pourrait hypothétiquement expliquer une surestimation de l'effet et un sens opposé à notre hypothèse de recherche (405). Le sous-groupe où les formes cliniques ne sont pas identifiées est donc très difficilement interprétable. Les résultats obtenus sur le groupe de sclérose en plaques récurrente-rémittente doivent également être reproduits pour s'assurer de leur véracité.

Pour l'article 2, même si la survenue de diabète de type 2 restait relativement rare à l'adolescence durant cette période d'incidence (1985-1993) (63, 388), nous ne pouvons exclure un biais de classement parmi nos 442 cas dits de diabète de type 1. En effet, de potentiels cas de diabète de type 2 ainsi que des diabètes MODY (*Maturity-onset diabetes of the young*) auraient pu survenir entre 10 et 18 ans sans que nous ayons la possibilité de les caractériser. Ce biais de classement serait néanmoins non différentiel dans le cadre de notre cohorte et l'association tendrait alors vers la valeur nulle. L'impact d'un tel biais de classement sur les mesures d'association serait toutefois limité étant donné la très faible incidence des diabètes de type 2 et MODY dans une population de cet âge dans les années 1980s et 1990s. De plus, l'utilisation d'une période d'exclusion de 2 ans, à partir de la disponibilité de nos données (1983), avec une troncature en 1985, nous a certes permis de circonscrire les cas incidents dans la régression proportionnelle de Cox, mais nous a fait éliminer 335 cas de diabète survenant entre 1983 et 1985 ou en amont de cette période. Ceci pourrait apparaître au premier abord comme une sélection de notre population. Pourtant il n'en est rien. Nous avons certes sélectionné la période d'étude entre 1985 et 1993 mais nous n'avons en aucun cas induit un biais de sélection dans cette même période. L'ensemble des cas incidents de cette période ont été convenablement identifiés par la définition de Nakhla. Nous avons seulement éliminé en amont une période dans laquelle il était impossible de distinguer les cas prévalents et incidents. Ceci nous a permis d'éviter un classement hasardeux de cas prévalents survenus en amont de 1983, en cas soit-disant « incident » dans les deux années 1983 et 1984. Les mesures d'associations de l'article 3 apportent également un éclairage sur les résultats de l'article 2, sur le diabète de type 1 à l'adolescence. Nous observons en effet avec une puissance statistique plus limitée, pour ce phénotype, qu'il ne semble exister aucune association inverse avant 30 ans mais que celle-ci se manifeste bien après cet âge. Outre la probable différence de pathogénèse de ce

phénotype entre ces deux âges, ce résultat corrobore la mesure d'association établie dans l'adolescence.

Pour l'article 3, la définition CCDSS a permis d'identifier un très grand nombre de cas incidents de diabète (n=11 108) dans l'échantillon éligible, et l'échantillon des individus couvert par l'assurance médicaments (n=9 874). Néanmoins, notre caractérisation phénotypique ne permettait d'établir que 50,3% et 56,5% des formes cliniques en présence, laissant près de 40 à 50% des cas de diabète de l'échantillon éligible et de l'échantillon des individus couvert par l'assurance médicaments, pratiquement inutilisables dans nos régressions. La circonscription de nos analyses principales aux personnes couvertes pour les services pharmaceutiques durant la majeure partie du suivi (échantillon 3, N=160 622), le plus fiable au niveau des informations pharmaceutiques, permet d'utiliser l'ensemble des cas en présence : 3 331 avec une caractérisation phénotypique et 2 766 étant potentiellement des diabètes de type 2 non-observants (au total n=6 096 cas incidents de diabètes). De plus, le redressement par pondérations inverses stabilisées nous a permis une inférence de nos estimations à la population cible, et un redressement du biais de sélection. Néanmoins, le nombre de cas de diabète LADA caractérisés est faible. Le ratio de prévalence entre les diabètes LADA identifiés et tous les cas de diabète de notre étude témoigne d'une proportion au mieux de 1,3%, bien inférieure aux prévisions estimées entre 2% et 10% par des auteurs comme Nishimura et al. (6). Certains cas de diabète LADA restent donc incorrectement classés en diabète de type 1 ou de type 2, voire inclus dans la fraction où il ne nous était pas possible d'appliquer notre caractérisation phénotypique. Ce biais de classement pourrait être différentiel selon nos expositions principales et nous faire observer un sur-risque à l'opposé de notre hypothèse de recherche pour ce phénotype clinique (405). C'est pourquoi, il faut dès lors l'interpréter avec précaution.

Les diagnostics phénotypiques reposaient sur les données médico-administratives et/ou pharmaceutiques. Ils ne pouvaient prendre en compte certaines réalités cliniques faites en-dehors des standards de prescription des antidiabétiques. Ces quelques cas particuliers, non répertoriés, ont pu également engendrer des biais de classification, toujours non différentiel dans nos données. Leur impact serait donc présent mais probablement minimal.

Les informations médico-administratives n'étaient pas disponibles avant 1983 (et même 1987 pour les hospitalisations). Ce manque d'information ne nous a pas permis de détecter de potentiels cas

pédiatriques de sclérose en plaques et de diabète de type 1 entre la naissance et les 9-10 ans des membres de la cohorte. Néanmoins, pour la sclérose en plaques les cas pédiatriques ( $\leq 16$  ans) sont relativement rares et leur recherche dans notre période d'étude 1983-1996, établissait un nombre  $< 5$ . Pour le diabète de type 1, ce manque d'information nous a obligé à centrer notre étude sur l'adolescence. Cette période inclut le pic d'incidence de cette maladie (69), nous ayant permis d'identifier 442 cas et faisant de notre étude la plus vaste à ce jour, avec une très bonne généralisation de nos résultats à la période d'adolescence. Pour le troisième article, notre volonté était de circonscrire l'étude à une période où les membres de la cohorte étaient tous adultes pour compléter les connaissances manquantes sur ce sujet. L'absence d'information avant 1983 n'était donc pas du tout limitante.

## **9.6 Forces de nos trois études**

Les mesures d'association établies dans nos trois articles le sont avec une puissance statistique inégalée, des mesures fiables tant de l'exposition principale que du phénotype clinique d'intérêt, et un traitement méticuleux de la confusion (337). La présente agrégation de la Cohorte pour l'immunité et la santé a ainsi été multipliée par 5 depuis les premiers travaux de notre groupe (63, 303) et les données du registre au BCG se sont avérées complètes et fiables (279). Lorsque cela était possible toutes nos régressions ont pris en compte les phénotypes cliniques de nos événements d'intérêt. Ceci est particulièrement important pour la sclérose en plaques où les cas sont suspectés d'avoir des processus inflammatoires partiellement différents (23). Cela l'est également pour les diabètes dont les pathogénèses et les déterminants sont établis et différents (27, 199).

Dans deux de nos articles nous nous appuyons pour la première fois sur des données médico-administratives, en utilisant des définitions validées, et sur des données pharmaceutiques dont l'utilisation est très spécifique de la forme clinique des maladies d'intérêt. Dans le cadre de la sclérose en plaques, Culpepper et al. avaient déjà démontré que l'utilisation de ces données pharmaceutiques augmentait la sensibilité des définitions médico-administratives (353). Pour chacun des articles 1 et 3, nous avons en effet bénéficié d'une large couverture de l'assurance médicaments publique (65-70%) permettant de garantir une très bonne représentativité de notre population cible. De plus, pour pallier la sélection d'une partie des échantillons éligibles des deux études, nous avons utilisé des redressements par pondérations inverses stabilisée, favorisant ainsi

la correction d'un possible biais de sélection. Le manque de couverture de l'assurance médicaments publique en amont de 1997 (seulement 5% de la cohorte a été couverte durant au moins une période avant 1997), n'a pas du tout été limitant dans nos analyses. Avant 1997, l'utilisation des médicaments modificateurs d'effet de la maladie pour la sclérose en plaques n'était pas autorisée. Pour l'article 3, la restriction à des âges adultes entre 1997 et 2014 nous a permis de ne pas être limité par les 5% de couverture de l'assurance médicaments publique entre 1983 et 1996.

Dans la période 1997-2014, les 70.4% d'individus ayant une couverture médicaments (N=282,081) avaient une durée médiane cumulative de 6,9 ans contre 12,6 ans pour ceux ayant recours aux médicaments modificateurs d'effet. Dans les 18 années de suivi, les utilisateurs de ces médicaments modificateurs d'effet étaient donc couverts 70% (12,6/18) du temps en moyenne. Dans la période 1997-2014, les 65.7% d'individus ayant une couverture médicaments (N<sub>2</sub>=260,440) avaient une durée médiane cumulative de 5,9 ans contre 12,9 ans pour les individus ayant recours à au moins 1 médicament antidiabétique. Dans les 18 années de suivi, les utilisateurs de ces médicaments antidiabétiques étaient donc couverts 72% (12,9/18) du temps en moyenne.

La prescription de médicaments par les médecins peut parfois être détournée de l'étiquette, c'est-à-dire de l'agrément donné par Santé Canada pour une ou des indication(s) spécifique(s) d'un médicament donné. Environ 20% des prescriptions canadiennes étaient détournées de l'étiquette en 2005 (407). Malgré la très grande spécificité des médicaments étudiés, médicaments modificateurs d'effet et antidiabétiques, la probabilité d'une utilisation sans indication officielle restait donc possible mais certainement très rare. En effet, les médicaments modificateurs d'effet étaient hautement novateurs et spécialisés entre 1997 et 2014 pour la sclérose en plaques récurrente-rémittente, selon leurs étiquettes. De plus, notre définition prenait en compte  $\geq 1$  médicaments modificateurs d'effet. Dans le cas de la sclérose en plaques, les interférons bêta 1a et 1b étaient ainsi toujours associés à l'utilisation de l'acétate de glatiramère sur la période de couverture des individus. Ces patients pouvaient même avoir essayé jusqu'à 7 classes thérapeutiques de façon concomitante ou juxtaposée durant leur couverture à l'assurance médicaments (acétate de glatiramère, interférons bêta 1a et 1b, anti-CD52, Anti-SP-1P, anti-intégrine alpha4, inhibiteurs de  $\Sigma$  de pyrimidine, bloqueurs des canaux K<sup>+</sup> (ce dernier n'étant pas spécifique de la sclérose en plaques mais associé aux autres classes)). Pour les antidiabétiques, notamment les plus anciens comme la metformine, l'agrément n'a jamais été renouvelé malgré

l'utilisation aujourd'hui dans d'autres indications comme le syndrome des ovaires polykystiques ou le prédiabète (408-410). Néanmoins comme les médicaments modificateurs d'effet, notre définition prenait en compte  $\geq 1$  antidiabétiques qui pouvaient être associés de façon concomitante ou juxtaposée durant leur couverture à l'assurance médicaments. La prescription en dehors de l'agrément est une pratique clinique qui ne pouvait donc pas être prise en compte dans notre étude. Néanmoins, elle devait être marginale aux vues de la spécificité et de l'association des différentes classes thérapeutiques utilisées.

Enfin, plusieurs sous-analyses et analyses de sensibilité ont été entreprises pour s'assurer de la robustesse de nos résultats et éclairer leur interprétation au niveau des trois articles.

### **9.7 Mécanismes immunologiques sous-jacents potentiels**

Certains scientifiques ont suggéré que l'immunité entraînée n'est que le terreau des maladies auto-immunes (299). Le sur-risque global établi dans la contribution 1, pourrait dès lors étayer leur hypothèse. Pourtant l'étude détaillée de ce sur-risque démontre qu'il s'explique par une absence d'association entre la vaccination au BCG et la survenue de sclérose en plaques récurrente-rémittente sur l'ensemble du suivi et d'un sur-risque estimé sur un sous-groupe présentant un biais de classement manifeste puisqu'il contient en toute vraisemblance tous les phénotypes cliniques de sclérose en plaques tels que les formes progressive secondaire, progressive primaire et récurrente-rémittente. Il faudrait donc estimer des associations par phénotype et notamment avec les formes progressive primaire et progressive secondaire pour affirmer que l'hypothèse de Arts et al. est plausible. Nos résultats suggèrent en effet que cette vaccination n'est pas dangereuse vis-à-vis de la forme la plus courante de sclérose en plaques. De plus, certains auteurs favorisent des processus inflammatoires partiellement différents notamment entre les deux formes principales de sclérose en plaques, i.e., progressive primaire et récurrente-rémittente (23). Or, si on fait l'hypothèse d'une pathogénèse et/ou de déterminants différents, l'association entre la vaccination au BCG et la sclérose en plaques devrait être estimée par phénotype, comme nous l'avons réalisé pour les différents diabètes, et non en considérant tous les cas globalement. Les associations pourraient en effet être différentes selon le phénotype considéré.

Pour la première fois, nous observons dans une étude populationnelle bien menée et d'envergure, une protection de la vaccination au BCG sur les survenues de diabète de type 1 et 2 à long terme.

Ces résultats concordent avec ceux d'essais cliniques thérapeutiques (40, 296, 297) et d'expérimentations animales (41-43) menés à plus court terme. Ces résultats s'opposent donc frontalement à l'hypothèse délétère de Arts et al. vis-à-vis de la vaccination au BCG et de l'immunité entraînée (299).

Parmi les potentiels mécanismes sous-jacents, certains auteurs proposent un changement de métabolisme du glucose, des cellules innées et adaptatives (37, 39, 41-43). Ce changement induit par leurs activations respectives est toutefois très courant et en rien spécifique de l'injection du BCG. De plus, comment des cellules effectrices pourraient-elle rester actives pour induire une réduction pérenne de l'hyperglycémie chronique, qui plus est à si long terme ? Le mécanisme régulateur proposé par Ristori et Faustman via les deux voies de signalisation du TNF- $\alpha$  (40, 296, 298), pourrait potentiellement expliquer nos résultats sur les deux diabètes d'origine auto-immune. Néanmoins, cette cytokine pro-inflammatoire (associée à l'IFN- $\gamma$ ) n'a rien d'un mécanisme régulateur au niveau du système nerveux central où elle accentuerait la destruction de la gaine de myéline, des oligodendrocytes et des neurones en renforçant la chronicisation des processus inflammatoires de la sclérose en plaques (84). Un tel mécanisme, au niveau du système nerveux central, irait dans le sens de l'hypothèse délétère de Arts et al (299). Pourtant nous observons une absence d'association avec la sclérose en plaques récurrente-rémittente, représentant 85% des formes cliniques initiale. Vis-à-vis du diabète de type 2, le TNF- $\alpha$  renforcerait également l'inflammation causée par l'obésité (potentialisation des adipokines (144)) induisant de fait un renforcement de la résistance et de l'insuffisance à l'insuline, par une destruction des cellules  $\beta$ . Pourtant là encore, nous observons une protection vis-à-vis du diabète de type 2.

Au final, seule l'immunité entraînée (31), dans un cadre vaccinal, pourrait rester assez pérenne, et ce jusqu'à 44 ans après l'injection néonatale (37, 39), car encryptée épigénétiquement dans l'ADN des progéniteurs hématopoïétiques des cellules myéloïdes. Pourtant, contrairement à Arts et al. (299), nous pensons que l'immunité entraînée ne serait pas délétère, bien au contraire. Nous émettons l'hypothèse d'une boucle de rétrocontrôle de l'immunité adaptative régulatrice, activée par des seuils de synthèse de cytokines pro-inflammatoires plus élevés induits par l'immunité entraînée. Cette immunomodulation par boucle de rétrocontrôle agirait à la fois sur les processus auto-immuns et/ou inflammatoires de la sclérose en plaques et des diabètes de type 1 et LADA, mais aussi sur les déterminants du diabète de type 2, en réduisant l'inflammation chronique. Elle n'agirait donc pas directement sur l'hyperglycémie chronique qui n'est que la résultante des trois

grands phénotypes de diabète. Néanmoins, cette hypothèse demande à être démontrée de façon expérimentale et mécanistique. Enfin, la confrontation de nos résultats entre les articles 2 et 3, permet de corroborer l'hypothèse d'une très grande hétérogénéité des processus immunologiques auto-immuns chez les individus développant un diabète de type 1 (85), comme celle établie dans la pathogénèse du diabète LADA (199).

## 10 Perspectives futures de recherche

---

Le caractère délétère des associations entre le BCG et la sclérose plaques, toutes formes confondues, amène à se poser de nombreuses questions auxquelles il faudra répondre. Il sera nécessaire de reproduire ces analyses par phénotype clinique avec une robuste caractérisation clinique. Il sera également nécessaire d'aller au-delà des 44 ans de suivi. En effet, les formes cliniques progressive primaire ou progressive secondaire, sur lesquelles des doutes persistent encore, surviennent plus tard, notamment pour la forme progressive primaire ayant un pic d'incidence aux alentours de 60 ans.

L'absence d'association dans l'adolescence avec la survenue de diabète de type 1 est une nouvelle pierre à l'édifice de l'évidence épidémiologique. Néanmoins, il sera nécessaire de mesurer cette association dans l'enfance afin de combler les connaissances sur ce sujet.

Vis-à-vis de l'article 3, il sera nécessaire d'effectuer des analyses après 44 ans au moment où le diabète de type 2 est le plus courant, puisque son pic d'incidence se situe entre 55 et 60 ans. Nos résultats mettent en lumière le caractère très hétérogène des pathogénèses tant du diabète de type 1 que du diabète LADA. Néanmoins, les associations délétères observées pour ce phénotype sont difficilement interprétables et généralisables, aux vues des limites méthodologiques que nous avons rencontrées. Là encore, nos résultats attirent l'attention sur la nécessité de les reproduire avec une caractérisation phénotypique adéquate ou des données cliniques robustes.

Les études reposant sur des données médico-administratives doivent donc être poursuivies à des âges plus avancés aussi bien pour la sclérose en plaques que pour les différents types de diabète. Néanmoins, ces études devront nécessairement être appuyées par des études cliniques et/ou populationnelles permettant une caractérisation robuste des formes cliniques.

## 11 Conclusion

---

Cette thèse contribue aux connaissances scientifiques sur la vaccination au BCG et son lien avec le développement de maladies inflammatoires chroniques d'origine auto-immune et/ou métabolique.

Pour le 1er objectif, nous avons observé une association délétère lorsque toutes les formes cliniques de sclérose en plaques étaient étudiées ensemble. Le second temps de nos analyses a révélé que l'étude par phénotype doit être prise en compte dans les futures publications puisque nous avons observé une absence d'association entre la vaccination au BCG et la sclérose en plaques récurrente-rémittente, la forme initiale la plus courante. Le sur-risque mis en évidence dans l'association globale demeurait dans l'analyse d'un sous-groupe où le regroupement des différentes formes cliniques était conservé par un défaut d'information.

Pour le 2e objectif, en comblant les défauts méthodologiques des précédentes études, nous n'avons observé aucune association entre la vaccination au BCG et la survenue de diabète de type 1 dans l'adolescence. Cet article confirme donc les précédentes études en levant un peu plus le doute sur l'absence d'association.

Pour le 3e objectif, les résultats obtenus dans la période d'analyse 1997-2014 ont suggéré pour la première fois, l'existence d'associations inverses entre la vaccination au BCG et la survenue des diabètes de type 1 et 2 à long terme. Ils comblent les manques de la littérature scientifique sur ce sujet et soutiennent l'hypothèse d'une immunomodulation adaptative par rétrocontrôle de l'inflammation accrue de l'immunité entraînée, et ce jusqu'à 44 ans après une seule injection vaccinale.

Autre que pour des effectifs réduits dans certaines sous-analyses, nos travaux ont surmonté les lacunes méthodologiques des précédentes études et ils ont scrupuleusement vérifié la potentielle modification de l'effet du BCG selon le sexe sur les mesures d'association. Aucune de ces analyses n'a révélé de modification de l'effet selon le sexe. La comparaison des résultats des articles 2 et 3 permet également de corroborer nos premières investigations menées dans l'adolescence et de suspecter une différence de pathogénèse et/ou de déterminants du diabète de type 1 avant et après 30 ans.

Les découvertes découlant de cette thèse soulignent l'importance de considérer les phénotypes cliniques dans les études futures portant sur la vaccination au BCG et son lien possible avec la sclérose en plaques ou le diabète.

## 12 Bibliographie

---

1. Antel J, Antel S, Caramanos Z, Arnold DL, Kuhlmann T. Primary progressive multiple sclerosis: part of the MS disease spectrum or separate disease entity? *Acta neuropathologica*. 2012;123(5):627-38.
2. McKay KA, Kwan V, Duggan T, Tremlett H. Risk factors associated with the onset of relapsing-remitting and primary progressive multiple sclerosis: a systematic review. *Biomed Res Int*. 2015;2015:817238.
3. Ricardo A, Cecilia Q, Bárbara E, Leila C, Berenice S, Cecilia P, et al. Clinical and demographic characteristics of primary progressive multiple sclerosis in Argentina: Argentinean registry cohort study (RelevarEM). *Neurol Sci*. 2020.
4. Galea I, Ward-Abel N, Heesen C. Relapse in multiple sclerosis. *Bmj*. 2015;350:h1765.
5. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. *Lancet*. 2018;391(10130):1622-36.
6. Nishimura A, Matsumura K, Kikuno S, Nagasawa K, Okubo M, Mori Y, et al. Slowly Progressive Type 1 Diabetes Mellitus: Current Knowledge And Future Perspectives. *Diabetes, metabolic syndrome and obesity : targets and therapy*. 2019;12:2461-77.
7. Banday MZ, Sameer AS, Nissar S. Pathophysiology of diabetes: An overview. *Avicenna J Med*. 2020;10(4):174-88.
8. Howard J, Trevick S, Younger DS. Epidemiology of Multiple Sclerosis. *Neurol Clin*. 2016;34(4):919-39.
9. Multiple Sclerosis International Federation. Atlas of MS 2013: Mapping Multiple Sclerosis Around the World London2013 [Available from: <http://www.msif.org/about-ms/publications-and-resources/>].
10. Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor BV, et al. Atlas of Multiple Sclerosis 2013: A growing global problem with widespread inequity. *Neurology*. 2014;83(11):1022-4.
11. Thomas NJ, Jones SE, Weedon MN, Shields BM, Oram RA, Hattersley AT. Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank. *The lancet Diabetes & endocrinology*. 2018;6(2):122-9.

12. Fox DA, Islam N, Sutherland J, Reimer K, Amed S. Type 1 diabetes incidence and prevalence trends in a cohort of Canadian children and youth. *Pediatr Diabetes*. 2018;19(3):501-5.
13. Mobasseri M, Shirmohammadi M, Amiri T, Vahed N, Hosseini Fard H, Ghojzadeh M. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *Health promotion perspectives*. 2020;10(2):98-115.
14. Amed S, Islam N, Sutherland J, Reimer K. Incidence and prevalence trends of youth-onset type 2 diabetes in a cohort of Canadian youth: 2002-2013. *Pediatr Diabetes*. 2018;19(4):630-6.
15. Wu H, Ballantyne CM. Metabolic Inflammation and Insulin Resistance in Obesity. *Circulation research*. 2020;126(11):1549-64.
16. Bowers E, Singer K. Obesity-induced inflammation: The impact of the hematopoietic stem cell niche. *JCI Insight*. 2021;6(3).
17. Sawcer S, Franklin RJ, Ban M. Multiple sclerosis genetics. *Lancet Neurol*. 2014;13(7):700-9.
18. Ascherio A. Environmental factors in multiple sclerosis. *Expert Rev Neurother*. 2013;13(12 Suppl):3-9.
19. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol*. 2017;13(1):25-36.
20. Lehmann PV, Rottlaender A, Kuerten S. The autoimmune pathogenesis of multiple sclerosis. *Die Pharmazie*. 2015;70(1):5-11.
21. Lemus HN, Warrington AE, Rodriguez M. Multiple Sclerosis: Mechanisms of Disease and Strategies for Myelin and Axonal Repair. *Neurol Clin*. 2018;36(1):1-11.
22. Saxena A, Martin-Blondel G, Mars LT, Liblau RS. Role of CD8 T cell subsets in the pathogenesis of multiple sclerosis. *FEBS Lett*. 2011;585(23):3758-63.
23. Lassmann H. Pathogenic Mechanisms Associated With Different Clinical Courses of Multiple Sclerosis. *Front Immunol*. 2018;9:3116.
24. Philips JC, Radermecker RP. [Type 1 diabetes: from genetic predisposition to hypothetical environmental triggers]. *Rev Med Liege*. 2012;67(5-6):319-25.
25. Nyaga DM, Vickers MH, Jefferies C, Perry JK, O'Sullivan JM. The genetic architecture of type 1 diabetes mellitus. *Mol Cell Endocrinol*. 2018;477:70-80.

26. Butalia S, Kaplan GG, Khokhar B, Rabi DM. Environmental Risk Factors and Type 1 Diabetes: Past, Present, and Future. *Can J Diabetes*. 2016;40(6):586-93.
27. Carlsson S. Environmental (Lifestyle) Risk Factors for LADA. *Curr Diabetes Rev*. 2019;15(3):178-87.
28. Kowalewicz-Kulbat M, Loch C. BCG and protection against inflammatory and auto-immune diseases. *Expert Rev Vaccines*. 2017;16(7):1-10.
29. Curtis N. Potential role for BCG in treatment of autoimmune diseases. *BMJ*. 2016;355:i6091.
30. Rousseau MC, Parent ME, St-Pierre Y. Potential health effects from non-specific stimulation of the immune function in early age: the example of BCG vaccination. *Pediatric allergy and immunology*. 2008;19(5):438-48.
31. Netea MG, van Crevel R. BCG-induced protection: effects on innate immune memory. *Semin Immunol*. 2014;26(6):512-7.
32. Freyne B, Marchant A, Curtis N. BCG-associated heterologous immunity, a historical perspective: intervention studies in animal models of infectious diseases. *Trans R Soc Trop Med Hyg*. 2015;109(1):52-61.
33. Arts RJW, Moorlag S, Novakovic B, Li Y, Wang SY, Oosting M, et al. BCG Vaccination Protects against Experimental Viral Infection in Humans through the Induction of Cytokines Associated with Trained Immunity. *Cell host & microbe*. 2018;23(1):89-100.e5.
34. Kleinnijenhuis J, van Crevel R, Netea MG. Trained immunity: consequences for the heterologous effects of BCG vaccination. *Trans R Soc Trop Med Hyg*. 2015;109(1):29-35.
35. Kaufmann E, Sanz J, Dunn JL, Khan N, Mendonca LE, Pacis A, et al. BCG Educates Hematopoietic Stem Cells to Generate Protective Innate Immunity against Tuberculosis. *Cell*. 2018;172(1-2):176-90.e19.
36. van der Meer JW, Joosten LA, Riksen N, Netea MG. Trained immunity: A smart way to enhance innate immune defence. *Mol Immunol*. 2015;68(1):40-4.
37. Ahmed A, Rakshit S, Adiga V, Dias M, Dwarkanath P, D'Souza G, et al. A century of BCG: Impact on tuberculosis control and beyond. *Immunological reviews*. 2021;301(1):98-121.
38. Netea MG, Joosten LAB, Latz E, Mills KHG, Natoli G, Stunnenberg HG, et al. Trained immunity: A program of innate immune memory in health and disease. *Science (New York, NY)*. 2016;352(6284):aaf1098-aaf.

39. Singh AK, Netea MG, Bishai WR. BCG turns 100: its nontraditional uses against viruses, cancer, and immunologic diseases. *The Journal of clinical investigation*. 2021;131(11).
40. Ristori G, Faustman D, Matarese G, Romano S, Salvetti M. Bridging the gap between vaccination with Bacille Calmette-Guerin (BCG) and immunological tolerance: the cases of type 1 diabetes and multiple sclerosis. *Curr Opin Immunol*. 2018;55:89-96.
41. Kuhlreiber WM, Tran L, Kim T, Dybala M, Nguyen B, Plager S, et al. Long-term reduction in hyperglycemia in advanced type 1 diabetes: the value of induced aerobic glycolysis with BCG vaccinations. *NPJ vaccines*. 2018;3:23.
42. Kuhlreiber WM, Faustman DL. BCG Therapy for Type 1 Diabetes: Restoration of Balanced Immunity and Metabolism. *Trends Endocrinol Metab*. 2019;30(2):80-92.
43. Arts RJW, Carvalho A, La Rocca C, Palma C, Rodrigues F, Silvestre R, et al. Immunometabolic Pathways in BCG-Induced Trained Immunity. *Cell reports*. 2016;17(10):2562-71.
44. Lippens C, Garnier L, Guyonvarc'h PM, Santiago-Raber ML, Hugues S. Extended Freeze-Dried BCG Instructed pDCs Induce Suppressive Tregs and Dampen EAE. *Front Immunol*. 2018;9:2777.
45. Chevalier MF, Schneider AK, Cesson V, Dartiguenave F, Lucca I, Jichlinski P, et al. Conventional and PD-L1-expressing Regulatory T Cells are Enriched During BCG Therapy and may Limit its Efficacy. *European urology*. 2018;74(5):540-4.
46. de Bree LCJ, Janssen R, Aaby P, van Crevel R, Joosten LAB, Benn CS, et al. The impact of sex hormones on BCG-induced trained immunity. *Journal of leukocyte biology*. 2018;104(3):573-8.
47. Jensen KJ, Larsen N, Biering-Sorensen S, Andersen A, Eriksen HB, Monteiro I, et al. Heterologous immunological effects of early BCG vaccination in low-birth-weight infants in Guinea-Bissau: a randomized-controlled trial. *J Infect Dis*. 2015;211(6):956-67.
48. Smith SG, Kleinnijenhuis J, Netea MG, Dockrell HM. Whole Blood Profiling of Bacillus Calmette–Guérin-Induced Trained Innate Immunity in Infants Identifies Epidermal Growth Factor, IL-6, Platelet-Derived Growth Factor-AB/BB, and Natural Killer Cell Activation. *Frontiers in Immunology*. 2017;8(644).
49. Sparaco M, Bonavita S. The role of sex hormones in women with multiple sclerosis: From puberty to assisted reproductive techniques. *Frontiers in neuroendocrinology*. 2021;60:100889.

50. Assad S, Khan HH, Ghazanfar H, Khan ZH, Mansoor S, Rahman MA, et al. Role of Sex Hormone Levels and Psychological Stress in the Pathogenesis of Autoimmune Diseases. *Cureus*. 2017;9(6):e1315.
51. Golden LC, Voskuhl R. The importance of studying sex differences in disease: The example of multiple sclerosis. *J Neurosci Res*. 2017;95(1-2):633-43.
52. Ramien C, Taenzer A, Lupu A, Heckmann N, Engler JB, Patas K, et al. Sex effects on inflammatory and neurodegenerative processes in multiple sclerosis. *Neurosci Biobehav Rev*. 2016;67:137-46.
53. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. 2016;16(10):626-38.
54. Andersen E, Isager H, Hyllested K. Risk factors in multiple sclerosis: tuberculin reactivity, age at measles infection, tonsillectomy and appendectomy. *Acta Neurol Scand*. 1981;63(2):131-5.
55. Berr C, Puel J, Clanet M, Ruidavets JB, Mas JL, Alperovitch A. Risk factors in multiple sclerosis: a population-based case-control study in Hautes-Pyrenees, France. *Acta Neurol Scand*. 1989;80(1):46-50.
56. Casetta I, Granieri E, Malagu S, Tola MR, Paolino E, Caniatti LM, et al. Environmental risk factors and multiple sclerosis: a community-based, case-control study in the province of Ferrara, Italy. *Neuroepidemiology*. 1994;13(3):120-8.
57. Bansil S, Singhal BS, Ahuja GK, Riise T, Ladiwala U, Behari M, et al. Multiple sclerosis in India: a case-control study of environmental exposures. *Acta Neurol Scand*. 1997;95(2):90-5.
58. Zorzon M, Zivadinov R, Nasuelli D, Dolfini P, Bosco A, Bratina A, et al. Risk factors of multiple sclerosis: a case-control study. *Neurol Sci*. 2003;24(4):242-7.
59. Pekmezovic T, Jarebinski M, Drulovic J. Childhood infections as risk factors for multiple sclerosis: Belgrade case-control study. *Neuroepidemiology*. 2004;23(6):285-8.
60. Farez MF, Correale J. Immunizations and risk of multiple sclerosis: systematic review and meta-analysis. *J Neurol*. 2011;258(7):1197-206.
61. Huppmann M, Baumgarten A, Ziegler AG, Bonifacio E. Neonatal Bacille Calmette-Guerin vaccination and type 1 diabetes. *Diabetes Care*. 2005;28(5):1204-6.
62. Parent ME, Siemiatycki J, Menzies R, Fritschi L, Colle E. Bacille Calmette-Guerin vaccination and incidence of IDDM in Montreal, Canada. *Diabetes Care*. 1997;20(5):767-72.

63. Rousseau MC, El-Zein M, Conus F, Legault L, Parent ME. Bacillus Calmette-Guerin (BCG) Vaccination in Infancy and Risk of Childhood Diabetes. *Paediatric and perinatal epidemiology*. 2016;30(2):141-8.
64. EURODIAB Substudy 2 Study Group. Infections and vaccinations as risk factors for childhood type I (insulin-dependent) diabetes mellitus: a multicentre case-control investigation. *Diabetologia*. 2000;43(1):47-53.
65. Wadsworth EJ, Shield JP, Hunt LP, Baum JD. A case-control study of environmental factors associated with diabetes in the under 5s. *Diabetic Medicine*. 1997;14(5):390-6.
66. Hummel M, Fuchtenbusch M, Schenker M, Ziegler AG. No major association of breast-feeding, vaccinations, and childhood viral diseases with early islet autoimmunity in the German BABYDIAB Study. *Diabetes Care*. 2000;23(7):969-74.
67. Blom L, Nystrom L, Dahlquist G. The Swedish childhood diabetes study. Vaccinations and infections as risk determinants for diabetes in childhood. *Diabetologia*. 1991;34(3):176-81.
68. Legault L, Polychronakos C. Annual incidence of type 1 diabetes in Quebec between 1989-2000 in children. *Clin Invest Med*. 2006;29(1):10-3.
69. Diaz-Valencia PA, Bougnères P, Valleron AJ. Global epidemiology of type 1 diabetes in young adults and adults: a systematic review. *BMC Public Health*. 2015;15:255.
70. Nakhla M, Simard M, Dube M, Larocque I, Plante C, Legault L, et al. Identifying pediatric diabetes cases from health administrative data: a population-based validation study in Quebec, Canada. *Clin Epidemiol*. 2019;11:833-43.
71. Marshall JS, Warrington R, Watson W, Kim HL. An introduction to immunology and immunopathology. *Allergy, Asthma & Clinical Immunology*. 2018;14(2):49.
72. Chaplin DD. Overview of the immune response. *J Allergy Clin Immunol*. 2010;125(2 Suppl 2):S3-S23.
73. Waisman A, Johann L. Antigen-presenting cell diversity for T cell reactivation in central nervous system autoimmunity. *Journal of molecular medicine (Berlin, Germany)*. 2018;96(12):1279-92.
74. Savinov AY, Wong FS, Stonebraker AC, Chervonsky AV. Presentation of antigen by endothelial cells and chemoattraction are required for homing of insulin-specific CD8+ T cells. *J Exp Med*. 2003;197(5):643-56.

75. Moorlag S, Arts RJW, van Crevel R, Netea MG. Non-specific effects of BCG vaccine on viral infections. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2019;25(12):1473-8.
76. Deng Q, Luo Y, Chang C, Wu H, Ding Y, Xiao R. The Emerging Epigenetic Role of CD8+T Cells in Autoimmune Diseases: A Systematic Review. *Frontiers in immunology*. 2019;10:856-.
77. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. *Blood*. 2008;112(5):1570-80.
78. Geginat J, Paroni M, Maglie S, Alfen JS, Kastirr I, Gruarin P, et al. Plasticity of human CD4 T cell subsets. *Front Immunol*. 2014;5:630.
79. Thorsby E. A short history of HLA. *Tissue Antigens*. 2009;74(2):101-16.
80. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med*. 2000;343(10):702-9.
81. Gruen JR, Weissman SM. Human MHC class III and IV genes and disease associations. *Front Biosci*. 2001;6:D960-72.
82. Jokubaitis VG, Butzkueven H. A genetic basis for multiple sclerosis severity: Red herring or real? *Mol Cell Probes*. 2016;30(6):357-65.
83. Rojas M, Restrepo-Jiménez P, Monsalve DM, Pacheco Y, Acosta-Ampudia Y, Ramírez-Santana C, et al. Molecular mimicry and autoimmunity. *J Autoimmun*. 2018;95:100-23.
84. Kaskow BJ, Baecher-Allan C. Effector T Cells in Multiple Sclerosis. *Cold Spring Harbor perspectives in medicine*. 2018;8(4).
85. Clark M, Kroger CJ, Ke Q, Tisch RM. The Role of T Cell Receptor Signaling in the Development of Type 1 Diabetes. *Front Immunol*. 2020;11:615371.
86. Newton K, Dixit VM. Signaling in innate immunity and inflammation. *Cold Spring Harb Perspect Biol*. 2012;4(3):a006049.
87. Cronkite DA, Strutt TM. The Regulation of Inflammation by Innate and Adaptive Lymphocytes. *Journal of immunology research*. 2018;2018:1467538-.
88. Furman D, Campisi J, Verdin E, Carrera-Bastos P, Targ S, Franceschi C, et al. Chronic inflammation in the etiology of disease across the life span. *Nature Medicine*. 2019;25(12):1822-32.

89. Palavra F, Almeida L, Ambrosio AF, Reis F. Obesity and brain inflammation: a focus on multiple sclerosis. *Obes Rev.* 2016;17(3):211-24.
90. Vilarrasa N, San Jose P, Rubio M, Lecube A. Obesity in Patients with Type 1 Diabetes: Links, Risks and Management Challenges. *Diabetes, metabolic syndrome and obesity : targets and therapy.* 2021;14:2807-27.
91. Guerrero-Garcia JJ, Carrera-Quintanar L, Lopez-Roa RI, Marquez-Aguirre AL, Rojas-Mayorquin AE, Ortuno-Sahagun D. Multiple Sclerosis and Obesity: Possible Roles of Adipokines. Mediators of inflammation. 2016;2016:4036232.
92. Wirth K, Klenk J, Brefka S, Dallmeier D, Faehling K, Roqué IFM, et al. Biomarkers associated with sedentary behaviour in older adults: A systematic review. *Ageing Res Rev.* 2017;35:87-111.
93. Favere K, Bosman M, Delputte PL, Favoreel HW, Van Craenenbroeck EM, De Sutter J, et al. A systematic literature review on the effects of exercise on human Toll-like receptor expression. *Exerc Immunol Rev.* 2021;27:84-124.
94. Ruskovska T, Budić-Leto I, Corral-Jara KF, Ajdžanović V, Arola-Arnal A, Bravo FI, et al. Systematic Bioinformatic Analyses of Nutrigenomic Modifications by Polyphenols Associated with Cardiometabolic Health in Humans-Evidence from Targeted Nutrigenomic Studies. *Nutrients.* 2021;13(7).
95. Wild CP. The exposome: from concept to utility. *International journal of epidemiology.* 2012;41(1):24-32.
96. Poppe AY, Wolfson C, Zhu B. Prevalence of multiple sclerosis in Canada: a systematic review. *Can J Neurol Sci.* 2008;35(5):593-601.
97. Beck CA, Metz LM, Svenson LW, Patten SB. Regional variation of multiple sclerosis prevalence in Canada. *Mult Scler.* 2005;11(5):516-9.
98. Warren SA, Svenson LW, Warren KG. Contribution of incidence to increasing prevalence of multiple sclerosis in Alberta, Canada. *Mult Scler.* 2008;14(7):872-9.
99. Marrie RA, Fisk JD, Stadnyk KJ, Yu BN, Tremlett H, Wolfson C, et al. The incidence and prevalence of multiple sclerosis in Nova Scotia, Canada. *Can J Neurol Sci.* 2013;40(6):824-31.
100. Amezcua L, McCauley JL. Race and ethnicity on MS presentation and disease course. *Mult Scler.* 2020;26(5):561-7.

101. Broła W, Sobolewski P, Żak M, Flaga S, Fudala M, Siutka D, et al. Profile of Polish patients with primary progressive multiple sclerosis. *Mult Scler Relat Disord*. 2019;33:33-8.
102. Orton SM, Herrera BM, Yee IM, Valdar W, Ramagopalan SV, Sadovnick AD, et al. Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol*. 2006;5(11):932-6.
103. Bostrom I, Landtblom AM. Does the changing sex ratio of multiple sclerosis give opportunities for intervention? *Acta Neurol Scand*. 2015;132(199):42-5.
104. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012;380(9859):2224-60.
105. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128(1):305-10.
106. Compston A, Coles A. Multiple sclerosis. *Lancet*. 2008;372(9648):1502-17.
107. Xiang Z, Yang Y, Chang C, Lu Q. The epigenetic mechanism for discordance of autoimmunity in monozygotic twins. *J Autoimmun*. 2017.
108. Patsopoulos NA, De Jager PL. Genetic and gene expression signatures in multiple sclerosis. *Mult Scler*. 2020;26(5):576-81.
109. Rivas-Rodriguez E, Amezcua L. Ethnic Considerations and Multiple Sclerosis Disease Variability in the United States. *Neurol Clin*. 2018;36(1):151-62.
110. Wade BJ. Spatial analysis of global prevalence of multiple sclerosis suggests need for an updated prevalence scale. *Mult Scler Int*. 2014;2014:124578.
111. Didonna A, Oksenberg JR. Genetic determinants of risk and progression in multiple sclerosis. *Clin Chim Acta*. 2015;449:16-22.
112. Isobe N, Gourraud PA, Harbo HF, Caillier SJ, Santaniello A, Khankhanian P, et al. Genetic risk variants in African Americans with multiple sclerosis. *Neurology*. 2013;81(3):219-27.
113. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nature genetics*. 2015;47(10):1107-13.
114. Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O, et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am J Hum Genet*. 2004;74(1):160-7.

115. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* 2013;9(11):e1003926.
116. Qiu W, James I, Carroll WM, Mastaglia FL, Kermode AG. HLA-DR allele polymorphism and multiple sclerosis in Chinese populations: a meta-analysis. *Mult Scler.* 2011;17(4):382-8.
117. the International Multiple Sclerosis Genetics C, Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nature Genetics.* 2015;47:1107.
118. Yoshimura S, Isobe N, Yonekawa T, Matsushita T, Masaki K, Sato S, et al. Genetic and infectious profiles of Japanese multiple sclerosis patients. *PloS one.* 2012;7(11):e48592.
119. Briggs FB, Acuna B, Shen L, Ramsay P, Quach H, Bernstein A, et al. Smoking and risk of multiple sclerosis: evidence of modification by NAT1 variants. *Epidemiology.* 2014;25(4):605-14.
120. Wagner M, Sobczynski M, Karabon L, Bilinska M, Pokryszko-Dragan A, Pawlak-Adamska E, et al. Polymorphisms in CD28, CTLA-4, CD80 and CD86 genes may influence the risk of multiple sclerosis and its age of onset. *J Neuroimmunol.* 2015;288:79-86.
121. Mpandzou G, Ait Ben Haddou E, Regragui W, Benomar A, Yahyaoui M. Vitamin D deficiency and its role in neurological conditions: A review. *Rev Neurol (Paris).* 2016;172(2):109-22.
122. Barry EL, Rees JR, Peacock JL, Mott LA, Amos CI, Bostick RM, et al. Genetic Variants in CYP2R1, CYP24A1, and VDR Modify the Efficacy of Vitamin D3 Supplementation for Increasing Serum 25-Hydroxyvitamin D Levels in a Randomized Controlled Trial. *The Journal of Clinical Endocrinology & Metabolism.* 2014;99(10):E2133-E7.
123. Tavakolpour S. Interleukin 7 receptor polymorphisms and the risk of multiple sclerosis: A meta-analysis. *Mult Scler Relat Disord.* 2016;8:66-73.
124. Svejgaard A. The immunogenetics of multiple sclerosis. *Immunogenetics.* 2008;60(6):275-86.
125. Zhang Z, Wang L, Sun X, Zhang L, Lu L. Association of IL4 and IL4R polymorphisms with multiple sclerosis susceptibility in Caucasian population: A meta-analysis. *J Neurol Sci.* 2016;363:107-13.
126. Cavanillas ML, Alcina A, Nunez C, de las Heras V, Fernandez-Arquero M, Bartolome M, et al. Polymorphisms in the IL2, IL2RA and IL2RB genes in multiple sclerosis risk. *Eur J Hum Genet.* 2010;18(7):794-9.

127. Wagner M, Bilinska M, Pokryszko-Dragan A, Sobczynski M, Cyrul M, Kusnierczyk P, et al. ALCAM and CD6--multiple sclerosis risk factors. *J Neuroimmunol.* 2014;276(1-2):98-103.
128. Giménez-Orenga K, Oltra E. Human Endogenous Retrovirus as Therapeutic Targets in Neurologic Disease. *Pharmaceuticals (Basel, Switzerland).* 2021;14(6).
129. Belbasis L, Bellou V, Evangelou E, Ioannidis JP, Tzoulaki I. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. *Lancet Neurol.* 2015;14(3):263-73.
130. Mentis AA, Dardiotis E, Grigoriadis N, Petinaki E, Hadjigeorgiou GM. Viruses and endogenous retroviruses in multiple sclerosis: From correlation to causation. *Acta Neurol Scand.* 2017.
131. Pender MP, Csurhes PA, Burrows JM, Burrows SR. Defective T-cell control of Epstein-Barr virus infection in multiple sclerosis. *Clin Transl Immunology.* 2017;6(1):e126.
132. Makhani N, Banwell B, Tellier R, Yea C, McGovern S, O'Mahony J, et al. Viral exposures and MS outcome in a prospective cohort of children with acquired demyelination. *Mult Scler.* 2016;22(3):385-8.
133. Sundqvist E, Bergstrom T, Daialhosein H, Nystrom M, Sundstrom P, Hillert J, et al. Cytomegalovirus seropositivity is negatively associated with multiple sclerosis. *Mult Scler.* 2014;20(2):165-73.
134. Langer-Gould A, Wu J, Lucas R, Smith J, Gonzales E, Amezcua L, et al. Epstein-Barr virus, cytomegalovirus, and multiple sclerosis susceptibility: A multiethnic study. *Neurology.* 2017;89(13):1330-7.
135. Almohmeed YH, Avenell A, Aucott L, Vickers MA. Systematic review and meta-analysis of the sero-epidemiological association between Epstein Barr virus and multiple sclerosis. *PLoS One.* 2013;8(4):e61110.
136. Myhr KM, Frost P, Gronning M, Midgard R, Kalland KH, Helland DE, et al. Absence of HTLV-1 related sequences in MS from high prevalence areas in western Norway. *Acta Neurol Scand.* 1994;89(1):65-8.
137. Oger J. HTLV-1 infection and the viral etiology of multiple sclerosis. *J Neurol Sci.* 2007;262(1-2):100-4.
138. Wattel E, Mariotti M, Bignon JD, Agis F, Gordien E, Muller JY, et al. No evidence of HTLV-I infection in French patients with multiple sclerosis using the polymerase chain reaction. *J Clin Pathol.* 1991;44(10):871-2.

139. Agut H, Bonnafous P, Gautheret-Dejean A. Update on infections with human herpesviruses 6A, 6B, and 7. *Med Mal Infect.* 2017;47(2):83-91.
140. Leibovitch EC, Jacobson S. Evidence linking HHV-6 with multiple sclerosis: an update. *Curr Opin Virol.* 2014;9:127-33.
141. Kang JH, Sheu JJ, Kao S, Lin HC. Increased risk of multiple sclerosis following herpes zoster: a nationwide, population-based study. *J Infect Dis.* 2011;204(2):188-92.
142. Procaccini C, Pucino V, Mantzoros CS, Matarese G. Leptin in autoimmune diseases. *Metabolism.* 2015;64(1):92-104.
143. Hedstrom AK, Lima Bomfim I, Barcellos L, Gianfrancesco M, Schaefer C, Kockum I, et al. Interaction between adolescent obesity and HLA risk genes in the etiology of multiple sclerosis. *Neurology.* 2014;82(10):865-72.
144. Tsigalou C, Vallianou N, Dalamaga M. Autoantibody Production in Obesity: Is There Evidence for a Link Between Obesity and Autoimmunity? *Current obesity reports.* 2020;9(3):245-54.
145. Cantorna MT, Snyder L, Lin YD, Yang L. Vitamin D and 1,25(OH)<sub>2</sub>D regulation of T cells. *Nutrients.* 2015;7(4):3011-21.
146. Rolf L, Muris AH, Hupperts R, Damoiseaux J. Illuminating vitamin D effects on B-cells - the multiple sclerosis perspective. *Immunology.* 2016.
147. Lucas RM, Ponsonby AL, Dear K, Valery PC, Pender MP, Taylor BV, et al. Sun exposure and vitamin D are independent risk factors for CNS demyelination. *Neurology.* 2011;76(6):540-8.
148. Baarnhielm M, Hedstrom AK, Kockum I, Sundqvist E, Gustafsson SA, Hillert J, et al. Sunlight is associated with decreased multiple sclerosis risk: no interaction with human leukocyte antigen-DRB1\*15. *Eur J Neurol.* 2012;19(7):955-62.
149. Cortese M, Riise T, Bjornevik K, Holmoy T, Kampman MT, Magalhaes S, et al. Timing of use of cod liver oil, a vitamin D source, and multiple sclerosis risk: The EnvIMS study. *Mult Scler.* 2015;21(14):1856-64.
150. Švajger U, Rožman PJ. Recent discoveries in dendritic cell tolerance-inducing pharmacological molecules. *International immunopharmacology.* 2020;81:106275.
151. Weston M, Constantinescu CS. What role does tobacco smoking play in multiple sclerosis disability and mortality? A review of the evidence. *Neurodegener Dis Manag.* 2015;5(1):19-25.

152. Nakahama T, Hanieh H, Nguyen NT, Chinen I, Ripley B, Millrine D, et al. Aryl hydrocarbon receptor-mediated induction of the microRNA-132/212 cluster promotes interleukin-17-producing T-helper cell differentiation. *Proc Natl Acad Sci U S A*. 2013;110(29):11964-9.
153. Handel AE, Williamson AJ, Disanto G, Dobson R, Giovannoni G, Ramagopalan SV. Smoking and multiple sclerosis: an updated meta-analysis. *PloS one*. 2011;6(1):e16149.
154. Nishanth K, Tariq E, Nzvere FP, Miqdad M, Cancarevic I. Role of Smoking in the Pathogenesis of Multiple Sclerosis: A Review Article. *Cureus*. 2020;12(8):e9564.
155. Barragan-Martinez C, Speck-Hernandez CA, Montoya-Ortiz G, Mantilla RD, Anaya JM, Rojas-Villarraga A. Organic solvents as risk factor for autoimmune diseases: a systematic review and meta-analysis. *PloS one*. 2012;7(12):e51506.
156. Procaccini C, Pucino V, De Rosa V, Marone G, Matarese G. Neuro-endocrine networks controlling immune system in health and disease. *Front Immunol*. 2014;5:143.
157. Gustavsen S, Sondergaard HB, Oturai DB, Laursen B, Laursen JH, Magyari M, et al. Shift work at young age is associated with increased risk of multiple sclerosis in a Danish population. *Mult Scler Relat Disord*. 2016;9:104-9.
158. Hedstrom AK, Akerstedt T, Hillert J, Olsson T, Alfredsson L. Shift work at young age is associated with increased risk for multiple sclerosis. *Ann Neurol*. 2011;70(5):733-41.
159. Riise T, Mohr DC, Munger KL, Rich-Edwards JW, Kawachi I, Ascherio A. Stress and the risk of multiple sclerosis. *Neurology*. 2011;76(22):1866-71.
160. Briones-Buixassa L, Milà R, M<sup>a</sup> Aragonès J, Bufill E, Olaya B, Arrufat FX. Stress and multiple sclerosis: A systematic review considering potential moderating and mediating factors and methods of assessing stress. *Health psychology open*. 2015;2(2):2055102915612271.
161. Balbuena LD, Middleton RM, Tuite-Dalton K, Pouliou T, Williams KE, Noble GJ. Sunshine, Sea, and Season of Birth: MS Incidence in Wales. *PloS one*. 2016;11(5):e0155181.
162. Rodríguez Cruz PM, Matthews L, Boggild M, Cavey A, Constantinescu CS, Evangelou N, et al. Time- and Region-Specific Season of Birth Effects in Multiple Sclerosis in the United Kingdom. *Times and Regions of Birth Effects in Multiple Sclerosis*. *JAMA Neurology*. 2016;73(8):954-60.
163. Watad A, Azrielant S, Soriano A, Bracco D, Abu Much A, Amital H. Association between seasonal factors and multiple sclerosis. *Eur J Epidemiol*. 2016.

164. Haase S, Wilck N, Kleinewietfeld M, Muller DN, Linker RA. Sodium chloride triggers Th17 mediated autoimmunity. *J Neuroimmunol.* 2019;329:9-13.
165. Palacios N, Munger KL, Fitzgerald KC, Hart JE, Chitnis T, Ascherio A, et al. Exposure to particulate matter air pollution and risk of multiple sclerosis in two large cohorts of US nurses. *Environment international.* 2017;109:64-72.
166. Bai L, Burnett RT, Kwong JC, Hystad P, van Donkelaar A, Brook JR, et al. Long-term exposure to air pollution and the incidence of multiple sclerosis: A population-based cohort study. *Environmental research.* 2018;166:437-43.
167. Turk Boru U, Boluk C, Tasdemir M, Gezer T, Serim VA. Air pollution, a possible risk factor for multiple sclerosis. *Acta Neurol Scand.* 2020;141(5):431-7.
168. Yuchi W, Sbihi H, Davies H, Tamburic L, Brauer M. Road proximity, air pollution, noise, green space and neurologic disease incidence: a population-based cohort study. *Environmental health : a global access science source.* 2020;19(1):8.
169. Hedstrom AK, Hillert J, Olsson T, Alfredsson L. Nicotine might have a protective effect in the etiology of multiple sclerosis. *Mult Scler.* 2013;19(8):1009-13.
170. Sharif K, Watad A, Bragazzi NL, Adawi M, Amital H, Shoenfeld Y. Coffee and autoimmunity: More than a mere hot beverage! *Autoimmun Rev.* 2017;16(7):712-21.
171. Hedstrom AK, Mowry EM, Gianfrancesco MA, Shao X, Schaefer CA, Shen L, et al. High consumption of coffee is associated with decreased multiple sclerosis risk; results from two independent studies. *J Neurol Neurosurg Psychiatry.* 2016;87(5):454-60.
172. Wijnands JM, Kingwell E. Time to wake up and smell the coffee? Coffee consumption and multiple sclerosis. *J Neurol Neurosurg Psychiatry.* 2016;87(5):453.
173. Wang YJ, Li R, Yan JW, Wan YN, Tao JH, Chen B, et al. The epidemiology of alcohol consumption and multiple sclerosis: a review. *Neurol Sci.* 2015;36(2):189-96.
174. Salzer J, Hallmans G, Nystrom M, Stenlund H, Wadell G, Sundstrom P. Vitamin A and systemic inflammation as protective factors in multiple sclerosis. *Mult Scler.* 2013;19(8):1046-51.
175. Torkildsen O, Loken-Amsrud KI, Wergeland S, Myhr KM, Holmoy T. Fat-soluble vitamins as disease modulators in multiple sclerosis. *Acta Neurol Scand Suppl.* 2013(196):16-23.
176. Sedaghat F, Jessri M, Behrooz M, Mirghotbi M, Rashidkhani B. Mediterranean diet adherence and risk of multiple sclerosis: a case-control study. *Asia Pac J Clin Nutr.* 2016;25(2):377-84.

177. Probstel AK, Baranzini SE. The Role of the Gut Microbiome in Multiple Sclerosis Risk and Progression: Towards Characterization of the "MS Microbiome". *Neurotherapeutics*. 2018;15(1):126-34.
178. Mirza A, Mao-Draayer Y. The gut microbiome and microbial translocation in multiple sclerosis. *Clin Immunol*. 2017.
179. Ochoa-Reparaz J, Kasper LH. The influence of gut-derived CD39 regulatory T cells in CNS demyelinating disease. *Transl Res*. 2017;179:126-38.
180. Kamm CP, Uitdehaag BM, Polman CH. Multiple Sclerosis: Current Knowledge and Future Outlook. *European neurology*. 2014;72(3-4):132-41.
181. Thompson A. Overview of primary progressive multiple sclerosis (PPMS): similarities and differences from other forms of MS, diagnostic criteria, pros and cons of progressive diagnosis. *Mult Scler*. 2004;10 Suppl 1:S2-7.
182. Kuhlmann T, Ludwin S, Prat A, Antel J, Brück W, Lassmann H. An updated histological classification system for multiple sclerosis lesions. *Acta neuropathologica*. 2017;133(1):13-24.
183. Kerlero de Rosbo N, Milo R, Lees MB, Burger D, Bernard CC, Ben-Nun A. Reactivity to myelin antigens in multiple sclerosis. Peripheral blood lymphocytes respond predominantly to myelin oligodendrocyte glycoprotein. *The Journal of clinical investigation*. 1993;92(6):2602-8.
184. Haase S, Linker RA. Inflammation in multiple sclerosis. *Therapeutic advances in neurological disorders*. 2021;14:17562864211007687.
185. Simkins TJ, Duncan GJ, Bourdette D. Chronic Demyelination and Axonal Degeneration in Multiple Sclerosis: Pathogenesis and Therapeutic Implications. *Curr Neurol Neurosci Rep*. 2021;21(6):26.
186. Geladaris A, Häusler D, Weber MS. Microglia: The Missing Link to Decipher and Therapeutically Control MS Progression? *Int J Mol Sci*. 2021;22(7).
187. Hiltensperger M, Korn T. The Interleukin (IL)-23/T helper (Th)17 Axis in Experimental Autoimmune Encephalomyelitis and Multiple Sclerosis. *Cold Spring Harbor perspectives in medicine*. 2018;8(1).
188. Napier J, Rose L, Adeoye O, Hooker E, Walsh KB. Modulating acute neuroinflammation in intracerebral hemorrhage: the potential promise of currently approved medications for multiple sclerosis. *Immunopharmacology and immunotoxicology*. 2019;41(1):7-15.

189. Prod'homme T, Zamvil SS. The Evolving Mechanisms of Action of Glatiramer Acetate. *Cold Spring Harbor perspectives in medicine*. 2019;9(2).
190. Baker D, Amor S. Experimental autoimmune encephalomyelitis is a good model of multiple sclerosis if used wisely. *Mult Scler Relat Disord*. 2014;3(5):555-64.
191. Mockus TE, Munie A, Atkinson JR, Segal BM. Encephalitogenic and Regulatory CD8 T Cells in Multiple Sclerosis and Its Animal Models. *J Immunol*. 2021;206(1):3-10.
192. Häusser-Kinzel S, Weber MS. The Role of B Cells and Antibodies in Multiple Sclerosis, Neuromyelitis Optica, and Related Disorders. *Frontiers in Immunology*. 2019;10(201).
193. Quinn J, Axtell R. Emerging Role of Follicular T Helper Cells in Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis. *International Journal of Molecular Sciences*. 2018;19(10):3233.
194. Lochhead JJ, Yang J, Ronaldson PT, Davis TP. Structure, Function, and Regulation of the Blood-Brain Barrier Tight Junction in Central Nervous System Disorders. *Front Physiol*. 2020;11:914.
195. Savino W, Chaves B, Bonomo AC, Cotta-de-Almeida V. Integrin-directed antibody-based immunotherapy: focus on VLA-4. *Immunotherapy Advances*. 2021;1(1).
196. Baker D, Pryce G, Herrod SS, Schmierer K. Potential mechanisms of action related to the efficacy and safety of cladribine. *Mult Scler Relat Disord*. 2019;30:176-86.
197. Galicia-Garcia U, Benito-Vicente A, Jebari S, Larrea-Sebal A, Siddiqi H, Uribe KB, et al. Pathophysiology of Type 2 Diabetes Mellitus. *Int J Mol Sci*. 2020;21(17).
198. Bell DS, Ovalle F. How long can insulin therapy be avoided in the patient with type 2 diabetes mellitus by use of a combination of metformin and a sulfonylurea? *Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists*. 2000;6(4):293-5.
199. Carlsson S. Etiology and Pathogenesis of Latent Autoimmune Diabetes in Adults (LADA) Compared to Type 2 Diabetes. *Front Physiol*. 2019;10:320.
200. Pieralice S, Pozzilli P. Latent Autoimmune Diabetes in Adults: A Review on Clinical Implications and Management. *Diabetes & metabolism journal*. 2018;42(6):451-64.
201. Buzzetti R, Tuomi T, Mauricio D, Pietropaolo M, Zhou Z, Pozzilli P, et al. Management of Latent Autoimmune Diabetes in Adults: A Consensus Statement From an International Expert Panel. *Diabetes*. 2020;69(10):2037-47.

202. Dickens LT, Thomas CC. Updates in Gestational Diabetes Prevalence, Treatment, and Health Policy. *Curr Diab Rep.* 2019;19(6):33.
203. Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet.* 2009;373(9677):1773-9.
204. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of Type 2 Diabetes - Global Burden of Disease and Forecasted Trends. *Journal of epidemiology and global health.* 2020;10(1):107-11.
205. Liu J, Ren Z-H, Qiang H, Wu J, Shen M, Zhang L, et al. Trends in the incidence of diabetes mellitus: results from the Global Burden of Disease Study 2017 and implications for diabetes mellitus prevention. *BMC Public Health.* 2020;20(1):1415.
206. Huebschmann AG, Huxley RR, Kohrt WM, Zeitler P, Regensteiner JG, Reusch JEB. Sex differences in the burden of type 2 diabetes and cardiovascular risk across the life course. *Diabetologia.* 2019;62(10):1761-72.
207. Mauvais-Jarvis F. Epidemiology of Gender Differences in Diabetes and Obesity. *Adv Exp Med Biol.* 2017;1043:3-8.
208. Ostman J, Lonnberg G, Arnqvist HJ, Blohme G, Bolinder J, Ekblom Schnell A, et al. Gender differences and temporal variation in the incidence of type 1 diabetes: results of 8012 cases in the nationwide Diabetes Incidence Study in Sweden 1983-2002. *J Intern Med.* 2008;263(4):386-94.
209. Ismail L, Materwala H, Al Kaabi J. Association of risk factors with type 2 diabetes: A systematic review. *Computational and structural biotechnology journal.* 2021;19:1759-85.
210. Xia Y, Xie Z, Huang G, Zhou Z. Incidence and trend of type 1 diabetes and the underlying environmental determinants. *Diabetes Metab Res Rev.* 2019;35(1):e3075.
211. Mussa J, Meltzer S, Bond R, Garfield N, Dasgupta K. Trends in National Canadian Guideline Recommendations for the Screening and Diagnosis of Gestational Diabetes Mellitus over the Years: A Scoping Review. *Int J Environ Res Public Health.* 2021;18(4).
212. Feig DS, Hwee J, Shah BR, Booth GL, Bierman AS, Lipscombe LL. Trends in incidence of diabetes in pregnancy and serious perinatal outcomes: a large, population-based study in Ontario, Canada, 1996-2010. *Diabetes Care.* 2014;37(6):1590-6.
213. Shao F, Zheng P, Yu D, Zhou Z, Jia L. Follicular helper T cells in type 1 diabetes. *The FASEB Journal.* 2020;34(1):30-40.

214. Samuel VT, Shulman GI. Mechanisms for insulin resistance: common threads and missing links. *Cell*. 2012;148(5):852-71.
215. Hjort R, Alfredsson L, Andersson T, Carlsson PO, Grill V, Groop L, et al. Family history of type 1 and type 2 diabetes and risk of latent autoimmune diabetes in adults (LADA). *Diabetes & metabolism*. 2017;43(6):536-42.
216. Weires MB, Tausch B, Haug PJ, Edwards CQ, Wetter T, Cannon-Albright LA. Familiality of diabetes mellitus. *Exp Clin Endocrinol Diabetes*. 2007;115(10):634-40.
217. Ornellas F, Carapeto PV, Mandarim-de-Lacerda CA, Aguila MB. Obese fathers lead to an altered metabolism and obesity in their children in adulthood: review of experimental and human studies. *J Pediatr (Rio J)*. 2017;93(6):551-9.
218. Zhang H, Pollin TI. Epigenetics Variation and Pathogenesis in Diabetes. *Curr Diab Rep*. 2018;18(11):121.
219. Rich SS, Erlich H, Concannon P. Genetics of Type 1 Diabetes. In: Cowie CC, Casagrande SS, Menke A, Cissell MA, Eberhardt MS, Meigs JB, et al., editors. *Diabetes in America*. Bethesda MD interest.2018.
220. Rich SS, Weitkamp LR, Barbosa J. Genetic heterogeneity of insulin-dependent (type I) diabetes mellitus: evidence from a study of extended haplotypes. *Am J Hum Genet*. 1984;36(5):1015-23.
221. Redondo MJ, Steck AK, Pugliese A. Genetics of type 1 diabetes. *Pediatric diabetes*. 2018;19(3):346-53.
222. Ilonen J, Kiviniemi M, Lempainen J, Simell O, Toppari J, Veijola R, et al. Genetic susceptibility to type 1 diabetes in childhood - estimation of HLA class II associated disease risk and class II effect in various phases of islet autoimmunity. *Pediatr Diabetes*. 2016;17 Suppl 22:8-16.
223. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2009;41(6):703-7.
224. Frommer L, Kahaly GJ. Type 1 Diabetes and Autoimmune Thyroid Disease-The Genetic Link. *Front Endocrinol (Lausanne)*. 2021;12:618213.
225. Jacobi T, Massier L, Klötting N, Horn K, Schuch A, Ahnert P, et al. HLA Class II Allele Analyses Implicate Common Genetic Components in Type 1 and Non-Insulin-Treated Type 2 Diabetes. *The Journal of clinical endocrinology and metabolism*. 2020;105(3).

226. Florez JC, Udler MS, Hanson RL. Genetics of Type 2 Diabetes. In: Cowie CC, Casagrande SS, Menke A, Cissell MA, Eberhardt MS, Meigs JB, et al., editors. Diabetes in America. Bethesda MD interest.2018.
227. Thomsen SK, Gloyn AL. The pancreatic  $\beta$  cell: recent insights from human genetics. Trends in endocrinology and metabolism: TEM. 2014;25(8):425-34.
228. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet. 2000;26(1):76-80.
229. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006;38(3):320-3.
230. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet. 2008;40(9):1098-102.
231. Principi N, Berioli MG, Bianchini S, Esposito S. Type 1 diabetes and viral infections: What is the relationship? J Clin Virol. 2017;96:26-31.
232. Geravandi S, Liu H, Maedler K. Enteroviruses and T1D: Is It the Virus, the Genes or Both which Cause T1D. Microorganisms. 2020;8(7).
233. Battaglia M, Buckner JH, Levings MK, Richardson SJ, Wong FS, Tree TI. Identifying the 'Achilles heel' of type 1 diabetes. Clin Exp Immunol. 2021;204(2):167-78.
234. Turk Wensveen T, Gašparini D, Rahelić D, Wensveen FM. Type 2 diabetes and viral infection; cause and effect of disease. Diabetes Res Clin Pract. 2021;172:108637.
235. Cardwell CR, Stene LC, Joner G, Davis EA, Cinek O, Rosenbauer J, et al. Birthweight and the risk of childhood-onset type 1 diabetes: a meta-analysis of observational studies using individual patient data. Diabetologia. 2010;53(4):641-51.
236. Harder T, Roepke K, Diller N, Stechling Y, Dudenhausen JW, Plagemann A. Birth weight, early weight gain, and subsequent risk of type 1 diabetes: systematic review and meta-analysis. Am J Epidemiol. 2009;169(12):1428-36.
237. Li S, Zhang M, Tian H, Liu Z, Yin X, Xi B. Preterm birth and risk of type 1 and type 2 diabetes: systematic review and meta-analysis. Obes Rev. 2014;15(10):804-11.

238. Crump C, Sundquist J, Sundquist K. Preterm birth and risk of type 1 and type 2 diabetes: a national cohort study. *Diabetologia*. 2020;63(3):508-18.
239. Waernbaum I, Dahlquist G, Lind T. Perinatal risk factors for type 1 diabetes revisited: a population-based register study. *Diabetologia*. 2019;62(7):1173-84.
240. Metsala J, Hakola L, Lundqvist A, Virta LJ, Gissler M, Virtanen SM. Perinatal factors and the risk of type 1 diabetes in childhood and adolescence-A register-based case-cohort study in Finland, years 1987 to 2009. *Pediatr Diabetes*. 2020.
241. Lee HY, Lu CL, Chen HF, Su HF, Li CY. Perinatal and childhood risk factors for early-onset type 1 diabetes: a population-based case-control study in Taiwan. *Eur J Public Health*. 2015;25(6):1024-9.
242. Maddatu J, Anderson-Baucum E, Evans-Molina C. Smoking and the risk of type 2 diabetes. *Transl Res*. 2017;184:101-7.
243. Hidayat K, Zou SY, Shi BM. The influence of maternal body mass index, maternal diabetes mellitus, and maternal smoking during pregnancy on the risk of childhood-onset type 1 diabetes mellitus in the offspring: Systematic review and meta-analysis of observational studies. *Obes Rev*. 2019;20(8):1106-20.
244. Infante M, Ricordi C, Sanchez J, Clare-Salzler MJ, Padilla N, Fuenmayor V, et al. Influence of Vitamin D on Islet Autoimmunity and Beta-Cell Function in Type 1 Diabetes. *Nutrients*. 2019;11(9).
245. Issa CM. Vitamin D and Type 2 Diabetes Mellitus. *Adv Exp Med Biol*. 2017;996:193-205.
246. Löfvenborg JE, Andersson T, Carlsson PO, Dorkhan M, Groop L, Martinell M, et al. Sweetened beverage intake and risk of latent autoimmune diabetes in adults (LADA) and type 2 diabetes. *European journal of endocrinology*. 2016;175(6):605-14.
247. Qin P, Li Q, Zhao Y, Chen Q, Sun X, Liu Y, et al. Sugar and artificially sweetened beverages and risk of obesity, type 2 diabetes mellitus, hypertension, and all-cause mortality: a dose-response meta-analysis of prospective cohort studies. *Eur J Epidemiol*. 2020;35(7):655-71.
248. Lamb MM, Frederiksen B, Seifert JA, Kroehl M, Rewers M, Norris JM. Sugar intake is associated with progression from islet autoimmunity to type 1 diabetes: the Diabetes Autoimmunity Study in the Young. *Diabetologia*. 2015;58(9):2027-34.
249. Sharif K, Watad A, Coplan L, Amital H, Shoenfeld Y, Afek A. Psychological stress and type 1 diabetes mellitus: what is the link? *Expert Rev Clin Immunol*. 2018;14(12):1081-8.

250. Karavanaki K, Tsoka E, Liacopoulou M, Karayianni C, Petrou V, Pippidou E, et al. Psychological stress as a factor potentially contributing to the pathogenesis of Type 1 diabetes mellitus. *J Endocrinol Invest*. 2008;31(5):406-15.
251. Hackett RA, Steptoe A. Type 2 diabetes mellitus and psychological stress - a modifiable risk factor. *Nat Rev Endocrinol*. 2017;13(9):547-60.
252. Kato M, Noda M, Inoue M, Kadowaki T, Tsugane S. Psychological factors, coffee and risk of diabetes mellitus among middle-aged Japanese: a population-based prospective study in the JPHC study cohort. *Endocr J*. 2009;56(3):459-68.
253. Natella F, Scaccini C. Role of coffee in modulation of diabetes risk. *Nutr Rev*. 2012;70(4):207-17.
254. Kolb H, Martin S, Kempf K. Coffee and Lower Risk of Type 2 Diabetes: Arguments for a Causal Relationship. *Nutrients*. 2021;13(4).
255. Löfvenborg JE, Andersson T, Carlsson PO, Dorkhan M, Groop L, Martinell M, et al. Coffee consumption and the risk of latent autoimmune diabetes in adults--results from a Swedish case-control study. *Diabet Med*. 2014;31(7):799-805.
256. Carlsson S, Midthjell K, Tesfamarian MY, Grill V. Age, overweight and physical inactivity increase the risk of latent autoimmune diabetes in adults: results from the Nord-Trøndelag health study. *Diabetologia*. 2007;50(1):55-8.
257. Smith AD, Crippa A, Woodcock J, Brage S. Physical activity and incident type 2 diabetes mellitus: a systematic review and dose-response meta-analysis of prospective cohort studies. *Diabetologia*. 2016;59(12):2527-45.
258. Nixon SJ, Lewis B. Clarifying the neurobehavioral sequelae of moderate drinking lifestyles and acute alcohol effects with aging. *Int Rev Neurobiol*. 2019;148:39-78.
259. Li XH, Yu FF, Zhou YH, He J. Association between alcohol consumption and the risk of incident type 2 diabetes: a systematic review and dose-response meta-analysis. *Am J Clin Nutr*. 2016;103(3):818-29.
260. Rasouli B, Andersson T, Carlsson PO, Dorkhan M, Grill V, Groop L, et al. Alcohol and the risk for latent autoimmune diabetes in adults: results based on Swedish ESTRID study. *European journal of endocrinology*. 2014;171(5):535-43.
261. Chaudhury A, Duvoor C, Reddy Dendi VS, Kraleti S, Chada A, Ravilla R, et al. Clinical Review of Antidiabetic Drugs: Implications for Type 2 Diabetes Mellitus Management. *Front Endocrinol (Lausanne)*. 2017;8:6.

262. Calabrese CM, Valentini A, Calabrese G. Gut Microbiota and Type 1 Diabetes Mellitus: The Effect of Mediterranean Diet. *Front Nutr.* 2020;7:612773.
263. Liu X, Zhang Y, Wu H, Zhu P, Mo X, Ma X, et al. Intake of polyunsaturated fatty acids and risk of preclinical and clinical type 1 diabetes in children-a systematic review and meta-analysis. *Eur J Clin Nutr.* 2019;73(1):1-8.
264. Schwingshackl L, Hoffmann G, Lampousi AM, Knüppel S, Iqbal K, Schwedhelm C, et al. Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. *Eur J Epidemiol.* 2017;32(5):363-75.
265. Löfvenborg JE, Andersson T, Carlsson PO, Dorkhan M, Groop L, Martinell M, et al. Fatty fish consumption and risk of latent autoimmune diabetes in adults. *Nutr Diabetes.* 2014;4(10):e139.
266. Löfvenborg JE, Ahlqvist E, Alfredsson L, Andersson T, Groop L, Tuomi T, et al. Consumption of red meat, genetic susceptibility, and risk of LADA and type 2 diabetes. *European Journal of Nutrition.* 2021;60(2):769-79.
267. Lebenthal Y, de Vries L, Phillip M, Lazar L. Familial type 1 diabetes mellitus - gender distribution and age at onset of diabetes distinguish between parent-offspring and sib-pair subgroups. *Pediatr Diabetes.* 2010;11(6):403-11.
268. Cardwell CR, Carson DJ, Patterson CC. Parental age at delivery, birth order, birth weight and gestational age are associated with the risk of childhood Type 1 diabetes: a UK regional retrospective cohort study. *Diabet Med.* 2005;22(2):200-6.
269. Szalecki M, Wysocka-Mincewicz M, Ramotowska A, Mazur A, Lisowicz L, Ben-Skowronek I, et al. Epidemiology of type 1 diabetes in Polish children: A multicentre cohort study. *Diabetes Metab Res Rev.* 2018;34(2).
270. Castillo-Reinado K, Maier W, Holle R, Stahl-Pehe A, Baechle C, Kuss O, et al. Associations of area deprivation and urban/rural traits with the incidence of type 1 diabetes: analysis at the municipality level in North Rhine-Westphalia, Germany. *Diabet Med.* 2020;37(12):2089-97.
271. Verbeeten KC, Elks CE, Daneman D, Ong KK. Association between childhood obesity and subsequent Type 1 diabetes: a systematic review and meta-analysis. *Diabet Med.* 2011;28(1):10-8.
272. Virtanen SM, Räsänen L, Aro A, Ylönen K, Lounamaa R, Akerblom HK, et al. Is children's or parents' coffee or tea consumption associated with the risk for type 1 diabetes mellitus in children? Childhood Diabetes in Finland Study Group. *Eur J Clin Nutr.* 1994;48(4):279-85.

273. Padhi S, Nayak AK, Behera A. Type II diabetes mellitus: a review on recent drug based therapeutics. *Biomed Pharmacother.* 2020;131:110708.
274. Lobo N, Brooks NA, Zlotta AR, Cirillo JD, Boorjian S, Black PC, et al. 100 years of *Bacillus Calmette-Guérin* immunotherapy: from cattle to COVID-19. *Nat Rev Urol.* 2021:1-12.
275. Tan SY, Kwok E. Albert Calmette (1863-1933): originator of the BCG vaccine. *Singapore Med J.* 2012;53(7):433-4.
276. Bégué P. Albert Calmette: à propos du 150e anniversaire de sa naissance. *Bulletin de l'Académie Nationale de Médecine.* 2013;197(7):1475-84.
277. Desrosiers G. [Joseph-Albert Baudoin (1875-1962): professor of hygiene]. *Can Bull Med Hist.* 1993;10(2):251-68.
278. Payment P. L'oeuvre d'Armand Frappier: 75 ans de recherche et d'enseignement pour l'amélioration de la santé. 2014.
279. Rousseau MC, Conus F, Li J, Parent ME, El-Zein M. The Quebec BCG Vaccination Registry (1956-1992): assessing data quality and linkage with administrative health databases. *BMC Medical Informatics and Decision Making.* 2014;14(2).
280. Ministère de la Santé. Dates de début et de cessation des principaux programmes de vaccination gratuite au Québec 2021 [Available from: <https://msss.gouv.qc.ca/professionnels/vaccination/piq-programmes-et-noms-commerciaux-des-vaccins/dates-de-debut-et-de-cessation-des-principaux-programmes-de-vaccination-gratuite-au-quebec/>].
281. Lancione S, Alvarez JV, Alsdurf H, Pai M, Zwerling AA. Tracking changes in national BCG vaccination policies and practices using the BCG World Atlas. *BMJ Glob Health.* 2022;7(1).
282. Rousseau M-C, Conus F, Kâ K, El-Zein M, Parent M-É, Menzies D. *Bacillus Calmette-Guérin (BCG)* vaccination patterns in the province of Québec, Canada, 1956–1974. *Vaccine.* 2017;35:4777-84.
283. Bright MR, Curtis N, Messina NL. The role of antibodies in *Bacille Calmette Guérin*-mediated immune responses and protection against tuberculosis in humans: A systematic review. *Tuberculosis (Edinb).* 2020:101947.
284. Mangtani P, Abubakar I, Ariti C, Beynon R, Pimpin L, Fine PE, et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis.* 2014;58(4):470-80.

285. Abubakar I, Pimpin L, Ariti C, Beynon R, Mangtani P, Sterne JA, et al. Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette-Guérin vaccination against tuberculosis. *Health Technol Assess*. 2013;17(37):1-372, v-vi.
286. Nguipdop-Djomo P, Heldal E, Rodrigues LC, Abubakar I, Mangtani P. Duration of BCG protection against tuberculosis and change in effectiveness with time since vaccination in Norway: a retrospective population-based cohort study. *Lancet Infect Dis*. 2016;16(2):219-26.
287. Schoenmakers A, Mieras L, Budiawan T, van Brakel WH. The State of Affairs in Post-Exposure Leprosy Prevention: A Descriptive Meta-Analysis on Immuno- and Chemo-Prophylaxis. *Res Rep Trop Med*. 2020;11:97-117.
288. Biering-Sørensen S, Aaby P, Lund N, Monteiro I, Jensen KJ, Eriksen HB, et al. Early BCG-Denmark and Neonatal Mortality Among Infants Weighing <2500 g: A Randomized Controlled Trial. *Clin Infect Dis*. 2017;65(7):1183-90.
289. Arts RJW, Moorlag S, Novakovic B, Li Y, Wang SY, Oosting M, et al. BCG Vaccination Protects against Experimental Viral Infection in Humans through the Induction of Cytokines Associated with Trained Immunity. *Cell host & microbe*. 2018;23(1):89-100 e5.
290. Podder I, Bhattacharya S, Mishra V, Sarkar TK, Chandra S, Sil A, et al. Immunotherapy in viral warts with intradermal Bacillus Calmette-Guerin vaccine versus intradermal tuberculin purified protein derivative: A double-blind, randomized controlled trial comparing effectiveness and safety in a tertiary care center in Eastern India. *Indian J Dermatol Venereol Leprol*. 2017;83(3):411.
291. Stögerer T, Stäger S. Innate Immune Sensing by Cells of the Adaptive Immune System. *Frontiers in Immunology*. 2020;11.
292. Fabié A, Mai LT, Dagenais-Lussier X, Hammami A, van Grevenynghe J, Stäger S. IRF-5 Promotes Cell Death in CD4 T Cells during Chronic Infection. *Cell reports*. 2018;24(5):1163-75.
293. Silva-Barrios S, Smans M, Duerr CU, Qureshi ST, Fritz JH, Descoteaux A, et al. Innate Immune B Cell Activation by *Leishmania donovani* Exacerbates Disease and Mediates Hypergammaglobulinemia. *Cell reports*. 2016;15(11):2427-37.
294. Sfakianos JP, Salome B, Daza J, Farkas A, Bhardwaj N, Horowitz A. Bacillus Calmette-Guerin (BCG): Its fight against pathogens and cancer. *Urol Oncol*. 2021;39(2):121-9.
295. Mi Y, Han J, Zhu J, Jin T. Role of the PD-1/PD-L1 Signaling in Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis: Recent Insights and Future Directions. *Mol Neurobiol*. 2021;58(12):6249-71.

296. Faustman DL. TNF, TNF Inducers and TNFR2 Agonists: A New Path to Type 1 Diabetes Treatment. *Diabetes Metab Res Rev.* 2018;34(1).
297. Faustman DL, Wang L, Okubo Y, Burger D, Ban L, Man G, et al. Proof-of-concept, randomized, controlled clinical trial of Bacillus-Calmette-Guerin for treatment of long-term type 1 diabetes. *PLoS ONE [Electronic Resource].* 2012;7(8):e41756.
298. Faustman D, Davis M. TNF receptor 2 pathway: drug target for autoimmune diseases. *Nat Rev Drug Discov.* 2010;9(6):482-93.
299. Arts RJW, Joosten LAB, Netea MG. The Potential Role of Trained Immunity in Autoimmune and Autoinflammatory Disorders. *Frontiers in Immunology.* 2018;9(298).
300. Bennett JM, Reeves G, Billman GE, Sturmberg JP. Inflammation-Nature's Way to Efficiently Respond to All Types of Challenges: Implications for Understanding and Managing "the Epidemic" of Chronic Diseases. *Front Med (Lausanne).* 2018;5:316.
301. Chen L, Deng H, Cui H, Fang J, Zuo Z, Deng J, et al. Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget.* 2017;9(6):7204-18.
302. Manuck TA, Rice MM, Bailit JL, Grobman WA, Reddy UM, Wapner RJ, et al. Preterm neonatal morbidity and mortality by gestational age: a contemporary cohort. *Am J Obstet Gynecol.* 2016;215(1):103 e1- e14.
303. Rousseau M-C, El-Zein M, Conus F, Benedetti A, Parent M-E. Cohort Profile: The Québec Birth Cohort on Immunity and Health (QBCIH). *International journal of epidemiology.* 2018:1-10.
304. Institut de la Statistique du Québec. *Méthodologie de l'appariement probabiliste à l'Institut de la statistique du Québec.* 2010.
305. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. *Int J Epidemiol.* 2016;45(3):954-64.
306. Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M, et al. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics.* 2001;108(2):E35.
307. Li J, Menzies D, Landry JS, Benedetti A, Rousseau MC. Determinants of Bacillus Calmette-Guerin (BCG) vaccination among Quebec children. *Preventive medicine.* 2014;66(Sept):87-94.

308. Al-Sakran LH, Marrie RA, Blackburn DF, Knox KB, Evans CD. Establishing the Incidence and Prevalence of Multiple Sclerosis in Saskatchewan. *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques*. 2018;45(3):295-303.
309. Marrie RA, O'Mahony J, Maxwell C, Ling V, Yeh EA, Arnold DL, et al. Incidence and prevalence of MS in children: A population-based study in Ontario, Canada. *Neurology*. 2018;91(17):e1579-e90.
310. Marrie RA, Yu N, Blanchard J, Leung S, Elliott L. The rising prevalence and changing age distribution of multiple sclerosis in Manitoba. *Neurology*. 2010;74(6):465-71.
311. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol*. 2011;69(2):292-302.
312. Sadaka Y, Verhey LH, Shroff MM, Branson HM, Arnold DL, Narayanan S, et al. 2010 McDonald criteria for diagnosing pediatric multiple sclerosis. *Ann Neurol*. 2012;72(2):211-23.
313. RxTx, Therapeutic choices, Multiple Sclerosis [Internet]. Canadian pharmacist association. 2019 [cited 2019-03-06]. Available from: [https://www.myrxtx.ca/print/new/documents/CHAPTER/fr/multiple\\_sclerosis](https://www.myrxtx.ca/print/new/documents/CHAPTER/fr/multiple_sclerosis).
314. Méridor M, Sylvestre MP, Marrie RA, Rousseau MC. Does age at diagnosis influence the use of health services for multiple sclerosis? *Mult Scler Relat Disord*. 2020;46:102555.
315. Guttmann A, Nakhla M, Henderson M, To T, Daneman D, Cauch-Dudek K, et al. Validation of a health administrative data algorithm for assessing the epidemiology of diabetes in Canadian children. *Pediatric diabetes*. 2010;11(2):122-8.
316. James RC, Blanchard JF, Campbell D, Clottey C, Osei W, Svenson LW, et al. A model for non-communicable disease surveillance in Canada: the prairie pilot diabetes surveillance system. *Chronic diseases in Canada*. 2004;25(1):7-12.
317. Blanchard JF, Ludwig S, Wajda A, Dean H, Anderson K, Kendall O, et al. Incidence and prevalence of diabetes in Manitoba, 1986-1991. *Diabetes Care*. 1996;19(8):807-11.
318. Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes cases from administrative data: a population-based validation study. *BMC health services research*. 2018;18(1):316.
319. Lix LM, Ayles J, Bartholomew S, Cooke CA, Ellison J, Emond V, et al. The Canadian Chronic Disease Surveillance System: A model for collaborative surveillance. *International journal of population data science*. 2018;3(3):433.

320. Rajkumar V, Levine SN. Latent Autoimmune Diabetes. StatPearls. Treasure Island FL: © 2021, StatPearls Publishing LLC.; 2021.
321. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53(282):457-81.
322. Allison P. Survival Analysis Using SAS: A Practical Guide, Second Edition: SAS Publishing; 2010.
323. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B (Methodological). 1972;34(2):187-220.
324. Therneau TM, Grambsch PM. The Cox Model. Modeling Survival Data: Extending the Cox Model. New York, NY: Springer New York; 2000. p. 39-77.
325. Sai LMR, Stedman. A SAS® Macro for Covariate Specification in Linear, Logistic, or Survival Regression 2017 [Available from: <https://support.sas.com/resources/papers/proceedings17/1223-2017.pdf>].
326. Cox DR, Snell EJ. A General Definition of Residuals. Journal of the Royal Statistical Society Series B (Methodological). 1968;30(2):248-75.
327. Lin DY, Wei LJ, Ying Z. Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. Biometrika. 1993;80(3):557-72.
328. Schoenfeld D. Partial Residuals for The Proportional Hazards Regression Model. Biometrika. 1982;69(1):239-41.
329. Grambsch PM, Therneau TM. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. Biometrika. 1994;81(3):515-26.
330. Therneau TM, Grambsch PM. Testing Proportional Hazards. Modeling Survival Data: Extending the Cox Model. New York, NY: Springer New York; 2000. p. 127-52.
331. In J, Lee DK. Survival analysis: part II - applied clinical data analysis. Korean J Anesthesiol. 2019;72(5):441-57.
332. Le Bourdais C, Piché V. La démographie Québécoise, chapitre 7, Un siècle d'immigration au Québec : de la peur à l'ouverture.2003.
333. Pampalon R, Hamel D, Gamache P, Philibert MD, Raymond G, Simpson A. An area-based material and social deprivation index for public health in Quebec and Canada. Can J Public Health. 2012;103(8 Suppl 2):S17-22.

334. Gamache P, Hamel D, PAMPALON R. L'indice de défavorisation matérielle et sociale: en bref. *INSPQ*; 2015.
335. Pampalon R, Hamel D, Gamache P, Philibert MD, Raymond G, Simpson A. An area-based material and social deprivation index for public health in Québec and Canada. *Canadian journal of public health = Revue canadienne de sante publique*. 2012;103(8 Suppl 2):S17-S22.
336. Pampalon R, Hamel D, Gamache P. Indice de défavorisation matérielle et sociale du Québec: suivi méthodologique de 1991 à 2006: Institut national de santé publique du Québec; 2010.
337. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615-25.
338. Hernán MA, Robins JM. *Causal Inference: What If*: Boca Raton: Chapman & Hall/CRC; 2020.
339. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol*. 2008;8:70.
340. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodol*. 2015;15:30.
341. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*. 2007;16(3):219-42.
342. Harel O, Mitchell EM, Perkins NJ, Cole SR, Tchetgen Tchetgen EJ, Sun B, et al. Multiple Imputation for Incomplete Data in Epidemiologic Studies. *Am J Epidemiol*. 2018;187(3):576-84.
343. Perkins NJ, Cole SR, Harel O, Tchetgen Tchetgen EJ, Sun B, Mitchell EM, et al. Principled Approaches to Missing Data in Epidemiologic Studies. *Am J Epidemiol*. 2018;187(3):568-75.
344. Cottrell G, Cot M, Mary JY. L'imputation multiple des données manquantes aléatoirement : concepts généraux et présentation d'une méthode Monte-Carlo. *Revue d'Épidémiologie et de Santé Publique*. 2009;57(5):361-72.
345. Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol*. 2009;60:549-76.
346. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45(3):1 - 67.

347. Loulergue P, Pulcini C, Massin S, Bernhard M, Fonteneau L, Levy-Bruhl D, et al. Validity of self-reported vaccination status among French healthcare students. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2014;20(12):O1152-4.
348. Wijnands JMA, Zhu F, Kingwell E, Zhao Y, Evans C, Fisk JD, et al. Prodrome in relapsing-remitting and primary progressive multiple sclerosis. *Eur J Neurol*. 2019;26(7):1032-6.
349. Kavaliunas A, Manouchehrinia A, Danylaite Karrenbauer V, Gyllensten H, Glaser A, Alexanderson K, et al. Income in Multiple Sclerosis Patients with Different Disease Phenotypes. *PloS one*. 2017;12(1):e0169460.
350. Kiselev I, Bashinskaya V, Baulina N, Kozin M, Popova E, Boyko A, et al. Genetic differences between primary progressive and relapsing-remitting multiple sclerosis: The impact of immune-related genes variability. *Mult Scler Relat Disord*. 2019;29:130-6.
351. Rakshit S, Ahmed A, Adiga V, Sundararaj BK, Sahoo PN, Kenneth J, et al. BCG revaccination boosts adaptive polyfunctional Th1/Th17 and innate effectors in IGRA+ and IGRA- Indian adults. *JCI Insight*. 2019;4(24).
352. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-79.
353. Culpepper WJ, Marrie RA, Langer-Gould A, Wallin MT, Campbell JD, Nelson LM, et al. Validation of an algorithm for identifying MS cases in administrative health claims datasets. *Neurology*. 2019;92(10):e1016-e28.
354. Ristori G, Buzzi MG, Sabatini U, Giugni E, Bastianello S, Viselli F, et al. Use of Bacille Calmette-Guerin (BCG) in multiple sclerosis. *Neurology*. 1999;53(7):1588-9.
355. Ristori G, Romano S, Cannoni S, Visconti A, Tinelli E, Mendozzi L, et al. Effects of Bacille Calmette-Guerin after the first demyelinating event in the CNS. *Neurology*. 2014;82(1):41-8.
356. Stratos K, McGarragle K, Thistle J, Vyas MV, Lee L. Non-compliance with disease modifying therapies in patients with Multiple Sclerosis: A qualitative analysis. *Mult Scler Relat Disord*. 2020;41:102016.
357. Kern DM, Cepeda MS. Treatment patterns and comorbid burden of patients newly diagnosed with multiple sclerosis in the United States. *BMC Neurol*. 2020;20(1):296.

358. La Mantia L, Munari LM, Lovati R. Glatiramer acetate for multiple sclerosis. The Cochrane database of systematic reviews. 2010(5):Cd004678.
359. Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). *Ann Neurol*. 1996;39(3):285-94.
360. Paty DW, Li DK. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group. *Neurology*. 1993;43(4):662-7.
361. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. The IFNB Multiple Sclerosis Study Group. *Neurology*. 1993;43(4):655-61.
362. Coles AJ, Twyman CL, Arnold DL, Cohen JA, Confavreux C, Fox EJ, et al. Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. *Lancet*. 2012;380(9856):1829-39.
363. Gold R, Kappos L, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K, et al. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. *N Engl J Med*. 2012;367(12):1098-107.
364. Kappos L, Radue EW, O'Connor P, Polman C, Hohlfeld R, Calabresi P, et al. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *N Engl J Med*. 2010;362(5):387-401.
365. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med*. 2006;354(9):899-910.
366. O'Connor P, Wolinsky JS, Confavreux C, Comi G, Kappos L, Olsson TP, et al. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. *N Engl J Med*. 2011;365(14):1293-303.
367. INESSS. *Extrait d'avis au ministre-REBIF* 1998 [updated 1998-01-07. Available from: <https://www.inesss.qc.ca/thematiques/medicaments/medicaments-evaluation-aux-fins-dinscription/extrait-davis-au-ministre/rebif-511.html?DemandePluginController%5Bonglet%5D=0&cHash=00ffdd9fa17d8655f8ba588589587dc1>.

368. Peretti-Watel P, Cortaredona S, Ly EY, Seror V, Ndiaye S, Gaye I, et al. Determinants of childhood immunizations in Senegal: Adding previous shots to sociodemographic background. *Hum Vaccin Immunother.* 2019;1-8.
369. Sheikh N, Sultana M, Ali N, Akram R, Mahumud RA, Asaduzzaman M, et al. Coverage, Timelines, and Determinants of Incomplete Immunization in Bangladesh. *Trop Med Infect Dis.* 2018;3(3).
370. Hu Y, Chen Y, Liang H, Wang Y. An Overview of Coverage of BCG Vaccination and Its Determinants Based on Data from the Coverage Survey in Zhejiang Province. *Int J Environ Res Public Health.* 2018;15(6).
371. Hajizadeh M. Decomposing socioeconomic inequality in child vaccination in the Gambia, the Kyrgyz Republic and Namibia. *Vaccine.* 2019;37(44):6609-16.
372. Farzad F, J AR, Yamamoto E, Hamajima N. Socio-economic and demographic determinants of full immunization among children of 12-23 months in Afghanistan. *Nagoya J Med Sci.* 2017;79(2):179-88.
373. Berg-Hansen P, Celius EG. Socio-economic factors and immigrant population studies of multiple sclerosis. *Acta Neurol Scand.* 2015;132(199):37-41.
374. Bjernevik K, Riise T, Benjaminsen E, Celius EG, Dahl OP, Kampman MT, et al. Level of education and multiple sclerosis risk over a 50-year period: Registry-based sibling study. *Mult Scler.* 2017;23(2):213-9.
375. Montgomery SM, Lambe M, Olsson T, Ekbom A. Parental age, family size, and risk of multiple sclerosis. *Epidemiology.* 2004;15(6):717-23.
376. Bergamaschi R, Cortese A, Pichiecchio A, Berzolari FG, Borrelli P, Mallucci G, et al. Air pollution is associated to the multiple sclerosis inflammatory activity as measured by brain MRI. *Mult Scler.* 2018;24(12):1578-84.
377. Hader WJ, Yee IM. The prevalence of familial multiple sclerosis in Saskatoon, Saskatchewan. *Mult Scler Int.* 2014;2014:545080.
378. Harirchian MH, Fatehi F, Sarraf P, Honarvar NM, Bitarafan S. Worldwide prevalence of familial multiple sclerosis: A systematic review and meta-analysis. *Mult Scler Relat Disord.* 2018;20:43-7.
379. Parnell GP, Booth DR. The Multiple Sclerosis (MS) Genetic Risk Factors Indicate both Acquired and Innate Immune Cell Subsets Contribute to MS Pathogenesis and Identify Novel Therapeutic Opportunities. *Front Immunol.* 2017;8:425.

380. Fernandez-Menendez S, Fernandez-Moran M, Fernandez-Vega I, Perez-Alvarez A, Villafani-Echazu J. Epstein-Barr virus and multiple sclerosis. From evidence to therapeutic strategies. *J Neurol Sci.* 2016;361:213-9.
381. Fragoso YD, Stoney PN, McCaffery PJ. The evidence for a beneficial role of vitamin A in multiple sclerosis. *CNS drugs.* 2014;28(4):291-9.
382. Abdolahi M, Yavari P, Honarvar NM, Bitarafan S, Mahmoudi M, Saboor-Yaraghi AA. Molecular Mechanisms of the Action of Vitamin A in Th17/Treg Axis in Multiple Sclerosis. *J Mol Neurosci.* 2015;57(4):605-13.
383. Radenkovic M, Uvebrant K, Skog O, Sarmiento L, Avartsson J, Storm P, et al. Characterization of resident lymphocytes in human pancreatic islets. *Clinical & Experimental Immunology.* 2017;187(3):418-27.
384. Miller KM, Hart PH, de Klerk NH, Davis EA, Lucas RM. Are low sun exposure and/or vitamin D risk factors for type 1 diabetes? *Photochem Photobiol Sci.* 2017;16(3):381-98.
385. Piescik-Lech M, Chmielewska A, Shamir R, Szajewska H. Systematic Review: Early Infant Feeding and the Risk of Type 1 Diabetes. *J Pediatr Gastroenterol Nutr.* 2017;64(3):454-9.
386. Leong I. BCG vaccination for type 1 diabetes mellitus. *Nat Rev Endocrinol.* 2018.
387. Kleinnijenhuis J, Quintin J, Preijers F, Benn CS, Joosten LA, Jacobs C, et al. Long-lasting effects of BCG vaccination on both heterologous Th1/Th17 responses and innate trained immunity. *Journal of innate immunity.* 2014;6(2):152-8.
388. Gregory E, Crouse DL, Krishnamoorthy P, Legault L. Comparisons of Type 1 and 2 Diabetes Socioeconomic Characteristics in a Montreal Pediatric Clinic. *Canadian Journal of Diabetes.* 2010;34(4):340-5.
389. Chakhtoura M, Azar ST. The role of vitamin d deficiency in the incidence, progression, and complications of type 1 diabetes mellitus. *Int J Endocrinol.* 2013;2013:148673.
390. Patterson CC, Gyurus E, Rosenbauer J, Cinek O, Neu A, Schober E, et al. Seasonal variation in month of diagnosis in children with type 1 diabetes registered in 23 European centers during 1989-2008: little short-term influence of sunshine hours or average temperature. *Pediatr Diabetes.* 2015;16(8):573-80.
391. Tosur M, Redondo MJ. Heterogeneity of Type 1 Diabetes: The Effect of Ethnicity. *Curr Diabetes Rev.* 2018;14(3):266-72.

392. Peet A, Kool P, Ilonen J, Knip M, Tillmann V. Birth weight in newborn infants with different diabetes-associated HLA genotypes in three neighbouring countries: Finland, Estonia and Russian Karelia. *Diabetes Metab Res Rev.* 2012;28(5):455-61.
393. Chen G, Khan N, Walker R, Quan H. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract.* 2010;89(2):189-95.
394. Canada Go. Canadian Chronic Disease Surveillance System (CCDSS) infobase 2016 [Available from: <https://health-infobase.canada.ca/ccdss/data-tool/Age?G=00&V=1&M=5>].
395. García-Pérez LE, Alvarez M, Dilla T, Gil-Guillén V, Orozco-Beltrán D. Adherence to therapies in patients with type 2 diabetes. *Diabetes Ther.* 2013;4(2):175-94.
396. Ward LM, Pendergrass DB, Chima CC, Thorpe RJ, Jr., Bruce MA, Blackburn HF, et al. Access to medications among adults with type 2 diabetes using pharmacy- or clinic-based medication assistance programs: A systematic review. *J Am Pharm Assoc (2003).* 2020;60(6):e411-e21.
397. Corsenac P, Parent M, Benedetti A, Richard H, Stäger S, Rousseau MC. Association between Bacillus Calmette-Guerin vaccination and type 1 diabetes in adolescence: A population-based birth cohort study in Quebec, Canada. *Prev Med.* 2022;154:106893.
398. Castillo-Reinado K, Maier W, Holle R, Stahl-Pehe A, Baechle C, Kuss O, et al. Associations of area deprivation and urban/rural traits with the incidence of type 1 diabetes: analysis at the municipality level in North Rhine-Westphalia, Germany. *Diabet Med.* 2020.
399. Mattsson K, Jonsson I, Malmqvist E, Larsson HE, Rylander L. Maternal smoking during pregnancy and offspring type 1 diabetes mellitus risk: accounting for HLA haplotype. *Eur J Epidemiol.* 2015;30(3):231-8.
400. Nyaga DM, Vickers MH, Jefferies C, Perry JK, O'Sullivan JM. The genetic architecture of type 1 diabetes mellitus. *Mol Cell Endocrinol.* 2018.
401. Okada H, Kuhn C, Feillet H, Bach JF. The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clin Exp Immunol.* 2010;160(1):1-9.
402. Liu B, Xiang Y, Liu Z, Zhou Z. Past, present and future of latent autoimmune diabetes in adults. *Diabetes Metab Res Rev.* 2020;36(1):e3205.
403. Hjort R, Löfvenborg JE, Ahlqvist E, Alfredsson L, Andersson T, Grill V, et al. Interaction Between Overweight and Genotypes of HLA, TCF7L2, and FTO in Relation to the Risk of Latent Autoimmune Diabetes in Adults and Type 2 Diabetes. *The Journal of clinical endocrinology and metabolism.* 2019;104(10):4815-26.

404. Chen W, Chen X, Zhang M, Huang Z. The association of human leukocyte antigen class II (HLA II) haplotypes with the risk of Latent autoimmune diabetes of adults (LADA): Evidence based on available data. *Gene*. 2021;767:145177.
405. Kesmodel US. Information bias in epidemiological studies with a special focus on obstetrics and gynecology. *Acta Obstet Gynecol Scand*. 2018;97(4):417-23.
406. Petráš M, Lesná IK, Dáňová J, Čelko AM. Can Vaccination Trigger Autoimmune Disorders? A Meta-Analysis. *Vaccines (Basel)*. 2021;9(8).
407. de Paulsen N. The regulatory gap: Off-label drug use in Canada. *U Toronto Fac L Rev*. 2005;63:183.
408. Hostalek U, Campbell I. Metformin for diabetes prevention: update of the evidence base. *Curr Med Res Opin*. 2021;37(10):1705-17.
409. Magzoub R, Kheirelseid EAH, Perks C, Lewis S. Does metformin improve reproduction outcomes for non-obese, infertile women with polycystic ovary syndrome? Meta-analysis and systematic review. *Eur J Obstet Gynecol Reprod Biol*. 2022;271:38-62.
410. Song Y, Wang H, Zhu Z, Huang H. Effects of Metformin and Exercise in Polycystic Ovary Syndrome: Systematic Review and Meta-Analysis. *Horm Metab Res*. 2021;53(11):738-45.