The association between environmental factors and the risk of paediatric inflammatory bowel disease

Michael Elten

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Epidemiology

School of Epidemiology and Public Health
Faculty of Medicine
University of Ottawa

Acknowledgements

Thank you to my co-supervisors, Dr. Éric Lavigne and Dr. Eric Benchimol. My professional accomplishments highlighted in this thesis would not have been possible without your continual guidance and support. Your mentorship has led me to many triumphs over the course of this degree, and you have nurtured in me a love and joy of conducting good and relevant research. Thank you also to Dr. Deshayne Fell who comprised the third member of my thesis advisory committee. You made many meaningful contributions to this work, and to my personal growth as a researcher.

To my colleagues in the MSc. Epidemiology program, the Exposure Assessment Section at Health Canada, and those that shared the student room with me at ICES uOttawa – thank you for your collaboration and conversation. You made coming into work everyday interesting and engaging. Best of luck for the future.

Lastly, thanks to my family and friends for supporting me through this entire process. You helped me celebrate the successes, and cope with the defeats that have happened along this journey. To my parents: thank you for your unconditional love and support – you have always inspired me to challenge myself and aim higher. Your sacrifice made this a reality. To Brydie: thank you for your love and for always being there for me. You have helped me be persistently positive, a quality that has been invaluable throughout this chapter of my life, and one will serve me well in the future.

Thank you all for your contributions, this achievement is as much yours as it is mine.

Table of Contents

List of Tables

List of Figures

List of Supplementary Tables and Figures

Preface

Student's contribution:

Michael Elten was the primary author of all content included in this thesis. Study design was done in collaboration with Drs. Éric Lavigne and Eric Benchimol. The cohort was assembled by Michael Elten and Glenys Smith. All analyses were conducted by Michael Elten with critical feedback from Drs. Éric Lavigne, Eric Benchimol, and Deshayne Fell. All authors reviewed the final version of the respective manuscripts presented within this thesis.

Approvals to conduct this research:

This thesis project has been approved by the Ottawa Health Science Network Research Ethics Board (Protocol ID: 20180021-01H), and the Health Canada / Public Health Agency of Canada Research Ethics Board. This project was also approved by an ICES Privacy Officer.

Abbreviations

BORN	Better Outcomes and Registry Network
CANUE	Canadian Urban Environmental Health Research Consortium
CD	Crohn's disease
CIHI	Canadian Institute for Health Information
CI	Confidence interval
DAD	Discharge Abstract Database
DAG	Directed acyclic graph
DLNM	Distributed lag non-linear model
GA	Gestational age
GI	Gastrointestinal
HR	Hazard ratio
IBD	Inflammatory bowel disease
IBD-U	Inflammatory bowel disease – type unclassifiable
IKN	ICES key number
IRR	Incident rate ratio
NAPS	National Air Pollution Surveillance
NDVI	Normalized difference vegetation index
NO_2	Nitrogen dioxide
NO_x	Nitrogen oxides
NPV	Negative predictive value
O_3	Ozone
OCCC	Ontario Crohn's and Colitis Cohort
OHIP	Ontario Health Insurance Program
OR	Odds ratio
O_x	Oxidant capacity
$PM_{2.5}$	Particulate matter with diameter ≤ 2.5 μm
PM_{10}	Particulate matter with diameter ≤ 10 μm
PPV	Positive predictive value
RECORD	REporting of studies Conducted using Observational Routintely-collected Data
RPDB	Registered Persons Database
Th1	T helper type 1
Th2	T helper type 2
TNFα	Tumor necrosis factor α
UC	Ulcerative colitis
VIF	Variance inflation factor
VOC	Volatile Organic Compound
WHO	World Health Organization

Abstract

Background: The incidence of inflammatory bowel disease (IBD) in children is increasing. Environmental factors related to industrialization, including increased levels of air pollution and reduced greenspace may be partially responsible.

Methods: Health administrative data from Ontario were used to identify a cohort of children, and identify those diagnosed with IBD before the age of 18 years. Modelled exposures to air pollution (nitrogen dioxide, fine particulate matter, ozone, and oxidant capacity) and greenspace (through the normalized difference vegetation index) were obtained based on residential six-digit postal code.

Results: Exposure to oxidant capacity during childhood and mid-pregnancy was associated with increased risk of IBD <18 years. No association for other pollutants was found. Residential greenspace during childhood was protective of paediatric-onset IBD, but no association with in-utero exposures was found.

Conclusion: We observed significant relationships between environmental factors and paediatric-onset IBD. The novel finding of a protective effect of greenspace should be explored further.

Chapter 1: Background

1.1 Overview of inflammatory bowel disease

Inflammatory bowel disease (IBD) is a chronic immune-mediated condition comprised of two main subtypes: Crohn's disease (CD) and ulcerative colitis (UC).[1] IBD is characterized by inflammation and ulceration of the gastrointestinal (GI) tract that is caused by an inappropriate response of the host immune system.[2] IBD can lead to a variety of unpleasant and potentially debilitating symptoms including severe abdominal pain, rectal bleeding, fatigue, and vomiting. Some serious consequences of these primary symptoms that warrant concern include unintended weight loss, and anemia. More severe complications can affect other organs such as the liver, skin, eyes, and bones.[3]

CD and UC share similar risk genes,[4] but differences between the subtypes exist. Most notably, CD can affect the entire GI tract while UC is typically localised to the colon. While both subtypes can cause severe complications including hospitalization, surgery, infections and death, UC patients are typically diagnosed more quickly due to their presentation with bloody diarrhea. CD patients can often go months without seeking medical treatment or without being diagnosed due to the indolent nature of their symptoms. Due to their many similarities, it can sometimes be difficult to determine which of the two main subtypes are present in an IBD patient when only the colon is affected. In such cases, occurring in about 5-10% of patients, a diagnosis of IBD-type unclassified (IBD-U) is given.[1]

There is currently no cure for IBD; it is a chronic disease that must be managed with ongoing medication throughout a patient's life in close consultation with a gastroenterologist. The disease course of those with IBD typically cycles between disease flare-up (where the symptoms are severe) and remission (where inflammation and symptoms tend to be dormant and patients experience a normal quality of life). Treatments are commonly prescribed to try to induce and maintain a state of disease remission. Several classes of medications can be given with varying success and side effects, including corticosteroids, immunomodulators, 5-aminosalicylates, and biological therapies. The choice and success of a treatment relies on many factors including disease subtype, patient characteristics, and cost. In cases where medications are not effective at managing disease course, surgical interventions are often recommended as a last resort. Recent figures show 10-year surgery risks of 46.6% and 15.6% after diagnosis of CD and UC respectively.[5]

Those with IBD are at risk for having lower health-related quality of life when compared to the general population.[6] This is particularly true among those for whom treatments are not effective, as the greatest predictor of quality of life is IBD disease severity, even when other chronic conditions are present.[7] However, even when the disease is well managed, IBD patients can experience emotional distress that affects quality of life.[8] Some IBD-specific concerns that lead to this psychological stress include worries about loss of bowel control, fatigue, and fear of sexual inadequacy amongst many others.[9] In addition, patients with IBD are at increased risk for mental illness, especially anxiety, depression and substrance-related disorders.[10]

The typical age of onset of IBD is between adolescence and early adulthood, although IBD can be diagnosed at any age. Paediatric-onset IBD is of particular concern, as there are additional health and psychosocial considerations that negatively impacts quality of life.[11] Though childhood-onset IBD shares many susceptible genes with the adult variant,[12] phenotypically, IBD diagnosed in childhood tends to affect the bowel more extensively and may be more severe.[13] Not only are the typical symptoms worse, but there are also additional considerations when IBD manifests in adolescence or earlier. For instance, a well documented symptom in children with IBD is reduced linear growth.[14,15] This is before considering the educational and career-related challenges that are also experienced by children diagnosed with IBD during this vulnerable developmental stage.[16]

1.2 Epidemiology of IBD

IBD first rose to prominence in Western nations around the time of industrialization.[17] More recent studies have shown that this rise in incidence has appeared to plateau (especially in adults), but there has been a concerning increase in IBD incidence among middle- and low-income nations in Asia, Africa and South America.[17–19] Canada in particular is among the countries with the highest reported prevalence and incidence rates of IBD.[17,20] A recent report showed that there are about 270,000, or over 7 out of every 1000 Canadians living with IBD. Projections suggest that by 2030, this number is expected to jump to 400,000 or about 1% of Canadians.[21] Although incidence rates for adult IBD have generally stabilized in high-income nations, paediatric-onset IBD incident rates have continued to rise, particularly in early-onset

cases in children less than 5 years of age.[22] In fact, it has been reported that Ontario has one of the highest rates of paediatric IBD worldwide.[23]

The increasing prevalence of IBD in Canada is associated with significant direct and indirect healthcare costs. The direct total cost to treat all Canadians with IBD in 2018 was estimated at CAD$1.28 billion using data from a Manitoba study (inflation adjusted annual cost of CAD$4,731 per person multiplied by the 270,000 estimated Canadians with IBD in 2018) but calculations based on data from other provinces suggest this figure could be even higher, surpassing the CAD$2 billion mark.[24–26] Major sources of these expenditures were medications, surgery, hospitalizations and diagnostic procedures, which are usually conducted by specialists. The indirect costs of IBD are also notable, and are estimated to have about an equal value to the direct costs of IBD - about CAD$1.29 billion,[21] largely stemming from absenteeism, early retirement, premature mortality, and out of pocket costs.[27,28] It should be noted that this value is based on many assumptions and figures from other countries, though it also ignores other sources of indirect costs such as presenteeism, caregiver costs, and impacts to professional development.[29]

Childhood-onset cases of IBD can be especially costly; as IBD is a chronic condition, afflicted children will likely have to go through lifelong treatments and appointments with specialists in order to help manage their symptoms. Further, impacts to education can have long term consequences to individual earnings,[30] and affected children are more likely to need caregivers, which adds to the increased indirect costs.[31]

The demographics of IBD diagnosis in several provinces in Canada have been previously explored.[21] In terms of age at diagnosis, CD and UC were most likely to be diagnosed in the second and third decades of life. Among children, CD was more common than UC, whereas among seniors the opposite was true. In terms of biological sex, females were slightly more likely to be diagnosed with CD than males, but the proportion of men and women diagnosed with UC was about equal. Males tend to be more commonly diagnosed with UC and CD during childhood.

IBD etiology is complex and is an area of active investigation. The prevailing theory is some external environmental factors act to alter the intestinal microbiome in a genetically-predisposed person, triggering an abnormal and dysregulated immune response, which leads to the chronic inflammation seen in the inflammatory bowel diseases. This is depicted in Figure 1.

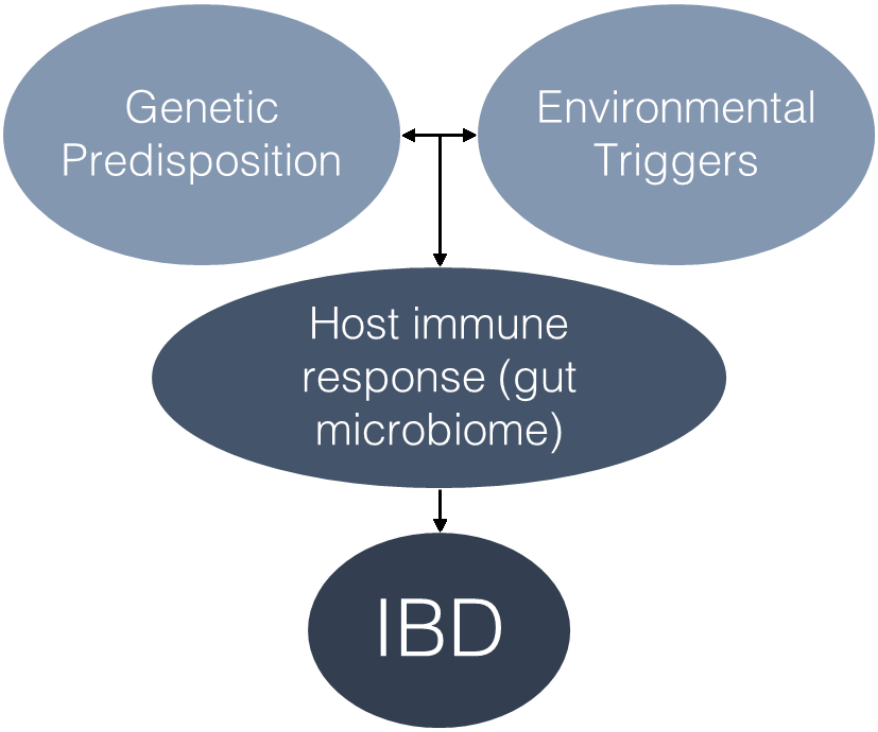


Figure 1. Overview of elements in IBD etiology

The first studies to identify a specific genetic link with IBD were published in 2001, and showed an association between mutations in the NOD2 gene and Crohn's disease.[32,33] Since then, there have been many more studies aimed at exploring the genetic components of IBD, with close to 200 genetic mutations found to be associated with development of CD and UC, many of which are common to both subtypes.[34–36] Many of the loci identified as risk factors for IBD are also common to other immune-mediated conditions such as type 1 diabetes, ankylosing spondylitis and psoriasis,[34] suggesting that there may be shared mechanisms and disease pathways among them.

The NOD2 gene that was first linked to Crohn's disease plays an important role in regulating the inflammatory response to bacterial triggers. This supports the importance of the gut microbiome in IBD etiology. Many other studies since then have provided further evidence for the importance of the gut microbiome.[37] Patients with IBD have shown reduced biodiversity of bacteria living in the gut, as well as relative decreases in bacteria with anti-inflammatory properties and increases in bacteria with inflammatory properties.[38] Children with severe UC also showed a reduced richness, evenness and biodiversity of the gut microbiome when compared to healthy controls.[39] Reduced biodiversity in the gut microbiome in IBD vs controls have been shown in other studies as well.[40–42]

Evidence for in-utero colonization of the gut microbiome is very limited, so it is likely the early life period in which the gut microbiome is established.[43] This may be a susceptible period to environmental exposures for the developing child, and may subsequently represent an important period of modulating risk to IBD.

1.3 Environmental factors related to IBD

Despite the increased effort to discover specific genetic risk factors for IBD, the magnitude of associations reported remain weak. This combined with the low rates of concordance among monozygotic twins, 10-15% for UC and 30-35% for CD,[44] suggest that environmental factors are an important component of IBD etiology, and can be targeted as a way to potentially reduce the incidence of IBD.

The geographic pattern of IBD rates which saw high-income nations experience increases in IBD incidence in the 20th century, while middle- and low-income nations are only now seeing similar increases, suggests that perhaps a factor related to Westernization or industrialization contributes to IBD incidence. Urbanization is another important consequence of economic development and has led to a drastic shift in the types of environmental exposures faced by the population. The potential contributions of a wide range of environmental factors have been reviewed comprehensively in the literature,[45,46] so only factors that might be related to features of Westernization and industrial activity will be discussed here.

1.3.1 Smoking

Cross-sectional survey data from the United States show that cigarette smoking came into prominence in the 1900's, around the time that the initial increased incidence of IBD was being reported.[47] Plausiblility of a potential relationship is supported by a toxicological study showed that chronic smoke exposure altered gut bacterial activity and structure, as well as gene expression in mice.[48]

The association between active smoking and IBD in adults has been studied extensively. Many studies have shown that the effects differ across the disease subtypes. In a meta-analysis of 15 studies of active smoking, it was reported that being a current smoker was positively associated with CD (pooled OR: 1.76, 95% CI: 1.40-2.22), and being a former smoker was positively associated with UC (pooled OR: 1.79, 95% CI: 1.37-2.34).[49] However, being a current smoker demonstrated an inverse association with UC, with all but one study reporting an OR less than 1 (pooled OR: 0.58, 95% CI: 0.45-0.75). These findings show that there may be important differences in the environmental risk factors across disease subtypes.

The literature seems less conclusive on the association between passive (or second-hand) smoking and risk of IBD. One study showed that children exposed to smoke exposure in-utero had increased odds of both CD (OR: 1.72, 95% CI: 1.10-2.71), and UC (OR: 1.53, 95% CI: 0.93-2.49), while those exposed during childhood had increased odds of CD (OR: 2.04, 95% CI: 1.28-3.31) but not UC (OR: 1.04, 95% CI: 0.66-1.64).[50] A meta-analysis of the effects of prenatal passive smoking on IBD did not find a significant association with both CD (pooled OR: 1.10, 95% CI: 0.67-1.80) and UC (pooled OR: 1.11, 95% CI: 0.63-1.97), although there was notable heterogeneity between the included studies.[51] In this same meta-analysis, passive smoking during childhood was also not associated with either CD (pooled OR: 1.10, 95% CI: 0.92-1.30) or UC (pooled OR: 1.01, 95% CI: 0.85-1.20).[51]

1.3.2 The hygiene hypothesis

A popular theory on the emergence of the group of immune-mediated conditions that include IBD is that they are related to the trend towards more sterile environments and activities, known as the "hygiene hypothesis".[52] It is thought that since human evolution occurred at a time where exposure to many microbes and parasites was common, the immune system has developed to respond to such infections. The sudden removal of these microbes from the environment may lead to an immune system that is poorly regulated and overactive. The evidence base for the hygiene hypothesis for paediatric-onset IBD remains mixed. Hygienic conditions can be difficult to measure and quantify, so proxy indicators such as the home environment and exposure to domestic animals are often used.

Several studies have been conducted examining the effects of factors such as presence of bathrooms, hot water, water drainage, septic tank usage, bedroom sharing and household crowding.[53–56] Contrary to what one would expect given the hygiene hypothesis, bedroom sharing (which would make a child more prone to infections) was found to be associated with an increased risk of both CD and UC across two studies.[53,57] Personal towel use was found to be protective of CD, also contrary to the hygiene hypothesis.[55] Evidence that supports the hygiene hypothesis includes studies that found that the use of tap water was associated with risk of CD,[53] that owning a pet cat was protective of CD,[56] and that having a larger family size and greater household crowding were both protective of CD.[55,56]

1.3.3 Rurality of residence

Another important change that occurred during the 20th century in high-income nations is the shift towards a more urban population.[58] More recently, middle- and low-income countries have mirrored this population shift, and projections suggest that these countries will continue this trend of urbanization, with Asia and Africa representing close to 90% of the projected increase in urban populations by 2050.[59] There is large overlap with the hygiene hypothesis; urban areas tend to have a lower diversity of microbes from sources such as agriculture and farm animals. However, there are many other important differences in lifestyles and environmental exposures between rural and urban residents that cannot be attributed to hygiene and that may also be responsible for potential differences in risk of developing IBD.

Residing in an urban or a rural area may have important consequences on the gut microbiome, and on the subsequent risk of IBD.[60] There have not been any studies to investigate urban vs rural differences in gut microbiome in those with the same genetic background. However, some studies that have compared the gut microbiome of one rural population with another urban population have noted that, in general, populations that reside in non-Western and/or rural areas have a higher biodiversity of bacterial species.[61–64]

There have been a number of studies investigating the effects of rural and urban residence on the risk of IBD. A systematic review and meta-analysis showed that living in urban areas was associated with an increased risk of UC [pooled incidence rate ratio (IRR): 1.17, 95% confidence interval (CI): 1.03-1.32] and an increased risk of CD

(IRR: 1.42, 95% CI: 1.26-1.60).[65] The authors of this meta-analysis reported a high amount of heterogeneity among the included studies, which suggests that the effects of urban living may not be consistent across all regions. A more recent study evaluated the effects of rural/urban residence in a Canadian population and found that the protective association of having a rural residence was strongest among those diagnosed at less than 10 years (IRR: 0.58, 95% CI: 0.43-0.73) and those 10-17.9 years (IRR: 0.72, 95% CI: 0.64-0.81).[66]

Exposures to microbes attained by working or living near farms may help explain these associations. A study in Manitoba, Canada showed that living on a farm was a protective factor for CD [odds ratio (OR): 0.62, 95% CI: 0.46-0.85].[56] Similarly, exposure to farm animals in the first year of life was protective of both childhood-onset CD (OR: 0.5, 95% CI: 0.3-0.9) and childhood-onset UC (OR: 0.4, 95% CI: 0.2-0.8) in a study from Germany.[67]

1.3.4 Air Pollution and IBD

Exposures to ambient air pollution increased due to industrial activities and widespread use of combustion engines in high-income countries in the 20th century. Though better regulation and technology has helped reduce pollution levels in high-income countries in recent years, the concentrations of air pollutants have continued to rise in middle- to low-income countries, whose mean concentrations are often in excess of the World Health Organization (WHO) guidelines.[68]

Polluted air may negatively impact the gut microbiome. Research conducted on mice showed that exposure to ambient air pollutants alters the gut microbial composition, intestinal permeability, and incites an inflammatory response.[69] The harmful changes to the gut microbiome can be spurred by ingested particulate matter, which may also decrease colonic motility and clearance, and alter the metabolic function of the gut microbiome.[70]

A case-control study conducted by Kaplan et al in 2010 investigated this potential relationship and found that there were increased odds of CD for those in the higher quintiles of nitrogen dioxide (NO_2) concentration in a subgroup of those up to 23 years of age (OR= 2.31, 95% CI: 1.25-4.28).[71] They also noted that there were increased odds of UC among those aged up to 25 years due to increased exposure to sulfur dioxide (SO_2), OR: 2.00, 95% CI: 1.08-3.72. However these associations were not present for adult-onset IBD. Another study by Opstelten et al. showed that increased traffic intensity on major roads within 100m of one's home was associated with increased odds of IBD in adults (OR: 1.60, 95% CI: 1.04-2.46), but there were also inverse associations reported for exposures to particulate matter (PM) in this study.[72] As nitrogen dioxide is often used as a proxy indicator for the group of traffic-related air pollutants, there is some consistency across the two studies, but both report several null associations for particulate matter of various sizes, and when other measures of traffic were used. Associations between air pollution and other immune-mediated conditions such as type 1 diabetes and asthma have also been reported.[57,73–77]

1.3.5 Greenspace and IBD

Greenspace is an emerging exposure of interest in environmental epidemiology. Greenspace can impact human health through many different pathways; physical activity, psychological improvements, social connectedness and exposure to increased biodiversity are some of the more prominent mechanisms through which living near greenspace can influence health. Urbanization has led to a decrease in plant diversity, caused mainly by anthropogenic activities, which may have been a contributing factor to the increase in IBD incidence that was noted in the high-income regions.[78]

The gut microbiome may also be impacted through many different pathways facilitated by greenspace. Vegetation is a major source of biodiversity in urban areas, and the associated exposure to a wide range of microbes might be essential to maintaining a healthy gut microbiome.[79] The physical activity that is associated with residential greenspace may also have independent effects on gut microbiome diversity.[80,81]

There have not been any studies published on the associations between greenspace/vegetation and the risk of IBD in children or adults. However, relationships between residential greenspace and other immune-mediated conditions have been documented. For instance, studies have shown that residential greenspace was associated with a reduced risk of atopic sensitization and both childhood asthma prevalence and incidence.[82–84]

1.3.6 Diet

Diet is another potential modifiable factor that can influence the risk of developing IBD. Changes spurred by industrialization and urbanization of the population have led to numerous shifts in diet, leading to what is commonly known as “the Western diet”.[85] The urban diet tends to comprise food higher in fat, more animal products, refined sugars, and processed food made further away from the home.[86] The diet can directly impact the gut microbiome and lead to changes in the diversity of different bacterial species. One study showed that children from a rural African village, where the diet is high in fiber content, had a significantly different gut microbiome composition when compared to European children, and attributed these differences due to differences in diet.[62]

Associations of many different elements of the diet with IBD have been noted in several epidemiological analyses with conflicting results.[54,87–89] In general, there seems to be a positive association between IBD risk and higher consumption of sugars/sweeteners,[87,88] fat and oil consumption,[88] and red meat consumption.[89] Fruits and vegetables, on the other hand, seem to have a negative association with IBD risk.[87] An issue in these studies is that the methodology varies considerably and exposure ascertainment is usually via food recall questionnaires which can be prone to many biases. Meta-analyses have shown that these trends are only somewhat consistent, with evidence strongest for the relationship between dietary meat consumption and increased risk of IBD.[90,91]

1.3.7 Directed acyclic graph (DAG):

A challenge in investigating the association of various environmental risk factors with IBD is that they often tend to co-vary, so it can be difficult to assign an observed effect to a single environmental factor. In order to try and visualize how these and other environmental factors interrelate, and to try and aid in determining a sufficient statistical model to control the possible confounding of other factors, a directed acyclic graph (DAG) was created (Figure 2).

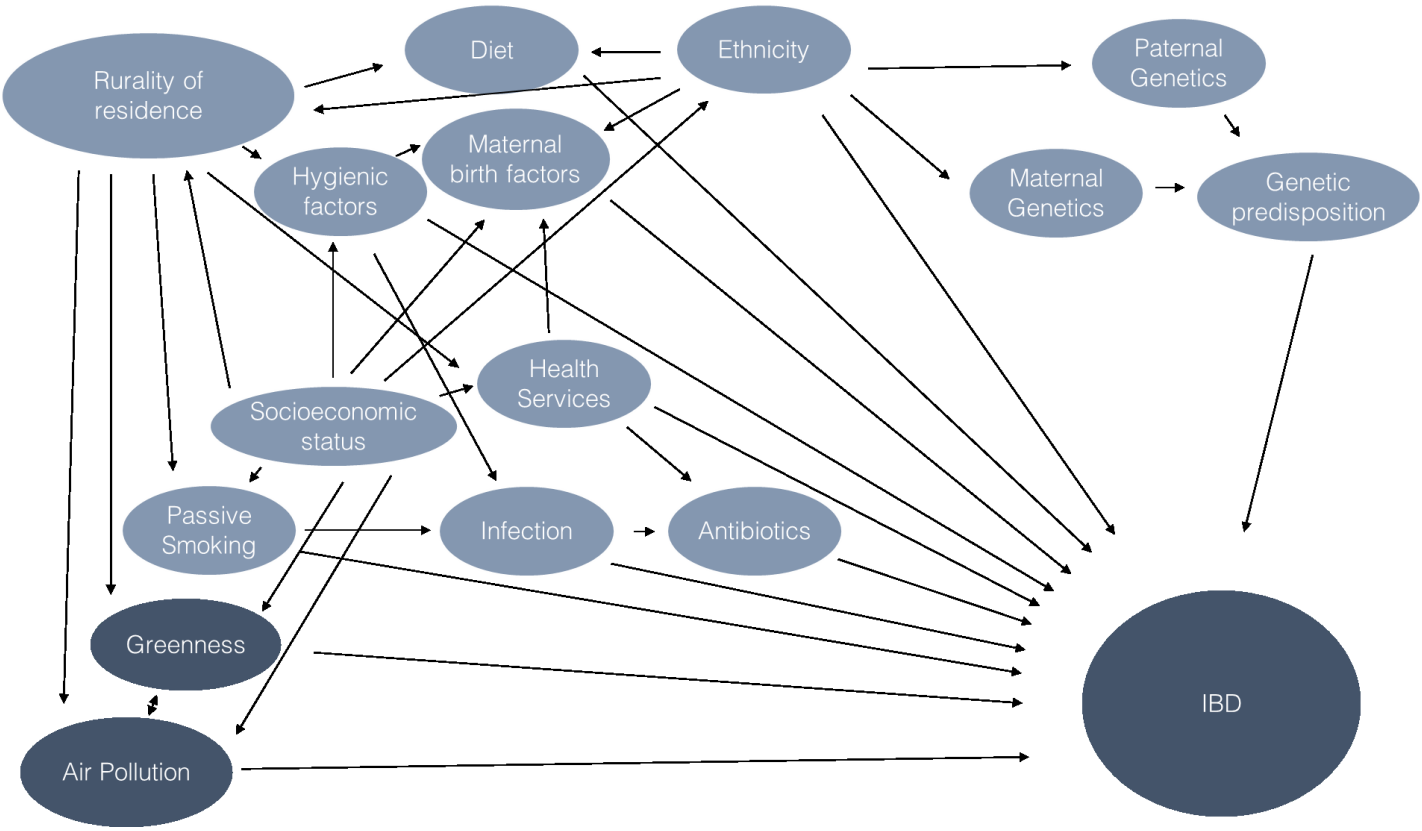


Figure 2. Example directed acyclic graph for IBD

1.4 Ambient air pollution

Ambient air pollution is the heterogenous mixture of harmful substances and chemicals present in the outdoor air that we breathe. Though most human exposure is through inhalation, there are also secondary pathways such as food contamination. Chronic and acute exposures to air pollutants have been associated with many physiological changes in the human body, most notably systemic inflammation, that lead to increased mortality.[92–95] As there is a large suite of air pollutants that are spatially correlated and likely to be exerting both independent and synergistic effects on health, elucidating the specific pathways relating these exposures to specific health outcomes is one of the greatest challenges faced by air pollution researchers. This ubiquity of effect on many different physiological pathways also means that there may be many health conditions related to air pollution, outside of the traditional respiratory and cardiovascular outcomes, that warrant attention. Of relevance to a potential relationship with IBD is the contribution of air pollution towards the epidemic of immune system-mediated chronic diseases.[77]

In general, Canada has relatively low levels of air pollution, with concentrations of pollutants generally on a downward trend, except for ozone in some regions.[68,96] Despite this, low-level concentrations may still be harmful, especially when the exposure is over a long period of time. Additionally, pollutants from other countries can travel through the atmosphere and affect the health of Canadians.[97] Health Canada estimates that 14,600 premature deaths can be attributed to NO_2, $PM_{2.5}$ and O_3 from anthropogenic sources every year.[98] There are also an additional 35 million acute

respiratory symptom days, 2.69 million asthma symptom days and 8,000 emergency room visit days due to these air pollutants every year. This leaves an estimated economic cost of $114 billion every year due to outdoor air pollution.[98]

These values were calculated using Canadian air pollution exposure data which are internationally valued for the ability to provide evidence on the health effects of chronic, low-level exposures. These robust exposure data are used in this thesis to assess both prenatal and postnatal exposures to ambient air pollution and their associated risk of paediatric IBD in Chapter 3. For this study, three pollutants: nitrogen dioxide (NO_2), particulate matter with diameter ≤ 2.5 μm ($PM_{2.5}$), and ozone (O_3), as well as a composite measure of oxidant capacity (O_x) were measured and evaluated. A brief overview of each of these pollutants is provided below.

1.4.1 Nitrogen dioxide (NO_2)

Nitrogen dioxide is a gaseous chemical that is primarily emitted from traffic-related sources, so much so that it is often considered a proxy for all traffic-related air pollution in health research studies. The concentration of NO_2 tends to decay fairly rapidly from the point source, so there is considerable spatial and temporal variation of this pollutant. When exposed to the sun, there is a photochemical reaction that occurs to convert a small amount of NO_2 to O_3. This may be important as it tends to lead to a slight negative correlation between the two pollutant concentrations.

In order to assess NO_2 in this thesis, data from a national land use regression (LUR) model were obtained. This model combined information from National Air Pollution

Surveillance (NAPS) fixed-ground monitors with additional variables known to predict NO_2 concentrations including road length within 10 kilometres, area of industrial land use within 2 kilometres, and mean summer rainfall. Subsequently, a kernel density function was applied to describe the density of roadways, in order to capture localised variations in the emissions from vehicles.

1.4.2 Particulate matter with a diameter ≤ 2.5 μm ($PM_{2.5}$)

$PM_{2.5}$, also known as fine particulate matter, describes the fraction of suspended atmospheric particles that are up to 2.5 μm in diameter. $PM_{2.5}$ is the result of both naturally occurring and man-made activity. Forest fires, transportation and residential combustion are the three largest sources of this pollutant.[99] The sources, and consequently, the composition of $PM_{2.5}$ can have a strong temporal and geographical variance.[100,101] Recent studies have suggested that the toxicity of particulate matter exposure may depend on more than just the overall mass concentration, but also on specific physical and chemical properties.[102,103]

In order to model exposures to $PM_{2.5}$ in this thesis, the overall mass of all particles was considered, as information on the source or other qualities of particulate matter was not available at a province-wide scale. Concentrations were estimated through the use of a satellite-derived estimate (available at a 1x1 kilometre resolution), that was augmented with data from ground-level monitors.[104]

1.4.3 Ozone (O_3)

Atmospheric ozone is a gaseous pollutant that is dissolved in the air. This secondary pollutant is formed by reactions of nitrogen oxides (NO_x, which include NO_2) and volatile organic compounds (VOCs) that occur in the presence of sunlight. The sources of NO_x and VOCs are largely from anthropogenic activities such as gasoline burning and industrial sources, though some VOCs are produced from natural sources. Due to the need for sunlight for O_3 to form, the concentrations tend to show a strong seasonal variation, with concentrations much higher in the summer than in the winter.

To characterize O_3 exposures in these analyses, a surface representing an average of daily 8-hour maximum concentrations in the warm seasons (May 1st to October 31st) at a 21km grid resolution was used. This surface, described in greater detail below, was created using an optimal interpolation technique.[105]

1.4.4 Oxidant capacity (O_x)

The oxidant capacity is a measure of pollution of recent interest that considers concentrations of both NO_2 and O_3 into a single summary measure. In this thesis, the redox-weighted equation for O_x was used, which weights the contribution of O_3 higher than that of NO_2 due to its higher oxidative capacity. The calculation is shown in Equation 1.

Equation 1. Redox-reweighted oxidant capacity calculation

$$Ox = \frac{(1.07 \times NO2) + (2.075 \times O3)}{3.145} \qquad (1)$$

There are several advantages in using this measure over considering each separately; firstly, both NO_2 and O_3 cause oxidative stress in the body, and their combined effects may be greater than the sum of their individual effects.[106,107] Secondly, exposure to the two pollutants tends to be correlated, due to a photochemical reaction that converts NO_2 to O_3 in the presence of sunlight. By combining the two into a single estimate, the statistical issues such as inflated standard errors, etc. that tend to arise from the collinearity caused by introducing both exposures in the same regression model can be avoided.

1.5 Greenspace

Exposure to greenspace has been known to confer many important health benefits,[108] such as reduced rates of obesity,[109] improved mental health,[110] better birth outcomes,[111] improved cardiovascular outcomes,[112] and lower overall mortality.[113] There are many implicated pathways that can mediate the health impacts of greenspace. These include: reduced exposure to air pollutants, noise mitigation, heat and humidity regulation, increased social support, increased levels of physical activity, or psychological stress reductions.[81,114] As is the case for air pollution exposures, it can be a challenge in epidemiological studies to isolate the effects of one specific mechanism, as they tend to all be present in some facet. It is very possible that several of the aforementioned pathways act simultaneously to influence health.

Although there may be several important aspects of vegetation that should be considered such as species, density, and access, exposure to plants in

epidemiological studies is typically represented as a crude overall exposure to greenspace. The Normalized Difference Vegetation Index (NDVI) is the most commonly used measurement of greenness. It represents one's overall exposure to green vegetation as viewed from a satellite. The advantage of using a measure such as NDVI to quantify exposure to vegetation is that in using a single numerical estimate available via satellite imagery, exposures can be feasibly assigned to a large study population spanning an entire province. The trade-off of using a crude metric such as NDVI is that potentially important features of vegetation, such as plant type, and access to the vegetation are not captured.

Even within NDVI, there are still many options for quantifying the exposure; there are annual vs growing season (i.e., May 1st to August 31st) averages, and a variety of distances to choose from. In this project, the measurement used to quantify greenness was the maximum of the mean growing season NDVI values that were located within a 250 m radius of the residential six-digit postal code area. The average of this NDVI value over pregnancy (based on weeks of gestation spent in each year when a pregnancy spanned two calendar years) and over childhood was assigned to each individual. This approach was preferred over a time-varying method since the recording of satellite imagery occurs only every 16 days across the growing season, thus the year-to-year calculation of NDVI may be vulnerable to getting adequate weather conditions on the sampling days. Temporal averaging of NDVI across years is a recommended method for generating more stable estimates that are robust to interannual annomalies.[115] As the spatial variation of greenspace is much greater than

the subtle changes to greenspace over time, the temporal averaging of greenspace was an acceptable step to take to ensure more accurate measurements.

Greenspace has not yet been investigated in relation to IBD. In Chapter 4, we present the results of an analysis that investigated the relationship between residential greenness during pregnancy or childhood, and the risk of developing paediatric-onset IBD.

1.6 Research objectives

There are two main research objectives for this thesis:

Objective 1: To evaluate the association between maternal and early-life exposures to ambient air pollution and the risk of paediatric inflammatory bowel disease

Objective 2: To evaluate the association between maternal and early-life exposures to residential greenspace and the risk of paediatric inflammatory bowel disease

To answer these research objectives, a retrospective cohort was formed using administrative health data. In separate analyses, the independent effects of air pollutants and greenspace were considered, while also controlling for other important environmental factors known to be associated with IBD.

1.7 Organization of the thesis

This thesis is written in the “thesis by manuscript” style. Immediately following this introduction chapter, Chapter 2 will describe the study cohort used to address the

research objectives. Chapters 3 and 4 are based on manuscripts formatted for publication. Chapter 3 investigates ambient air pollution as a potential risk factor for paediatric IBD. Chapter 4 looks at residential greenspace as a potential protective factor for paediatric IBD. Finally, Chapter 5 will discuss the thesis as a whole, and place into context the findings reported herein.

- End of Chapter 1 -

Chapter 2: Description of cohort

2.1 Data sources and linkage

Health administrative data were used to create the cohorts involved in answering both study objectives. This involved the linkage of several datasets housed and maintained at ICES uOttawa, a satellite site of ICES, which is a provincial not-for-profit research institute formerly known as the Institute for Clinical Evaluative Sciences (https://www.ices.on.ca/).

ICES assigns a unique encoded identifier known as an ICES Key Number (IKN), based on encrypted health card number, to every person eligible for provincial health insurance. The IKNs are then used to deterministically link individuals across databases, giving a broader picture of the health service usage and demographics of the population.

An overview of each of the data sources, as well as information on how they were linked is presented in this section.

2.1.1 MOMBABY dataset

The MOMBABY dataset is an ICES-derived dataset that is based on information available in the Canadian Institute of Health Information Discharge Abstract Database (CIHI-DAD), which contains administrative, clinical and demographic information on all hospital discharges, including obstetrical delivery admissions. It contains linked information on all delivering mothers and their newborns for hospital (not home) births.

Each record in MOMBABY refers to a delivering mother and her infant and contains IKNs for both the mother and child. We used variables available from MOMBABY including gestational age (available post 2002), infant sex, maternal age, and birthweight.

2.1.2 Registered Persons Database (RPDB)

The RPDB contains demographic information on anyone that has been assigned an Ontario health card number. These data are provided by the Ontario Ministry of Health and Long-Term Care, with some supplemental information added from in-house datasets at ICES. The RPDB contains important information such as age, sex, address, and IKNs for each individual. Of particular importance is the annual 6-digit postal code that is available, based on the estimated residence as of July 1st of each year. The information in the RPDB gets updated under several scenarios: (1) when the health card is renewed, typically occurring every 5 years, (2) voluntarily reported when a change of residential address occurs, and (3) during an interaction with the health care system. This database helps not only assign exposures based on 6-digit postal codes but also allowed us to capture residential mobility to some extent. As the study population changes addresses throughout early life, the environmental exposures can be obtained from their new reported postal code.

2.1.3 Ontario Crohn's and Colitis Cohort (OCCC)

The OCCC is another ICES-derived cohort comprising all incident and prevalent cases of IBD, based on several age-specific algorithms that have been validated within the province.[23,116] The algorithms used to identify IBD were derived from physician billing

data (from the Ontario Health Insurance Plan (OHIP) database), hospitalization data (from CIHI-DAD), and procedural data (from the OHIP database). Two algorithms were validated, depending on whether there was a recorded colonoscopy or sigmoidoscopy in their health records.[23] For children who were scoped, four physician contacts or two hospitalizations with an IBD diagnostic code (International Classification of Diseases [ICD]-9: 555, 556 or ICD-10: K50, K51) within a three-year period was indicative of an IBD diagnosis. For those who were not scoped, seven physician contacts or three hospitalizations with IBD codes within a three-year period were required to be classified as an incident case of IBD. The accuracy of these algorithms, when combined, had a sensitivity of 90.5%, specificity of >99.9%, positive predictive value (PPV) of 59.2% and a negative predictive value (NPV) of >99.9% among those aged <15 years old, while PPV increased to 76.0% for those aged <12 years at diagnosis in the algorithm derivation cohort. In a validation cohort of children <18 years at diagnosis, the sensitivity was 91.1%, specificity 99.5%, positive likelihood ratio 188.3 and negative likelihood ratio 0.090. Incident cases were available from April 1st, 1994 – March 31st, 2017, allowing for a 3-year lookback period to distinguish incident from prevalent cases with >90% accuracy. The recorded date of diagnosis for all children corresponds to the first date of contact with an IBD diagnostic code.

In addition to information on whether or not each child developed IBD, there was information on the disease subtype at time of diagnosis and the disease subtype based on the most recently available health records. A validated algorithm was used to determine this subtype classification, using diagnostic codes on the last five of seven

outpatient visits to distinguish CD from UC. Discrimination between the two disease types was fairly accurate, with a sensitivity of 95.1%, specificity of 86.0%, PPV of 92.0% and NPV of 91.2%. In the small proportion of cases (usually less than 10%) that the disease subtype cannot be determined a subtype of IBD-U was given.[23] For each case of incident IBD identified by the algorithm, the subtype determined at the initial diagnosis as well as the subtype based on the latest available health data are both available. The most recent subtype was used in the analyses presented in Chapters 3 and 4.

2.1.4 CENSUS database

The CENSUS database stored at ICES includes information from the previous Canadian censuses of the population (provided by Statistics Canada). For this study, information solely from the 2006 census was used. A single year was chosen to ensure consistency in the geographic variables, which often change from one census to the next. The year 2006 was selected as it is the census year that is closest to the midpoint of the study period. Geographic variables such as postal codes, census division, census tract, and census subdivision, as well as area-level variables such as median dissemination area household income and rural/urban residential status were linked to the main study population via 6-digit postal code at birth.

2.1.5 Air pollution datasets

Datasets containing the modelled surfaces for NO_2, $PM_{2.5}$ and O_3 were obtained from Health Canada. A spatiotemporal interpolation was then conducted with information from NAPS monitors in order to get weekly exposure estimates of each pollutant as

previously described.[117] For each air pollutant, the weekly mean concentration as measured at a NAPS monitor was divided by the previously derived long-term surface estimate at that monitor location to calculate a scaling factor. This was done for all available monitors. Using all of these scaling factors, a scaling surface was then created for each week by inverse distance weighting to all postal codes within 25 km of a NAPS monitor (where two NAPS monitors were in range, the closest one was taken). Weekly concentration estimates at six-digit postal codes were then generated by multiplying back these weekly scaling surfaces with the long-term estimates.

Datasets for each of the 3 air pollutants were imported into the ICES environment as project-specific data and were linked to the cohort via six-digit postal codes and dates. After linkage to the main cohort, weekly O_x estimates were generated by applying Equation 1 to the weekly estimates of NO_2 and O_3 where both were available, otherwise the estimate was coded as missing.

2.1.6 Greenness dataset

Annual datasets of Landsat NDVI values with varying methodology and buffer sizes were provided by the Canadian Urban Environmental Health Research Consortium (CANUE). The maximum of the mean growing season NDVI values within a 250m radius of each residential postal code was extracted for each year. A single composite dataset with all annual estimates was created and then imported into ICES as project-specific data. Variables include postal code and annual NDVI estimates from 1990-2015. Linkage to the main study population was achieved through six-digit postal code and dates of follow-up.

2.1.7 Dataset linkage

Linkage of all datasets was conducted at ICES uOttawa, through either the use of IKN or 6-digit postal codes. The MOMBABY dataset contained all eligible mother-infant pairs, and acted as the spine of the final dataset, with additional datasets being linked to add important information. Figure 3 illustrates how the databases were linked in order to assemble the study population.

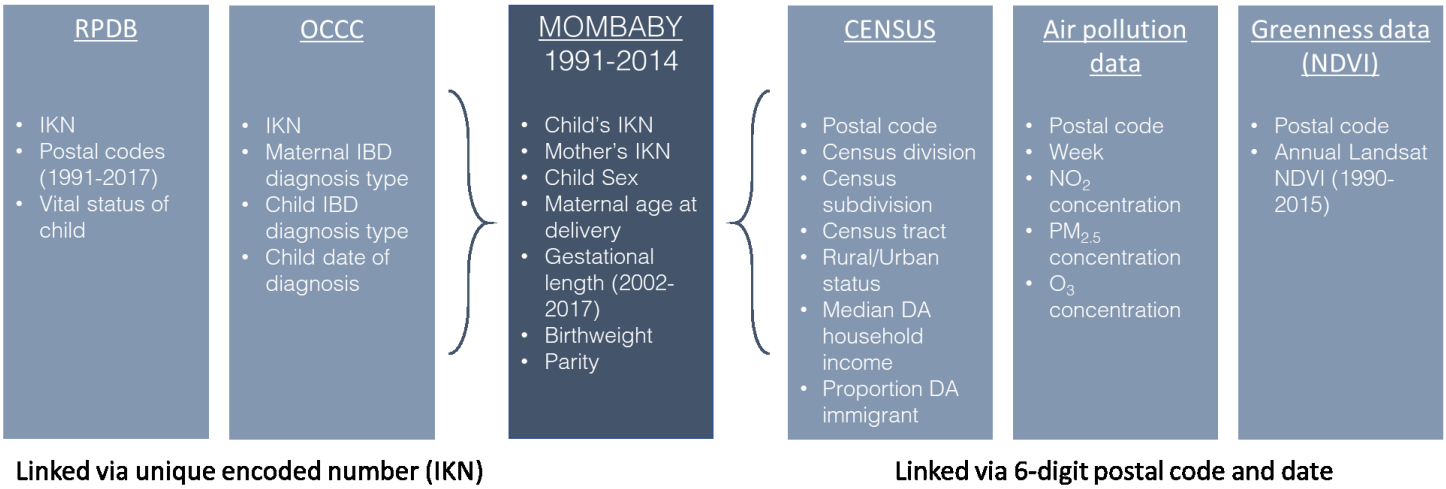


Figure 3. Database linkage diagram.

2.2 Cohort characteristics

As the same source population was used to address both study objectives, it will be described in some detail in this section. The source population comprised all registered births in Ontario between April 1st, 1991 and March 31st, 2014. These dates were chosen as they were the maximum range of birth dates that have information on outcomes from the OCCC. The latest available date in the OCCC at the time was until March 31st, 2017, so births were included up until three years prior to this date to ensure that there was a minimum of three years of health record information to allow all

at-risk children to qualify for the algorithm to define IBD. After obtaining information on all live births in this time frame, children were excluded for the following reasons:

1. Invalid mother or child identification number
2. Part of a multiple birth
3. Invalid information on biological sex
4. Ineligibility for provincial health insurance at birth
5. Missing information on covariates / exposures

The population flow-diagram (Figure 4) illustrates the exclusion steps that were applied to obtain the final study populations, with the numbers referring to the exclusion steps above.

(1) 103 675 (3.46%) of mother-infant pairs were excluded for not having a valid IKN. Without a valid IKN, information on outcomes, important covariates and exposures can not be assigned.

(2) 87 321 (3.02% of previously included cohort) children were removed for being part of a multiple birth (twins or higher-order multiple). This was done as children that are part of a multiple birth may have drastically different perinatal characteristics (birth weight, gestational age, etc.) and differences in prenatal development that makes it difficult to compare against singleton births.

(3) 86 (<0.01%) of births were excluded for having an invalid recorded biological sex, as this was to be a predictor in the models and is indicative of incomplete health records.

(4) 72 140 (2.57%) of the remaining population were excluded for being ineligible for provincial insurance at birth. This typically occurs through having a residential postal code outside of the province. If a child is ineligible for health insurance at birth, the outcome would never be able to be captured as the OCCC relies on information from provincial health records to determine outcome status. This dataset of 2,731,403 children will be referred to as the "source cohort" as this cohort was used in both analyses within this thesis.

(5) 13 987 (0.51%) individuals were missing information on rural/urban residence at birth, and/or neighbourhood income, and were excluded from both studies. An additional 498 627 (18.26%) of individuals had no information on any of the air pollutants during either pregnancy or childhood (or both). This was likely due to living in a remote postal code, or errors in NAPS monitors, and they were removed for Study population 1 which is used in Chapter 3 for the air pollution analyses as they were not able to be included in any of the regression models. 2 098 (0.08%) of individuals from the source cohort had no greenspace exposure estimate in pregnancy and childhood, and so were removed to form "Study Population 2", which was used in Chapter 4 for the greenspace exposures.

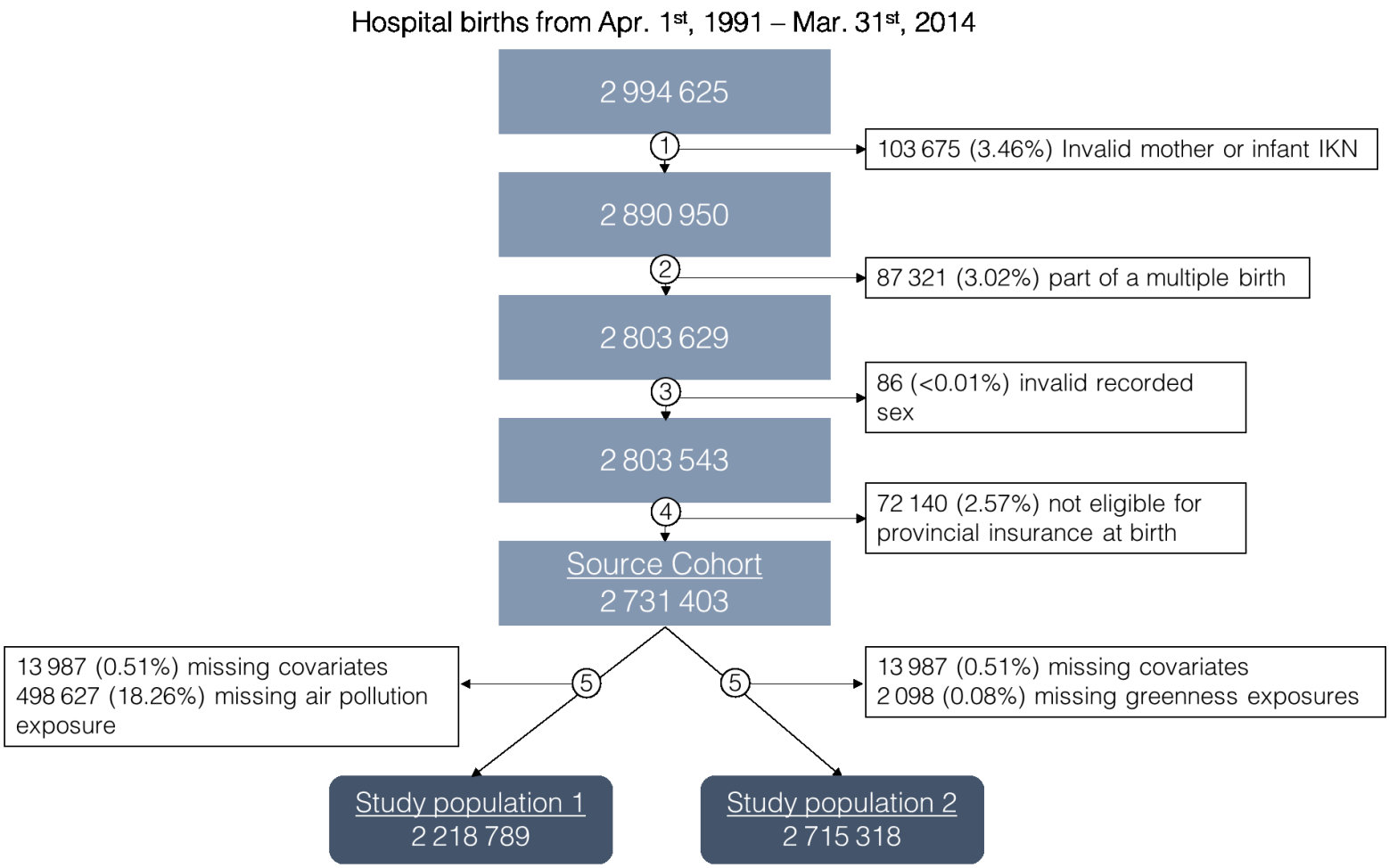


Figure 4. Flow diagram of study populations

The baseline characteristics of the cohort before exclusion for missing covariate and exposure values are shown in Table 1. To see the characteristics of the final study population after exclusions, refer to Chapter 3 and Chapter 4 respectively.

Table 1. Demographic characteristics of the source cohort at birth, stratified by IBD status

Characteristic	IBD cases (n=3 474)	Non-IBD cases (n=2 727 929)
Sex		
Female	1 478 (43%)	1 329 274 (49%)
Male	1 996 (57%)	1 398 655 (51%)
Median neighbourhood income quintile		
1st quintile (lowest)	568 (16.4%)	543 383 (19.9%)
2nd quintile	706 (20.3%)	543 127 (19.9%)
3rd quintile	678 (19.5%)	543 213 (19.9%)
4th quintile	694 (20.0%)	543 111 (19.9%)
5th quintile (highest)	808 (23.3%)	543 054 (19.9%)
Missing	20 (0.6%)	12 041 (0.4%)
Mother with IBD		
Yes	101 (3%)	11 988 (0.5%)
No	3 373 (97%)	2 715 941 (99.5%)
Residence at birth		
Urban	3 171 (91.3%)	2 418 918 (88.7%)
Rural	278 (8.0%)	295 424 (10.8%)
Missing	25 (0.7%)	13 587 (0.5%)
Maternal age (years)	30.0 (4.9)	29.5 (5.8)
Birth weight (grams)	3445 (567)	3411 (555)
Gestational age (weeks)	38.8 (1.5)	38.9 (1.6)

Results reported as counts (percentages) for categorical variables or means (standard deviations) for continuous variables. Percentages may not sum up to 100 because of rounding.

As noted in Table 1, there were a total of 3 474 cases of IBD that were captured in the source cohort. The distribution of disease subtypes at initial diagnosis and based on

the latest available health records for the source cohort are shown in Table 2. The majority of subtype classifications did not change from the initial diagnosis (92% of initial CD, and 85% of initial UC diagnoses did not change, while 50% of initial cases where IBD type was not classifiable by the algorithms were reclassified after considering further health records). For the analyses presented in this thesis, the latest available subtype classification was used. For subtype distributions of the final study populations, refer to Chapter 3 and Chapter 4.

Table 2. Changes in distribution of disease subtypes in the source cohort

	Latest CD	Latest UC	Latest IBD-U	Total
Initial CD	1 748	110	50	1 908
Initial UC	110	1 040	78	1 228
Initial IBD-U	66	104	168	338
Total	1 924	1 254	296	**3 474**

2.3 Handling of missing data

As is often the case when conducting large population-based studies using administrative data, there was a proportion of the study population that were missing information on exposures, covariates, and outcomes. In total, there were 512,614 (18.8% of the source cohort) individuals who were excluded for missing air pollution

exposures or for having missing covariates in Chapter 3, and 13,987 subjects (0.51% of the source cohort) that were excluded from Chapter 4 for having missing NDVI estimates or information on covariates.

2.3.1 Missing outcomes

Since IBD status was ascertained by review of health records, in order to have missing information on the outcomes, people would have to have an invalid identification number, or be ineligible for provincial insurance. Both of these groups were excluded in previous steps. Since there was no one explicitly missing information on their outcome, no action was needed for this.

2.3.2 Missing exposures

For air pollution exposures, the main source of missing data was the spatiotemporal interpolation that was used to take long-term surfaces of air pollutants and convert to weekly concentration estimates. Two main sources of missing exposure data are from having a remote residential six-digit postal code (more than 25km away from the closest NAPS monitor), or due to missing information at the NAPS monitors themselves. When missingness was due to an error in nearest NAPS monitor, information from the second nearest NAPS monitors was used to estimate weekly concentration where there was more than one located within 25 km.

Individuals that did not have a single valid weekly estimate to any of NO_2, $PM_{2.5}$ or O_3 during the pregnancy period or annual estimate during the childhood period were excluded from the study population. For instances where a week of exposure during

the pregnancy period was missing, a forward extrapolation technique was used where the previous week's estimate was substituted in the current week. If the previous week was also missing, a back-extrapolation technique was used to substitute the following week's exposure. For the cumulative time-varying air pollution exposures during childhood, when a child had a year of follow-up with no valid exposure estimate, their average exposure from birth up to that point was imputed. If a child was missing annual exposure estimates starting at birth, but then had valid estimates later on, the preceding missing years were left as missing, but the first year with a valid exposure estimate was considered as the average exposure up to that point, and used in the running average calculation thereafter in subsequent years. An indicator variable was created to capture all cases where an extrapolation of an exposure estimate had occured, and a sensitivity analysis excluding them was conducted. An example illustrating the air pollution exposure imputation is shown in Figure 5.

Example exposure data

ID	Follow-up	Year 1	Year 2	Year 3	Year 4	Year 5	IBD
1	5	.	.	20	24	.	1
2	5	12	20	19	.	.	0

Scenario 1: No exposure imputation

ID	time start	time stop	Cumulative NO_2	event	Used in analysis
1	0	1	.	0	N
1	1	2	.	0	N
1	2	3	.	0	N
1	3	4	.	0	N
1	4	5	.	1	N
2	0	1	12	0	Y
2	1	2	16	0	Y
2	2	3	17	0	Y
2	3	4	.	0	N
2	4	5	.	0	N

Scenario 2: Exposure imputation

ID	time start	time stop	Cumulative NO_2	event	Used in analysis
1	0	1	.	0	N
1	1	2	.	0	N
1	2	3	20	0	Y
1	3	4	21	0	Y
1	4	5	21	1	Y
2	0	1	12	0	Y
2	1	2	16	0	Y
2	2	3	17	0	Y
2	3	4	17	0	Y
2	4	5	17	0	Y

Figure 5. Example exposure estimation. Fictional exposures were created for ID1 and ID2 who each had 5 years of follow-up. ID_1 had two exposure estimates at years 3 and 4 while ID_2 had exposure estimates at years 1-3. In scenario 1 (without exposure imputation) no years of follow-up from ID_1 are used, and the last two years from ID_2 are not used. In scenario 2 (with exposure imputation) the last 3 years of follow-up from ID1 are able to be used and all years from ID_2 can be used. The extrapolated exposure estimates are shown in red.

Greenness estimates were less frequently missing due to the exposure estimates being based on long-term temporal averages of annual estimates during either pregnancy or childhood that made it less prone to missingness. If a year was missing, it was excluded in the long-term average calculation. Those without an NDVI estimate during pregnancy or childhood (and thus not included in any of the regression models) were excluded from the study population described in Chapter 4.

2.3.3 Missing covariates

There were also cases where some covariates used in the included studies were missing. There were 86 (<0.01%) children with invalid recorded biological sex who were excluded from the source cohort (Figure 4). In a similar way to how missing outcomes were not present in the study population, there were no instances of missing maternal IBD status, as those mothers with invalid identification numbers were excluded. There were 31 557 (1.2% of source cohort) individuals that were missing neighbourhood income data at the dissemination area-level. For these cases, the average median household income at the census subdivision was imputed if available. This left a total of 12 061 (0.4% of source cohort) individuals with missing income information. The individuals that had median income imputed were flagged so they could be excluded in a sensitivity analysis. There were also 13 612 (0.5% of source cohort) individuals that were missing information on rurality of residence at birth. For models where income or rurality were predictors, a listwise deletion methodology was applied so that any observation that had these covariates missing were excluded from those analyses.

2.3.4 Missing gestational age

The gestational age (GA) variable was available on both the child's and mother's health records for births that occurred after 2002 in the MOMBABY dataset, with a generally high agreement between the two measures, and were within 1 week of each other 98.7% of the time. This variable was based on a combination of clinical estimate based on ultrasound imaging, and on the reported last menstrual period. The child's GA estimate was used to assign weeks of pregnancy in the cohort, but in the small proportion of cases (<0.1%) where the baby's gestational age was missing, but the mother's was available, the mother's health record was used.

For births that occurred prior to 2002, no gestational age variable was available. To handle this, several easily implementable algorithms were evaluated in the proportion of the cohort that did have a valid GA estimate. The algorithm that was found to be the most accurate and unbiased at predicting recorded GA was then used to estimate GA for the births that occurred before 2002 in the source cohort for this thesis. The three steps of the chosen algorithm were as follows:

1. Preterm birth related diagnostic codes on the health records of either the mother (ICD-9: 644, ICD-10: O60) or the child (ICD-9: 765, ICD-10: P07) were used to separate the population into suspected preterm and term births.

2. For the suspected preterm births, the birthweight of each child was compared to a reference birthweight table for the Canadian population.[118] The gestational

age with the closest sex-specific median birthweight from the reference table was assigned to each infant.

3. For the suspected term births, a gestational age of 39 weeks was assigned.

The accuracy of this algorithm in estimating gestational age is shown in Figure 6. The within 1-week and within 2-week accuracies of this algorithm applied in the validation cohort were 76.7% and 96.6% respectively.

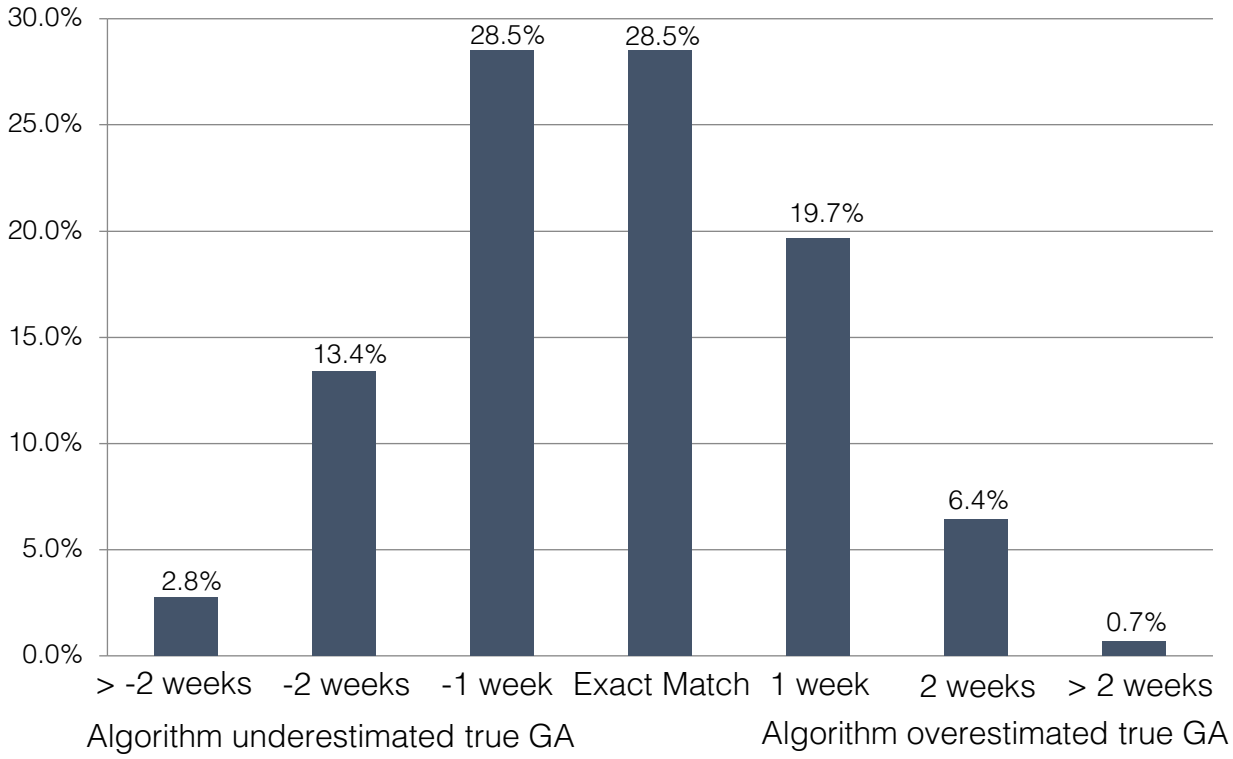


Figure 6. Accuracy of algorithm in estimating gestational age in a validation cohort.

- End of Chapter 2 -

Chapter 3: Air pollution and paediatric-onset inflammatory bowel disease

This chapter is based on the manuscript titled "*Ambient air pollution and the risk of pediatric-onset inflammatory bowel disease: a population-based cohort study*", which is formatted for submission to the journal *Environmental Health Perspectives*. It addresses Objective 1 of this thesis, which is to investigate the relationship between ambient air pollution and development of IBD in children. Ethics approval for this study was granted by the Research Ethics Boards of Health Canada and the Ottawa Health Science Network. The Reporting of studies Conducted using Observational Routinely-collected data (RECORD) guideline completed checklist for this manuscript is located in Appendix A – RECORD statement checklist for air pollution manuscript. Note that since this is an American-based journal, Americanized spelling of words such as "pediatric" and "neighborhood" are used exclusively in this chapter.

Authors: Michael Elten[1,2], Eric I. Benchimol, MD PhD[1,3,4,5,6], Deshayne B. Fell, PhD[1,3,4], M. Ellen Kuenzig, PhD[1,3,4,5], Glenys Smith, MSc.[4], Hong Chen, PhD[4,7,8,9], Gilaad G. Kaplan, MD MPH[10], Éric Lavigne, PhD[1,2,4]

1. School of Epidemiology and Public Health, University of Ottawa, Ottawa Canada
2. Air Health Sciences Division, Health Canada, Ottawa Canada
3. Children's Hospital of Eastern Ontario (CHEO) Research Institute, Ottawa Canada
4. ICES uOttawa, Ottawa, Canada

5. CHEO Inflammatory Bowel Disease Centre, Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Eastern Ontario, Ottawa, Canada

6. Department of Pediatrics, University of Ottawa, Ottawa Canada

7. Public Health Ontario, Toronto Canada

8. Dalla Lana School of Public Health, University of Toronto, Toronto Canada

9. Population Health Studies Division, Health Canada, Ottawa, Canada

10. Departments of Medicine and Community Health Sciences, University of Calgary, Calgary Canada

Author contributions: ME, EIB, and ÉL contributed to study design. ME and GS prepared the data. ME conducted all analysis. All authors helped interpret the results, and provided critical feedback and revisions to the final manuscript

3.1 Acknowledgements

This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). This study also received funding from Health Canada's Clean Air Regulatory Agenda. Parts of this material are based on data and information compiled and provided by MOHLTC and the Canadian Institute for Health Information (CIHI). The analyses, conclusions, opinions and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred. NDVI metrics, indexed to DMTI Spatial Inc. postal codes, were provided by CANUE (Canadian Urban Environmental Health Research Consortium).

M. Ellen Kuenzig was supported by a Post-Doctoral Fellowship Award from the Canadian Institutes of Health Research (CIHR), Canadian Association of Gastroenterology (CAG), and Crohn's and Colitis Canada, as well as a Mitacs Elevate Post-Doctoral Fellowship. Eric Benchimol was supported by a New Investigator Award from the CIHR, CAG and Crohn's and Colitis Canada. Eric Benchimol was also supported by the Career Enhancement Program of the Canadian Child Health Clinician Scientist Program.

3.2 Abstract

Background: High-income industrialized nations have the highest rates of inflammatory bowel disease (IBD). The incidence of pediatric-onset IBD is increasing faster than IBD

diagnosed in older individuals. Previous epidemiological studies have shown that air pollution might be a risk factor for the development of IBD, but results remain mixed.

Objectives: The objective of this study was to evaluate the associations between maternal and early-life exposures to nitrogen dioxide (NO_2), fine particulate matter ($PM_{2.5}$), ozone (O_3,) and oxidant capacity (O_x) and risk of pediatric-onset IBD diagnosis.

Methods: We conducted a retrospective cohort study using linked population-based health administrative data. Singleton livebirths in Ontario, Canada between April 1st, 1991 and March 31st, 2014 were included. We investigated the association between weekly exposures during pregnancy and annual exposures from birth until the age of 18 years, and IBD diagnosed <18 years of age using Cox proportional hazards models. Hazard ratios (HR) and 95% confidence intervals (CI) are for an associated increase in the interquartile range (IQR) of each pollutant. Models were mutually adjusted for exposures in both prenatal and postnatal periods, as well as for sex, rurality of residence at birth, maternal IBD, and neighborhood income.

Results: 2 218 789 newborns were included in this analysis, of whom 2 491 developed IBD during follow-up. Increased associations with pediatric-onset IBD were noted for childhood exposure to O_x (HR 1.08, 95% CI 1.01-1.16). An increased association was found for second trimester exposure to O_x (HR 1.21, 95% CI 1.03-1.42), but not for the overall pregnancy period (HR 1.12, 95% CI 0.79-1.59). No significant association was noted for exposures to NO_2, $PM_{2.5}$, or O_3.

Discussion: Exposure to O_x during childhood was associated with IBD <18 years. This suggests that air pollution may impact the developing child physiology in such a way that leads to early onset of IBD.

3.3 Introduction

Inflammatory bowel diseases (IBD) are a group of chronic immune-mediated digestive diseases comprised mainly of two subtypes: Crohn's disease (CD) and ulcerative colitis (UC). Both are characterized by inflammation of the gastro-intestinal tract, caused by an inappropriate response of the host immune system.[1] Incidence of inflammatory bowel disease increased in high income nations during industrialization, and have since plateaued. Lower income nations are currently experiencing a similar rise in incidence.[2] IBD diagnosed during childhood and adolescence has become more frequent worldwide, with incidence rates in Canada among the highest in the world.[2,3]

Although its etiology is complex and not fully understood, both environmental and genetic factors play a role in IBD.[4] Exposure to air pollution in-utero and throughout childhood is responsible for many adverse health outcomes including other immune-mediated diseases (e.g., type 1 diabetes, asthma)[5,6] and may similarly contribute to the development of IBD.[7] In previous observational studies, nitrogen dioxide (NO_2, commonly used as a proxy marker for traffic related air pollution) was associated with CD in a subgroup of individuals diagnosed before 23 years of age[8], as was a measure of nearby traffic intensity on major roads.[9] Null and inverse associations were also

reported in both of these studies, however. Air pollution is hypothesized to contribute to the risk of IBD through disturbances to the gut microbiome and immune system development.[10]

Several knowledge gaps remain on the potential effects of air pollution among childhood-onset IBD, and on the association between prenatal exposures of air pollution and IBD incidence. To address this, we conducted a large population-based retrospective cohort study to investigate the associations between maternal and childhood exposures to NO_2, fine particulate matter with a diameter ≤2.5 μm ($PM_{2.5}$), ozone (O_3), and redox-weighted oxidant capacity (O_x, a measure that takes into account the oxidative potential of both NO_2 and O_3), and risk of developing pediatric-onset IBD.

3.4 Methods

Ethics approval for this study was granted by the Research Ethics Boards of Health Canada and the Ottawa Health Science Network.

3.4.1 Study sources

We used population-based health administrative data from the province of Ontario, Canada (population: 13.6 million in 2014),[11] comprising all legal residents with a valid health card (>99% of the population). These data are housed at ICES, an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data. We used the following datasets: the Registered Persons Database (RPDB) (contains

demographic information on those registered for health insurance), MOMBABY (an ICES-derived cohort that links hospital admission records of mothers with their newborns for all hospital births in the province using the Discharge Abstract Database from the Canadian Institute for Health Information), and the Ontario Crohn's and Colitis Cohort (a validated cohort of all people diagnosed with IBD in Ontario)[3]. Additional area-level information was obtained from the CENSUS dataset that contains information collected from previous censuses of Canada. Information on the air pollution datasets are described below. The administrative datasets were linked using unique encoded identifiers and analyzed at ICES. Full anonymized and unedited datasets were available to researchers for this study.

3.4.2 Study design and population

This study was a retrospective cohort comprised of all live births occurring in hospitals in Ontario, Canada between April 1st, 1991 and March 31st, 2014. Children were excluded from the analyses if they had an invalid identification number (or a mother without a valid identification number), were part of a multiple birth, had an invalid recorded sex, or were ineligible for provincial health insurance at birth. 6-digit postal code, and exposure. As a final step, those that had missing or invalid covariate or missing or invalid exposure estimates to all pollutants during pregnancy or childhood were excluded from the analysis.

3.4.3 Outcome ascertainment

Children diagnosed with IBD before 18 years of age were identified from the Ontario Crohn's and Colitis Cohort using an algorithm that has been previously validated in the

Ontario pediatric population.[3] Children undergoing colonoscopy who had four physician contacts or two hospitalizations with an IBD diagnostic code (International Classification of Diseases [ICD]-9: 555, 556; ICD-10: K50, K51) within a 3-year period were classified as IBD cases. Children who did not have a colonoscopy were classified as IBD cases if they had seven physician contacts or three hospitalizations for IBD within a 3-year period. The date of the first code for IBD was considered the date of diagnosis. The algorithm identified children with IBD within Ontario's health administrative databases with a sensitivity of 91.1% and a specificity of 99.5%. In addition, a validated algorithm based on the diagnoses assigned to the last five of seven outpatient visits distinguished CD from UC patients with a sensitivity of 95.1% and a specificity of 86.0%. For a small proportion of IBD diagnoses (less than 10%), the disease subtype could not be determined from the health records and a subtype of IBD-unclassifiable was given. These were included in overall IBD analyses but excluded from subtype analyses.

3.4.4 Exposure assessment

Residential exposures to nitrogen dioxide (NO_2), fine particulate matter with a diameter ≤ 2.5 μm ($PM_{2.5}$), ozone (O_3), and redox-weighted oxidant capacity (O_x) were assigned to the study population at the geographic centre of each 6-digit postal code area. This assignment was facilitated by the Postal Code Conversion File Plus which was used to convert residential postal codes into geographic coordinates.[12] The mother's residence at the time of delivery was used for determining exposure assignment during

pregnancy, while the annual address as of July 1st in the RPDB was used for childhood exposures.

NO_2 was assessed based on a national land-use regression (LUR) model, using data from the Canadian National Air Pollution Surveillance (NAPS) monitoring network, combined with information on 2005-2006 satellite-derived NO_2 estimates, road lengths within 10km, area of industrial land use within 2 km and the mean summer rainfall.[13] Kernel density functions describing the density of the roadways were then applied to capture localised variations in vehicle emissions. This final spatial model predicted 73% of the variability in fixed ground monitor NO_2 concentrations. NO_2 estimates were available from April 1, 1990 - December 25, 2016.

A spatial $PM_{2.5}$ surface was derived using satellite-based estimates that were combined with ground-level monitor information, as described by van Donkelaar et al.[14] First, aerosol optical depth estimates were produced at a 1 kilometre (km) resolution across North America using an optimal estimation algorithm. A chemical transport model and geographically weighted regression techniques were then used to improve the accuracy of the exposure estimates, resulting in an overall agreement R^2 of 0.82. $PM_{2.5}$ estimates were available from Dec. 28, 1997 – Dec. 28, 2014.

O_3 was assessed using 21 km grid values, based on a surface that represents an average of daily 8 hour maximum concentrations in the warm seasons (May 1st to October 31st) using an optimal interpolation technique described previously.[15] O_3 estimates were available from April 1, 1990 - December 25, 2016.

As the modelled exposures are provided on an annual basis, spatiotemporal interpolation was conducted using information from monitors from the NAPS network to obtain estimates on a weekly scale. Briefly, a scaling factor was calculated for each NAPS ground monitor location by taking the weekly mean concentration measured by that monitor and dividing it by the previously described long-term surface estimates at the monitor location. A scaling surface was then created through a spatial interpolation technique that used inverse distance weighting to these assigned scaling factors to all postal codes within 25 km of a NAPS monitor. Finally, these scaling surfaces were multiplied back to the long-term estimates to obtain estimated weekly concentrations of each pollutant at the 6-digit postal code level.[16]

Oxidant capacity combines measures of NO_2 and O_3. These pollutants are often highly correlated and their combined exposure can cause greater oxidative stress than their individual contributions.[17,18] Using weekly NO_2 and O_3 surfaces derived previously, redox-weighted oxidant capacity was calculated as the weighted average according to the following formula: $O_x = [(1.07 \times NO_2) + (2.075 \times O_3)]/3.145)$.

Weekly exposures during pregnancy were assigned using the recorded gestational age available in the MOMBABY dataset. When a week of pregnancy was missing an exposure estimate, the previous week's estimate was extrapolated forwards. If the previous week was also missing, the following week's exposure estimate was extrapolated backwards. For exposures after birth, a time-varying exposure estimate for each pollutant was derived by taking a running average of the weekly exposures from birth until the end of each year of life. For instance, the estimate in the third year of

life is the average exposure from birth until the end of year three. When a year of exposure was missing, the average up to the previous year was extrapolated forwards. If a child's first years of exposure were missing, the first non-missing annual exposure estimate was considered to be their average exposure up to that year, but the previous missing years were left as missing. This was done to increase the amount of available observations to be modelled in the analyses as otherwise an individual missing even one week of exposure estimates during pregnancy would be excluded entirely from this study. We conducted a sensitivity analysis excluding those with extrapolated exposure estimates to see if this noticeably affected the observed associations.

3.4.5 Missing gestational age variable

For births occurring prior to 2002, no recorded gestational age (GA) is available, so we validated and used an algorithm to estimate GA in this subpopulation, as follows. If a diagnostic code for preterm birth was not present on the mother's (ICD-9: 765; ICD-10: P07) or child's (ICD-9: 644; ICD-10: O60) health record, a gestational week of 39 weeks was assumed. In mother-infant pairs that had at least one preterm birth code, the gestational age with the closest sex-specific median birthweight from a birthweight reference table derived from a Canadian population of singleton live births was selected as the estimated gestational age.[19] In the validation cohort of births born from fiscal years 2002 – 2015, this method was able to estimate gestational age with relative accuracy, and had a within 1-week accuracy of 77% and a within 2-week accuracy of 97%, with minimal bias.

3.4.6 Statistical analyses

Cox proportional hazards models were used to evaluate the associations between the continuous measure of each of the four air pollutants and incidence of childhood IBD. Separate models were fitted to obtain associations for disease subtypes. We used years since birth, rounded to one decimal as the underlying timescale. Children were followed until IBD diagnosis, their 18th birthday, death, ineligibility for provincial health insurance, or until the final available outcome ascertainment date of March 31st 2017. Maximum follow-up time was 3 years for the youngest children in the cohort (born in 2014) and 18 years for the oldest children (born in 1991). Results are expressed as the hazard ratio (HR) and 95% confidence interval (CI) corresponding to an increase across the interquartile range (IQR) of the pollutant of interest.

Both pregnancy and childhood exposure periods were included in the same mutually adjusted model in order to characterize which period was implicated for any observed association. Estimates were then further adjusted for potential confounding factors chosen *a priori* which have been associated with air pollution or pediatric IBD in previous research.[4] Potential confounders available in our study included sex, maternal IBD (identified from the Ontario Crohn's and Colitis Cohort using validated algorithms),[3,20] rural/urban area of residence at birth (using Statistics Canada definition of rurality in the 2006 Census),[21] and median neighborhood household income (also based on the 2006 Census, and calculated at the dissemination area level, which are small geographic units of 400-700 persons), which is a validated proxy for individual-level income, and was divided into quintiles.[22] For the 1.2% of the cohort that were

missing neighborhood income, the average median household income at the census subdivision level roughly equivalent to municipalities) were used if available, leaving only 0.4% missing.

An extension of the distributed lag non-linear model (DLNM) was used to simultaneously investigate each of the first 40 weeks of exposure during pregnancy (when pregnancies spanned less than 40 weeks, any weeks after their birth were given exposure estimates of zero).[23] Used previously in air pollution studies, this method allows for identification of critical windows of exposure for complex exposure-response relationships. To select the appropriate model, different lag structures (natural and B splines) and number of knots (2-5 knots) were used to define the crossbasis of pregnancy exposures. The crossbasis that minimized the Akaike Information Criterion (AIC) was selected as the final model. Estimates of association were obtained by calculating the cumulative hazard over weeks 0-12 (trimester 1), 13-26 (trimester 2), and 27-40 (trimester 3), or over weeks 0-40 (pregnancy average) from the DLNM models.

We investigated for potential effect modification by stratifying the main analyses by maternal IBD status (yes or no) and rural/urban residence at birth. We also quantitatively assessed for effect measure modification by adding a multiplicative interaction term between the main exposure and effect modifier along with first-order terms to the fully adjusted model. Effect measure modification was considered present if the Wald chi-square P value was <0.05. In pre-planned sensitivity analyses, we restricted to children diagnosed with IBD $\leq$10 years of age, as it has been

hypothesized that early-life exposures are the most important as it is during this time period that the gut microbiome is being established.[24]

Statistical analyses were conducted using R version 3.0.1,[25] using the *survival* (version 2.42-3) and *dlnm* (version 2.1.3) packages.[26,27]

3.5 Results

We identified 2 994 625 hospital mother-infant pairs occurring between April 1st, 1991 – March 31st, 2014. After initial exclusions, there were 2 731 403 children remaining before an additional 512 614 were removed for missing covariate or exposure information, leaving 2 218 789 eligible for inclusion in the study population (see Supplementary Figure 1 for study population flowchart). Of these, 2 491 children developed IBD during follow-up, including 1 375 cases of CD, 899 cases of UC, and 217 IBD cases that were unclassifiable using the administrative data. Characteristics of the final study population are shown in Table 3, while characteristics of those excluded for missing covariate or exposure information are shown in Supplementary Table 1. Children were followed-up for a total of 25 641 625.3 years, giving an average of 11.6 years per person. The average follow-up time for those born at least 18 years before the end of study date of April 1st 2017 (and thus theoretically able to attain full follow-up) was 16.9 years. Mean (standard deviation) exposures during pregnancy were 12.73 (7.43) ppb for NO_2, 7.96 (2.66) ug/m3 for $PM_{2.5}$, 24.14 (4.58) ppb for O_3, and 20.41 (3.10) ppb for O_x. Descriptive statistics for exposures to pollutants are described in Supplementary Table 2.

Figure 7 shows the estimated hazard for each pollutant as a function of gestational week after adjustment for childhood exposures, sex, rural/urban residence at birth, maternal IBD status, and neighborhood income. The crossbasis that minimized the AIC was a B-spline with two degrees of freedom and we used this for all four models. The graphs for O_3 and O_x suggest that there is a small association between exposures during the mid-pregnancy period and development of IBD. The graph for $PM_{2.5}$ exposure does not indicate any significant association, while NO_2, exposures in early pregnancy seemed to be associated with increased risk.

Table 4 shows the estimated hazard ratios for each pollutant and exposure period. Statistically significant positive associations were detected for second trimester exposure to O_x (HR 1.21, 95% CI 1.03–1.42) and childhood exposure to O_x (HR 1.08, 95% CI 1.01-1.16). First trimester exposure to NO_2 (HR 1.03, 95% CI 1.00-1.07) and second trimester exposure to O_3 (HR 1.08, 95% CI 1.00-1.16) were suggestive of an increased association but had confidence intervals that included the null. No other significant associations were detected for in-utero or childhood exposures to any of the pollutants.

Analyses stratified by disease subtype are presented in Table 5. Associations were not uniformly stronger for one disease subtype, although subtle differences in the effect estimates were present. Second trimester exposures to NO_2 were the only to differ in statistical significance (HR_{CD} 0.94, 95% CI 0.91-0.98; HR_{UC} 1.01, 95% CI 0.96-1.06), though the absolute difference in the effect estimates was small. Second trimester exposure to O_x (HR_{CD} 1.18, 95% CI 0.95-1.47; HR_{UC} 1.26, 95% CI 0.97-1.64) and

childhood exposure to O_x (HR_{CD} 1.05, 95% CI 0.96-1.16; HR_{UC} 1.10, 95% CI 0.98-1.24) were both slightly more strongly associated with UC.

Effect modification by rural/urban residence at birth was explored in Supplementary Table 3. First trimester exposure to NO_2 was associated with an increased risk of IBD among those living in rural areas at birth but not among those living in urban areas at birth (HR_{Rural} 1.42, 95% CI 1.10-1.84; HR_{Urban} 1.03, 95% CI 0.99-1.07; P=0.01). Second trimester exposures to O_x also had a stronger association with IBD among those born to rural residences (HR_{Rural} 3.08, 95% CI 1.35-7.04; HR_{Urban} 1.17, 95% CI 0.99-1.38; P=0.03). This pattern was also similar for second trimester exposures to O_3 (P=0.05).

Results of effect measure modification analyses by maternal IBD are presented in Supplementary Table 4. The most notable instance of effect modification was for $PM_{2.5}$ exposure during the childhood period (P=0.06), with a significant positive association being observed among children born to mothers with IBD (HR 2.03, 95% CI 1.17-3.52), but no association in those born to mothers without IBD (HR 0.95, 95% CI 0.83-1.09). Exposures to O_3 during overall pregnancy may also have been modified by maternal IBD status (P=0.10), with a negative association detected among those born to mothers with IBD (HR 0.25, 95% CI 0.06-0.98) and no association among those born to mothers without IBD (HR 0.93, 95% CI 0.75-1.14).

Sensitivity analyses restricted to early-onset IBD (diagnosed ≤10 years of age) are shown in Supplementary Figure 2 and Supplementary Table 5. All curves and associations were qualitatively similar to the main analysis. The sensitivity analysis that

looked at how extrapolation of exposures may have affected the results (Supplementary Table 6) was also very similar to the main analyses, with slightly increased effect sizes for the previously highlighted associations of NO_2 exposure during the first trimester (HR 1.06, 95% CI 1.01-1.11), O_3 and O_x exposure in the second trimester (HR 1.13, 95% CI 1.00-1.29 and HR 1.41, 95% CI 1.00-1.98 respectively), and O_x exposure during childhood (HR 1.14, 95% CI 0.94-1.37).

3.6 Discussion

In this large retrospective cohort, we observed an association between childhood exposure to oxidant capacity and increased risk of pediatric-onset IBD, but did not find an association for any of the other three pollutants investigated. We also detected an association between second trimester exposure to O_x with IBD, but when the effects of O_x across the whole pregnancy period was considered, the increased association was not conserved. These observed associations were slightly stronger for UC and persisted in a sensitivity analysis that looked at very early onset IBD cases, and in a sensitivity analysis that excluded extrapolated exposure estimates.

Childhood exposures to O_x may lead to increased risk of IBD through direct effects on epithelial tissue, dysregulation of the immune system, or through changes in the gut microbiota.[28] The potential effect of mid-pregnancy exposures to pollution are unlikely to cause direct changes to the fetal microbiome, however, as in-utero microbial colonization is extremely limited.[24] Instead, this oxidation and its resulting chronic inflammation may disrupt important processes in fetal development occurring during

mid-pregnancy, such as development of gut-associated lymphoid tissue that begins as early as 8-15 weeks of gestation.[29]

Our study did not confirm the results of a previous study which demonstrated an association between postnatal NO_2 exposure during and Crohn's disease in people under 23 years of age.[8] This could be due to differences in the way that NO_2 exposure was modelled, specifically, we modelled NO_2 as a continuous variable (instead of categorical) and using time-varying exposure estimates (instead of single exposure estimates), or due to other important differences such as the comparison of a pediatric vs adult population. This null association for NO_2 that we found was consistent with a different study that also reported null associations for NO_2.[9]

$PM_{2.5}$ was not found to be associated in either of the two previously mentioned studies on air pollution and IBD,[8,9] and was also generally not an important predictor of IBD in the current study. However, childhood exposures to $PM_{2.5}$ were associated with IBD among those born to mothers with IBD. This finding is surprising, but not unprecedented; smoking (which coincidentally is also a source of $PM_{2.5}$)[30] was found to be associated with CD among those with a family history of the disease, but not among those with no family history of the disease in a previous study.[31] This result suggests that genetic predisposition may be an important consideration in understanding the risk posed by air pollution, and other environmental exposures.

This study had several strengths worth noting. This study population is the largest retrospective birth cohort ever assembled to examine the association between

environmental risk factors and IBD, and our findings are generalizable to other similar populations that have comparable exposures to air pollution. The large sample size allowed for sufficient statistical power to detect potential associations that might be relatively small in magnitude, and to investigate smaller but relevant subgroups. We were also able to capture residential mobility during the childhood period to some extent. Using sophisticated statistical modelling, we were able to assess exposures on a weekly basis within pregnancy, a unique strategy in the IBD literature that can also help to identify critical windows of vulnerability and isolate developmental pathways that link exposure to disease.

There are also some limitations to consider when interpreting the results of this study. The health administrative data were not collected for the purpose of research, so there may be errors or changes in how the data were recorded over time. However, as air pollution levels in Ontario in general have decreased over the study period and IBD diagnoses have been increasing, this may have biased the results towards the null. Exposure misclassification is possible, from residential mobility during pregnancy and from only considering exposures at the residential area. However, mothers tend to move small distances in the same exposure regions so the level of misclassification is expected to be small.[32,33] Additionally, they would likely occur non-differentially across IBD outcomes, and would have biased our presented results towards the null, underestimating the associations between air pollutant exposures and IBD. Finally, we were able to adjust for some important covariates such as sex, maternal IBD, neighborhood income, and rurality of residence, but other important potential

confounders such as individual ethnicity were not available in the health administrative data.

In this population-based retrospective cohort study, we found some evidence of an association between exposure to oxidant capacity during childhood, and increased risk of being diagnosed with IBD before 18 years of age. This study highlights the importance of considering the contribution of multiple pollutants and environmental exposures. These findings suggest that there may be important effects of air pollution exposure on gut development, the developing immune system, or the gut microbiome that can subsequently result in IBD in childhood.

3.7 References for Chapter 3:

1 Sartor RB. Mechanisms of Disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nat Clin Pract Gastroenterol Hepatol* 2006; **3**: 390–407.

2 Ng SC, Yun Shi H, Hamidi N, *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *www.thelancet.com* 2017; **390**. DOI:10.1016/S0140-6736(17)32448-0.

3 Benchimol EI, Guttmann A, Griffiths AM, *et al.* Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. *Gut* 2009; **58**: 1490–7.

4 Aujnarain A, Mack DR, Benchimol EI. The Role of the Environment in the Development of Pediatric Inflammatory Bowel Disease. *Curr Gastroenterol Rep* 2013; **15**. DOI:10.1007/s11894-013-0326-4.

5 Malmqvist E, Elding Larsson H, Jönsson I, *et al.* Maternal exposure to air pollution and type 1 diabetes - Accounting for genetic factors. *Environ Res* 2015; **140**: 268–74.

6 Lavigne É, Bélair M-A, Rodriguez Duque D, *et al.* Effect modification of perinatal exposure to air pollution and childhood asthma incidence. *Eur Respir J* 2018; : 1701884.

7 Ritz SA. Air pollution as a potential contributor to the 'epidemic' of autoimmune

disease. *Med Hypotheses* 2010; **74**: 110–7.

8 Kaplan GG, Hubbard J, Korzenik J, *et al.* The inflammatory bowel diseases and ambient air pollution: A novel association. *Am J Gastroenterol* 2010; **105**: 2412–9.

9 Opstelten JL, Beelen RMJ, Leenders M, *et al.* Exposure to Ambient Air Pollution and the Risk of Inflammatory Bowel Disease: A European Nested Case–Control Study. *Dig Dis Sci* 2016; **61**: 2963–71.

10 Kish L, Hotte N, Kaplan GG, *et al.* Environmental Particulate Matter Induces Murine Intestinal Inflammatory Responses and Alters the Gut Microbiome. *PLoS One* 2013; **8**. DOI:10.1371/journal.pone.0062220.

11 Statistics Canada. Table 17-10-0005-01 Population estimates on July 1st, by age and sex. DOI:https://doi.org/10.25318/1710000501-eng.

12 Wilkins R, Peters PA. PCCF+ Version 5K User's Guide. 2012; **82F0086-XD**: 75.

13 Hystad P, Setton E, Cervantes A, *et al.* Creating national air pollution models for population exposure assessment in Canada. *Environ Health Perspect* 2011; **119**: 1123–9.

14 van Donkelaar A, Martin R V., Spurr RJD, Burnett RT. High-Resolution Satellite-Derived PM $_{2.5}$ from Optimal Estimation and Geographically Weighted Regression over North America. *Environ Sci Technol* 2015; **49**: 10482–91.

15 Robichaud A, Ménard R. Multi-year objective analyses of warm season ground-level ozone and PM2.5 over North America using real-time observations and Canadian operational air quality models. *Atmos Chem Phys* 2014; **14**: 1769–800.

16 Bechle MJ, Millet DB, Marshall JD. National Spatiotemporal Exposure Surface for NO2: Monthly Scaling of a Satellite-Derived Land-Use Regression, 2000-2010. *Environ Sci Technol* 2015; **49**: 12297–305.

17 Williams ML, Atkinson RW, Anderson HR, Kelly FJ. Associations between daily mortality in London and combined oxidant capacity, ozone and nitrogen dioxide. *Air Qual Atmos Heal* 2014; **7**: 407–14.

18 Weichenthal S, Lavigne É, Evans G, Pollitt K, Burnett RT. Ambient $PM_{2.5}$ and risk of emergency room visits for myocardial infarction: Impact of regional $PM_{2.5}$ oxidative potential: A case-crossover study. *Environ Heal* 2016; **15**: 46.

19 Kramer MS, Platt RW, Wen SW, *et al.* A New and Improved Population-Based Canadian Reference for Birth Weight for Gestational Age. *Pediatrics* 2001; **108**: e35.

20 Benchimol EI, Guttmann A, Mack DR, *et al.* Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. *J Clin Epidemiol* 2014; **67**: 887–96.

21 du Plessis V, Beshiri R, Bollman RD, Clemenson H. Rural and Small Town

Canada Analysis Bulletin. *Rural Small T Canada Anal Bull* 2001; **3**: 1–17.

22 Glazier RH, Creatore MI, Agha MM, Steele LS. Socioeconomic Misclassification in Ontario's Health Care Registry. *Can J Public Heal* 1996; **94**.

23 Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. *Stat Med* 2014; **33**: 881–99.

24 Perez-Muñoz ME, Arrieta MC, Ramer-Tait AE, Walter J. A critical assessment of the 'sterile womb' and 'in utero colonization' hypotheses: Implications for research on the pioneer infant microbiome. *Microbiome* 2017; **5**: 1–19.

25 R Core Team. R: A language and environment for statistical computing. 2019. https://www.r-project.org/.

26 Therneau TM. A Package for Survival Analysis in S. 2015. https://cran.r-project.org/package=survival.

27 Gasparrini A. Distributed Lag Linear and Non-Linear Models in R : The Package dlnm. *J Stat Softw* 2011; **43**: 1–20.

28 Beamish LA, Osornio-Vargas AR, Wine E. Air pollution: An environmental factor contributing to intestinal disease. *J Crohn's Colitis* 2011; **5**: 279–86.

29 West LJ. Defining critical windows in the development of the human immune system. *Hum Exp Toxicol* 2002; **21**: 499–505.

30 Invernizzi G, Ruprecht A, Mazza R, *et al.* Particulate matter from tobacco versus diesel car exhaust: An educational perspective. *Tob Control* 2004; **13**: 219–21.

31 Tuvlin JA, Raza SS, Bracamonte S, *et al.* Smoking and inflammatory bowel disease: Trends in familial and sporadic cohorts. *Inflamm Bowel Dis* 2007; **13**: 573–9.

32 Chen L, Bell EM, Caton AR, Druschel CM, Lin S. Residential mobility during pregnancy and the potential for ambient air pollution exposure misclassification. *Environ Res* 2010; **110**: 162–8.

33 Fell DB, Dodds L, King WD. Residential mobility during pregnancy. *Paediatr Perinat Epidemiol* 2004; **18**: 408–14.

3.8 Tables and figures

Table 3. Demographic characteristics of study population at birth stratified by IBD status.

Characteristic	IBD cases (n=2 491)	Non-IBD cases (n=2 216 298)
Sex		
Male	1 417 (56.9%)	1 137 224 (51.3%)
Female	1 074 (43.1%)	1 079 074 (48.7%)
Median neighborhood income quintile		
1st quintile (lowest)	428 (17.2%)	445 901 (20.1%)
2nd quintile	507 (20.4%)	445 899 (20.1%)
3rd quintile	504 (20.2%)	439 199 (19.8%)
4th quintile	505 (20.3%)	445 358 (20.1%)
5th quintile (highest)	547 (22.0%)	439 941 (19.9%)
Mother with IBD		
Yes	70 (2.8%)	9 982 (0.5%)
No	2 421 (97.2%)	2 206 316 (99.5%)
Residence at birth		
Urban	2 323 (93.3%)	1 994 073 (90.0%)
Rural	168 (6.7%)	222 225 (10.0%)
Maternal age (years)	29.9 (5.0)	29.5 (5.9)
Birth weight (grams)	34 29 (577)	3 406 (554)
Gestational age (weeks)	38.8 (1.6)	38.9 (1.6)

Results reported as counts (percentages) for categorical variables or means (standard deviations) for continuous variables. Percentages may not round to 100 due to rounding.

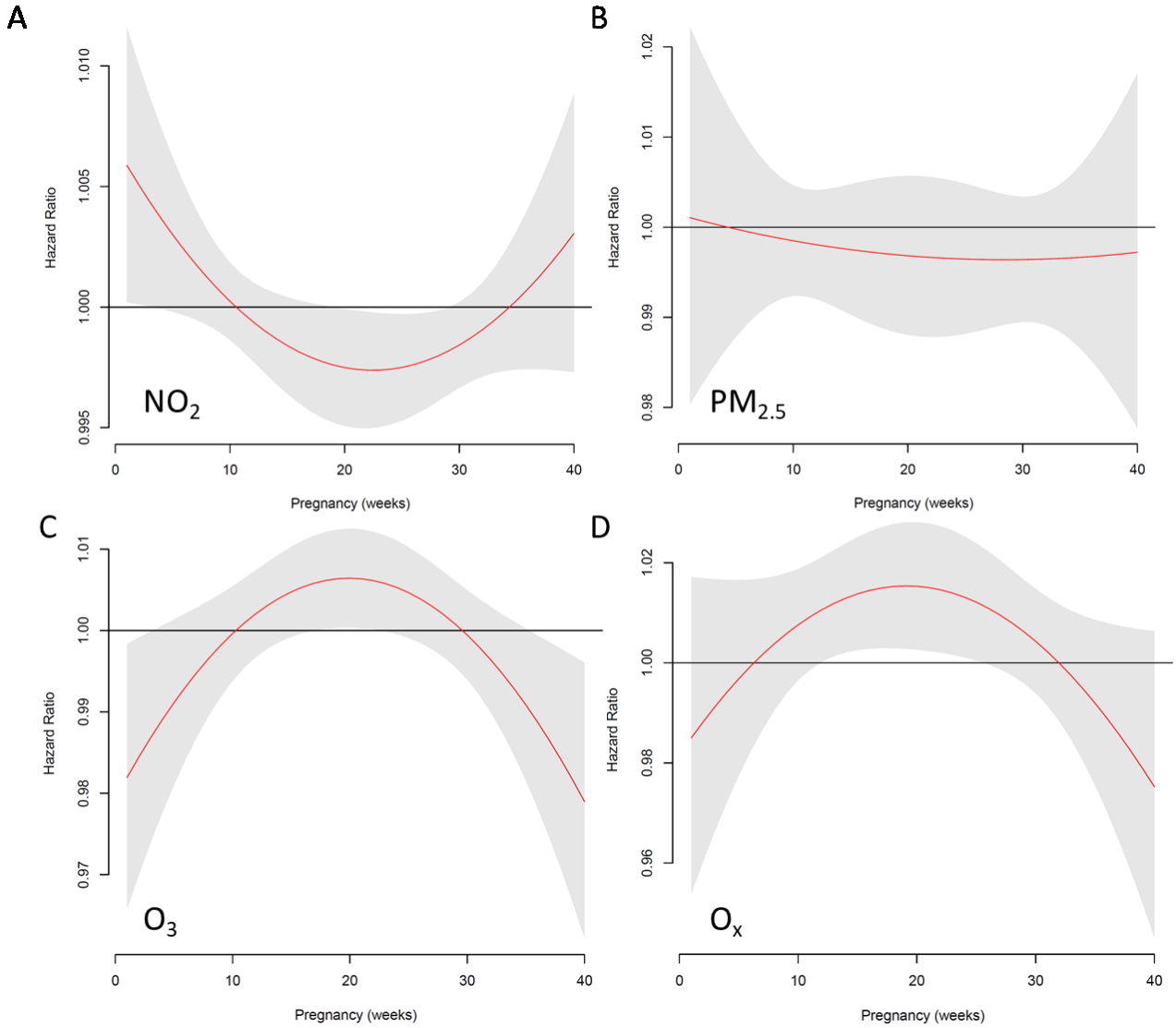


Figure 7. Curves of the associations between in-utero air pollutant exposures and pediatric-onset IBD. Hazard ratios (red lines) and 95% confidence intervals (shaded areas) present the risk of developing pediatric-onset IBD per interquartile range increase for (A) NO_2, 10.1 ppb, (B) $PM_{2.5}$, 3.5 ug/m3, (C) O_3, 5.8 ppb, and (D) O_x, 3.6 ppb during the pregnancy period after adjustment for childhood air pollutant exposures, sex, maternal IBD status, rural/urban residence and neighborhood income quintile.

Table 4. Associations between exposure to air pollutants and risk of overall IBD

Pollutant	Exposure period	Cohort size (Cases of IBD)	HR (95% CI)
NO_2	Trimester 1	1,523,588 (1645)	1.03 (1.00 - 1.07)
	Trimester 2		0.97 (0.94 - 1.00)
	Trimester 3		1.00 (0.97 - 1.03)
	Pregnancy		1.00 (0.96 - 1.03)
	Childhood		0.96 (0.86 - 1.06)
$PM_{2.5}$	Trimester 1	1,230,762 (793)	0.99 (0.86 - 1.15)
	Trimester 2		0.96 (0.86 - 1.07)
	Trimester 3		0.95 (0.84 - 1.08)
	Pregnancy		0.91 (0.79 - 1.04)
	Childhood		0.98 (0.86 - 1.12)
O_3	Trimester 1	1,671,124 (1748)	0.90 (0.80 - 1.02)
	Trimester 2		1.08 (1.00 - 1.16)
	Trimester 3		0.90 (0.80 - 1.01)
	Pregnancy		0.90 (0.73 - 1.10)
	Childhood		0.99 (0.94 - 1.04)
O_x	Trimester 1	956,421 (873)	0.97 (0.77 - 1.23)
	Trimester 2		1.21 (1.03 - 1.42)
	Trimester 3		0.93 (0.76 - 1.14)
	Pregnancy		1.12 (0.79 - 1.59)
	Childhood		1.08 (1.01 - 1.16)

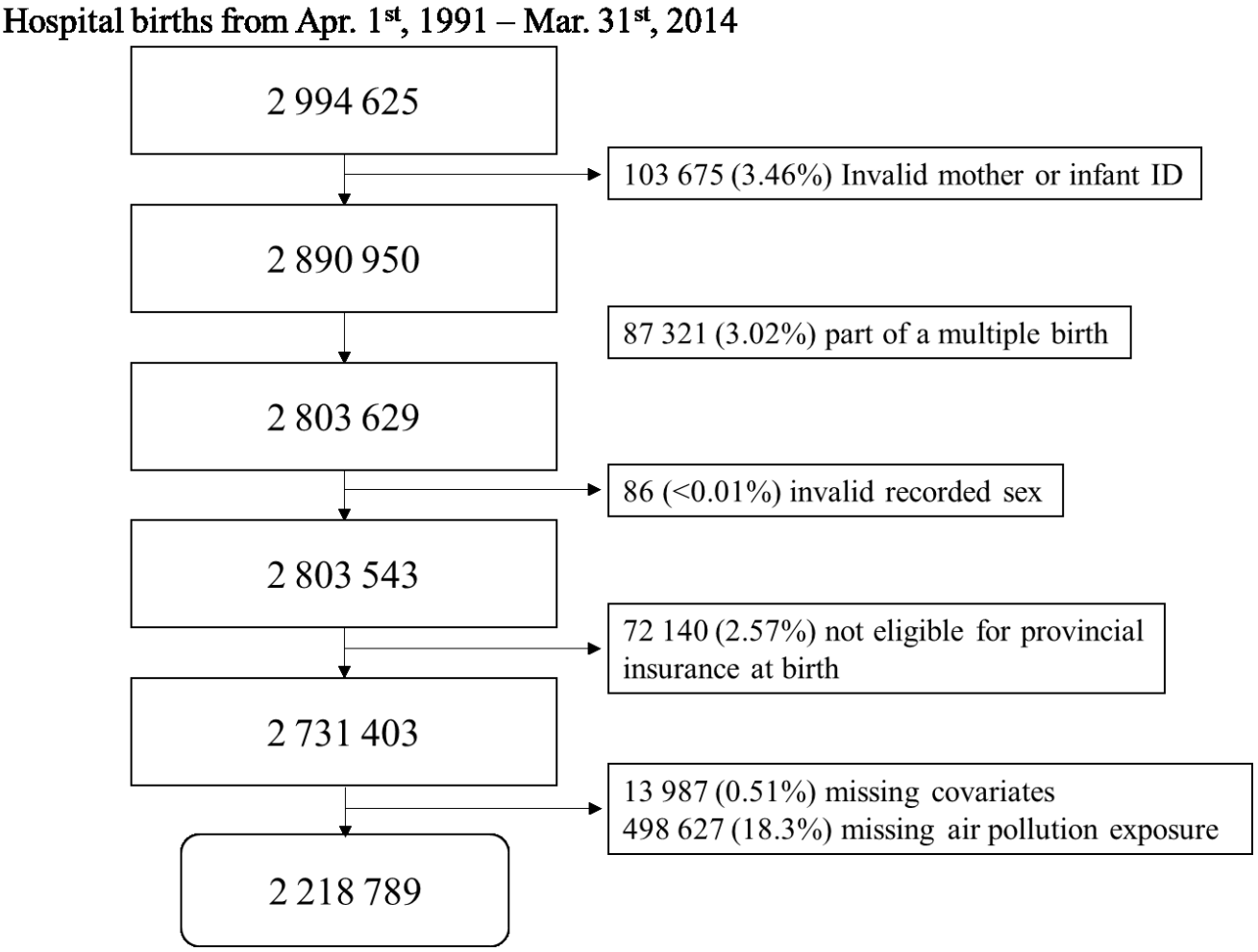
Associations are for an interquartile range increase in each combination of pollutant and exposure period. All models were adjusted for the crossbasis representing weekly pregnancy exposures, time-varying childhood exposures, sex, maternal IBD status, rural/urban residence, and neighborhood income quintile. HR: hazard ratio, CI: confidence interval.

Table 5. Associations between exposure to air pollutants and risk of Crohn's disease and ulcerative colitis

Pollutant	Exposure period	Crohn's disease		Ulcerative colitis	
		Cohort size (Cases of CD)	HR (95% CI)	Cohort size (Cases of UC)	HR (95% CI)
NO_2	Trimester 1	1,522,856 (913)	1.03 (0.98 - 1.09)	1,522,524 (581)	1.04 (0.98 - 1.11)
	Trimester 2		0.94 (0.91 - 0.98)		1.01 (0.96 - 1.06)
	Trimester 3		1.03 (0.98 - 1.08)		0.95 (0.90 - 1.01)
	Pregnancy		1.00 (0.95 - 1.04)		1.00 (0.94 - 1.06)
	Childhood		0.92 (0.80 - 1.06)		1.01 (0.85 - 1.20)
$PM_{2.5}$	Trimester 1	1,230,377 (408)	1.06 (0.87 - 1.29)	1,230,274 (305)	0.98 (0.77 - 1.23)
	Trimester 2		0.90 (0.77 - 1.04)		1.05 (0.87 - 1.25)
	Trimester 3		1.01 (0.85 - 1.19)		0.84 (0.68 - 1.02)
	Pregnancy		0.94 (0.79 - 1.14)		0.86 (0.69 - 1.07)
	Childhood		1.00 (0.83 - 1.19)		0.95 (0.77 - 1.17)
O_3	Trimester 1	1,670,329 (953)	1.02 (0.87 - 1.21)	1,670,023 (647)	0.83 (0.67 - 1.02)
	Trimester 2		1.06 (0.96 - 1.18)		1.09 (0.96 - 1.24)
	Trimester 3		0.94 (0.80 - 1.09)		0.89 (0.74 - 1.08)
	Pregnancy		1.02 (0.77 - 1.34)		0.84 (0.59 - 1.18)
	Childhood		1.00 (0.93 - 1.07)		0.94 (0.87 - 1.03)
O_x	Trimester 1	956,018 (470)	1.19 (0.86 - 1.63)	955,869 (321)	0.83 (0.56 - 1.22)
	Trimester 2		1.18 (0.95 - 1.47)		1.26 (0.97 - 1.64)
	Trimester 3		0.92 (0.70 - 1.21)		0.86 (0.61 - 1.21)
	Pregnancy		1.28 (0.80 - 2.05)		0.94 (0.52 - 1.70)
	Childhood		1.05 (0.96 - 1.16)		1.10 (0.98 - 1.24)

Hazard ratios (HR) and 95% confidence intervals (CI) are for an interquartile range increase in each combination of pollutant and exposure period. All models mutually adjusted for crossbasis representing weekly pregnancy exposures, and time-varying childhood exposures to the given pollutant, as well as child sex, maternal IBD status, rural/urban residence, and neighborhood income quintile. *indicate an association with a p-value <0.05.

3.9 Supplementary material for Chapter 3



Supplementary Figure 1. (Air Pollution) Study population flowchart

Supplementary Table 1. Demographic characteristics at birth for children that were excluded due to missing covariate or exposure information, stratified by IBD status.

Characteristic	IBD cases (n=983)	Non-IBD cases (n=511 631)
Sex		
Male	579 (59.9%)	261 431 (51.1%)
Female	404 (41.1%)	250 200 (48.9%)
Median neighbourhood income quintile		
1st quintile (lowest)	140 (14.2%)	97 482 (19.1%)
2nd quintile	199 (20.2%)	97 228 (19.0%)
3rd quintile	174 (17.7%)	104 014 (20.3%)
4th quintile	189 (19.2%)	97 753 (19.1%)
5th quintile (highest)	261 (26.6%)	103 113 (20.2%)
Missing	20 (2.0%)	12 041 (2.4%)
Mother with IBD		
Yes	31 (3.2%)	2 006 (0.4%)
No	952 (96.8%)	509 625 (99.6%)
Residence at birth		
Urban	848 (86.3%)	424 845 (83.0%)
Rural	110 (11.2%)	73 199 (14.3%)
Missing	25 (2.5%)	13 587 (2.7%)
Maternal age (years)	30.2 (4.9)	29.4 (5.4)
Birth weight (grams)	3 486 (540)	3 433 (560)
Gestational age (weeks)	38.8 (1.3)	38.8 (1.5)

Results reported as counts (percentages) for categorical variables or means (standard deviations) for continuous variables. Percentages may not sum to 100 due to rounding.

Supplementary Table 2. Descriptive statistics and Pearson correlation coefficients of pollutants across pre-specified exposure periods

Pollutant	Exposure period	Mean (SD)	IQR	Pearson Correlation Coefficients																			
				NO_2					$PM_{2.5}$					O_3					O_x					
				Trimester 1	Trimester 2	Trimester 3	Pregnancy	Childhood	Trimester 1	Trimester 2	Trimester 3	Pregnancy	Childhood	Trimester 1	Trimester 2	Trimester 3	Pregnancy	Childhood	Trimester 1	Trimester 2	Trimester 3	Pregnancy	Childhood
NO_2	Trimester1	12.70 (7.78)	10.46	1																			
	Trimester 2	12.60 (7.70)	10.31	0.85	1																		
	Trimester 3	12.39 (7.65)	10.19	0.73	0.85	1																	
	Pregnancy	12.73 (7.43)	10.08	0.92	0.96	0.92	1																
	Childhood	10.17 (5.80)	7.91	0.80	0.80	0.80	0.86	1															
$PM_{2.5}$	Trimester 1	7.96 (2.87)	3.75	0.41	0.44	0.43	0.46	0.36	1														
	Trimester 2	7.90 (2.81)	3.68	0.38	0.42	0.46	0.45	0.39	0.75	1													
	Trimester 3	7.91 (2.86)	3.73	0.43	0.39	0.42	0.44	0.39	0.73	0.76	1												
	Pregnancy	7.96 (2.66)	3.53	0.45	0.46	0.48	0.49	0.42	0.91	0.92	0.91	1											
	Childhood	7.49 (2.20)	2.99	0.36	0.37	0.37	0.39	0.42	0.64	0.67	0.68	0.73	1										
O_3	Trimester 1	24.16 (7.36)	11.66	-0.29	-0.32	0.04	-0.21	-0.11	-0.03	0.09	-0.01	0.02	-0.02	1									
	Trimester 2	24.37 (7.19)	11.35	0.07	-0.28	-0.32	-0.18	-0.09	-0.13	-0.03	0.08	-0.03	-0.01	0.11	1								
	Trimester 3	24.70 (7.23)	11.36	0.01	0.06	-0.29	-0.08	-0.14	0.00	-0.15	-0.03	-0.07	-0.01	-0.65	0.12	1							
	Pregnancy	24.14 (4.58)	5.79	-0.15	-0.36	-0.39	-0.32	-0.23	-0.11	-0.06	0.03	-0.05	-0.03	0.33	0.83	0.31	1						
	Childhood	25.52 (2.68)	2.89	-0.29	-0.21	-0.18	-0.24	-0.29	0.02	0	-0.03	0	-0.03	0.27	0.05	0.22	0.37	1					
O_x	Trimester 1	20.46 (4.62)	7.11	0.23	0.13	0.42	0.27	0.31	0.18	0.29	0.22	0.25	0.17	0.86	0.16	-0.65	0.26	0.12	1				
	Trimester 2	20.57 (4.52)	6.95	0.51	0.24	0.12	0.32	0.32	0.10	0.19	0.28	0.21	0.18	-0.05	0.86	0.15	0.65	-0.06	0.22	1			
	Trimester 3	20.71 (4.53)	6.94	0.39	0.49	0.21	0.39	0.27	0.21	0.08	0.18	0.18	0.18	-0.65	-0.03	0.87	0.12	0.13	-0.45	0.22	1		
	Pregnancy	20.41 (3.10)	3.63	0.65	0.50	0.43	0.57	0.52	0.29	0.33	0.40	0.37	0.31	0.10	0.57	0.21	0.60	0.11	0.45	0.83	0.44	1	
	Childhood	20.37 (2.09)	2.12	0.46	0.52	0.55	0.54	0.60	0.33	0.34	0.32	0.36	0.35	0.13	-0.03	0.07	0.12	0.55	0.38	0.24	0.35	0.56	1

Supplementary Table 3. Associations between exposure to air pollutants and risk of overall IBD stratified by rural / urban residence at birth

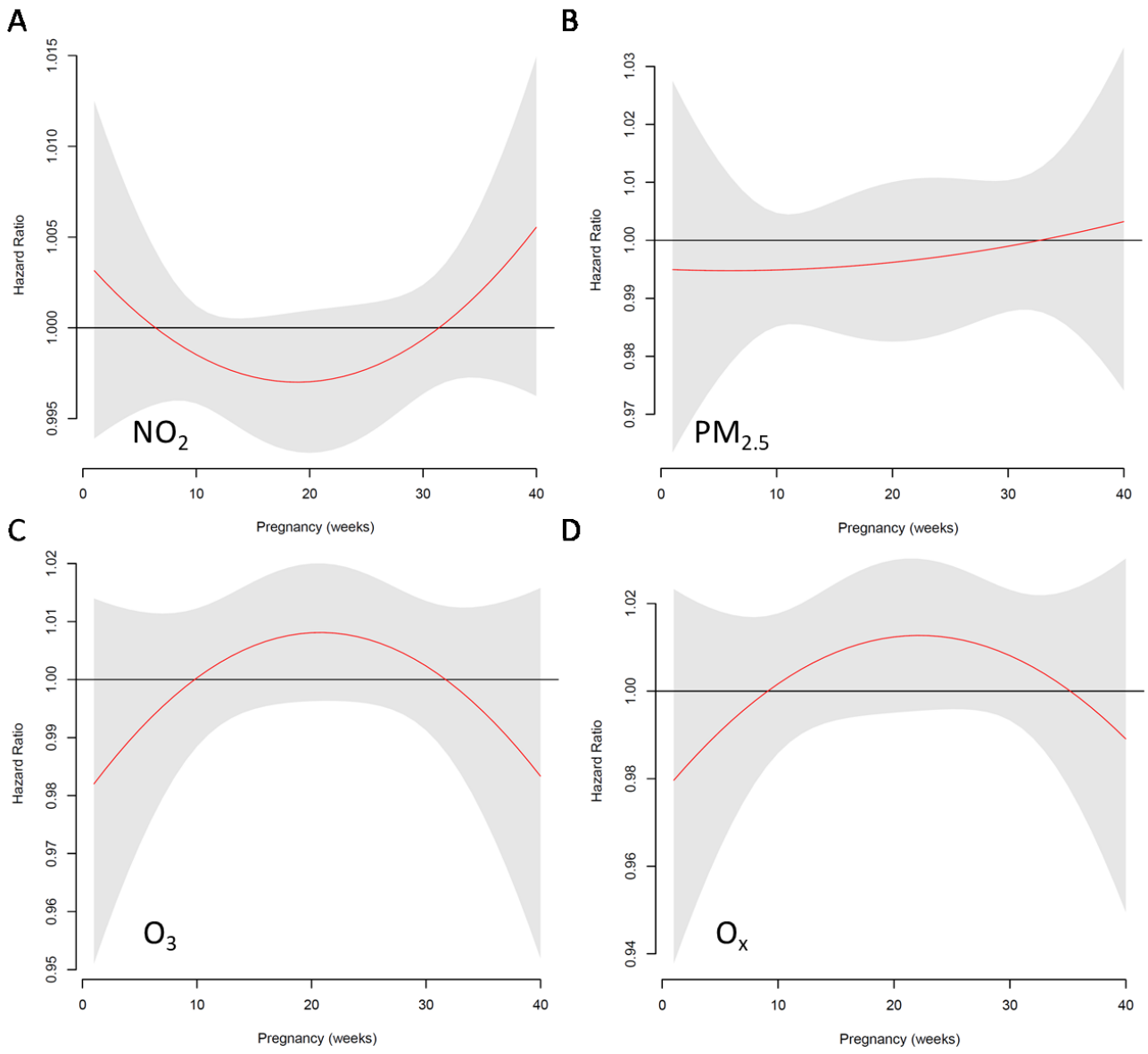
Pollutant	Exposure period	Urban		Rural		P value for effect measure modification
		Cohort size (IBD cases)	HR (95% CI)	Cohort size (IBD cases)	HR (95% CI)	
NO_2	Trimester 1	1 379 075 (1 526)	1.03 (0.99 - 1.07)	144 513 (119)	1.42 (1.10 - 1.84)	0.01
	Trimester 2		0.97 (0.94 - 1.00)		0.87 (0.71 - 1.06)	0.32
	Trimester 3		1.00 (0.97 - 1.04)		0.92 (0.72 - 1.18)	0.61
	Pregnancy		1.00 (0.96 - 1.03)		1.08 (0.82 - 1.42)	0.19
	Childhood		0.96 (0.86 - 1.07)		0.76 (0.35 - 1.66)	0.33
$PM_{2.5}$	Trimester 1	1 101 172 (743)	0.98 (0.84 - 1.13)	129 590 (50)	1.53 (0.73 - 3.23)	0.19
	Trimester 2		0.97 (0.86 - 1.08)		0.81 (0.45 - 1.45)	0.74
	Trimester 3		0.96 (0.85 - 1.09)		0.86 (0.44 - 1.70)	0.95
	Pregnancy		0.90 (0.79 - 1.04)		1.00 (0.50 - 1.99)	0.22
	Childhood		1.00 (0.87 - 1.14)		0.74 (0.39 - 1.43)	0.16
O_3	Trimester 1	1 519 251 (1653)	0.90 (0.80 - 1.03)	151 873 (95)	0.83 (0.50 - 1.39)	0.33
	Trimester 2		1.06 (0.98 - 1.14)		1.59 (1.11 - 2.27)	0.05
	Trimester 3		0.89 (0.79 - 1.00)		1.14 (0.71 - 1.84)	0.66
	Pregnancy		0.86 (0.70 - 1.07)		1.59 (0.70 - 3.60)	0.49
	Childhood		0.98 (0.93 - 1.03)		1.11 (0.91 - 1.35)	0.47
O_x	Trimester 1	881 232 (831)	0.99 (0.78 - 1.26)	75 189 (42)	0.84 (0.26 - 2.69)	0.66
	Trimester 2		1.17 (0.99 - 1.38)		3.08 (1.35 - 7.04)	0.03
	Trimester 3		0.92 (0.75 - 1.14)		1.24 (0.44 - 3.52)	0.73
	Pregnancy		1.08 (0.76 - 1.55)		3.56 (0.57 - 22.06)	0.25
	Childhood		1.08 (1.01 - 1.16)		1.06 (0.77 - 1.46)	0.43

All models were adjusted for the crossbasis representing weekly pregnancy exposures, time-varying childhood exposures, sex, maternal IBD and neighborhood income quintile. P values correspond to the multiplicative interaction term between the exposure and effect modifier (from a separate model).

Supplementary Table 4. Associations between exposure to air pollutants and overall IBD stratified by maternal IBD status

Pollutant	Exposure period	No maternal IBD		Maternal IBD		P value for effect measure modification
		Cohort size (IBD cases)	HR (95% CI)	Cohort size (IBD cases)	HR (95% CI)	
NO_2	Trimester 1	1 516 620 (1 601)	1.04 (1.00-1.08)	6 968 (44)	0.97 (0.77-1.22)	0.48
	Trimester 2		0.97 (0.94-1.00)		1.04 (0.86-1.26)	0.57
	Trimester 3		1.00 (0.96-1.03)		1.02 (0.82-1.26)	0.98
	Pregnancy		1.00 (0.96-1.03)		1.03 (0.82-1.30)	0.85
	Childhood		0.95 (0.86-1.06)		1.23 (0.62-2.43)	0.58
$PM_{2.5}$	Trimester 1	1 224 609 (768)	0.99 (0.86-1.14)	6 153 (25)	1.16 (0.50-2.71)	0.96
	Trimester 2		0.94 (0.85-1.05)		1.55 (0.82-2.95)	0.27
	Trimester 3		0.96 (0.84-1.08)		0.88 (0.45-1.75)	0.27
	Pregnancy		0.89 (0.78-1.02)		1.62 (0.75-3.47)	0.89
	Childhood		0.95 (0.83-1.09)		2.03 (1.17-3.52)	0.06
O_3	Trimester 1	1 663 515 (1700)	0.92 (0.81-1.04)	7 609 (48)	0.53 (0.24-1.14)	0.28
	Trimester 2		1.08 (1.00-1.17)		0.95 (0.59-1.51)	0.79
	Trimester 3		0.92 (0.82-1.03)		0.46 (0.22-0.94)	0.10
	Pregnancy		0.93 (0.75-1.14)		0.25 (0.06-0.98)	0.10
	Childhood		0.99 (0.94-1.04)		0.86 (0.63-1.18)	0.80
O_x	Trimester 1	956 421 (852)	0.99 (0.78-1.25)	4 482 (21)	0.55 (0.12-2.54)	0.81
	Trimester 2		1.21 (1.03-1.43)		1.09 (0.38-3.10)	0.90
	Trimester 3		0.94 (0.76-1.15)		0.56 (0.15-2.07)	0.80
	Pregnancy		1.14 (0.80-1.63)		0.37 (0.03-4.15)	0.82
	Childhood		1.09 (1.01-1.17)		0.84 (0.52-1.36)	0.51

All models were adjusted for the crossbasis representing weekly pregnancy exposures, time-varying childhood exposures, sex, rural/urban residence, and neighborhood income quintile. P values correspond to the multiplicative interaction term between the exposure and effect modifier (from a separate model).



Supplementary Figure 2. Curves of the associations between in-utero air pollutant exposures and pediatric-onset IBD before age 10. Hazard ratios (red lines) and 95% confidence interval bands (shaded areas) show the risk of developing IBD before the age of 10 per interquartile range increase for (A) NO_2, 10.1ppb, (B) $PM_{2.5}$, 3.5 ug/m3, (C) O_3, 5.8 ppb, and (D) O_x, 3.6 ppb during the pregnancy period after adjustment for childhood exposures, sex, maternal IBD status, rural/urban residence and neighborhood income

Supplementary Table 5. Associations between exposure to air pollutants and risk of IBD before age 10

Pollutant	Exposure period	Cohort size (Cases of IBD)	HR (95% CI)
NO_2	Trimester 1	1 517 393 (469)	1.01 (0.95 - 1.07)
	Trimester 2		0.96 (0.92 - 1.01)
	Trimester 3		1.02 (0.96 - 1.08)
	Pregnancy		0.98 (0.93 - 1.05)
	Childhood		0.90 (0.73 - 1.12)
$PM_{2.5}$	Trimester 1	1 230 496 (386)	0.94 (0.75 - 1.17)
	Trimester 2		0.95 (0.80 - 1.13)
	Trimester 3		1.01 (0.83 - 1.22)
	Pregnancy		0.90 (0.72 - 1.12)
	Childhood		0.92 (0.75 - 1.12)
O_3	Trimester 1	1 666 064 (512)	0.91 (0.71 - 1.16)
	Trimester 2		1.10 (0.95 - 1.28)
	Trimester 3		0.95 (0.76 - 1.19)
	Pregnancy		0.97 (0.63 - 1.48)
	Childhood		1.02 (0.92 - 1.12)
O_x	Trimester 1	952 298 (303)	0.90 (0.65 - 1.25)
	Trimester 2		1.17 (0.93 - 1.45)
	Trimester 3		1.03 (0.78 - 1.35)
	Pregnancy		1.11 (0.66 - 1.85)
	Childhood		1.20 (0.91 - 1.58)

Associations are for an interquartile range increase in each combination of pollutant and exposure period. All models were adjusted for the crossbasis representing weekly pregnancy exposures, the time-varying childhood exposure, sex, maternal IBD status, rural/urban residence, and neighborhood income quintile. HR: hazard ratio, CI: confidence interval.

Supplementary Table 6. Associations between exposure to air pollutants and risk of overall IBD while excluding extrapolated exposures

Pollutant	Exposure period	Cohort size (Cases of IBD)	HR (95% CI)
NO_2	Trimester 1	1 137 010 (665)	1.06 (1.01-1.11)
	Trimester 2		0.98 (0.94-1.01)
	Trimester 3		0.99 (0.94-1.03)
	Pregnancy		1.01 (0.96-1.06)
	Childhood		0.93 (0.75-1.15)
$PM_{2.5}$	Trimester 1	966 289 (273)	1.04 (0.75-1.42)
	Trimester 2		0.82 (0.64-1.04)
	Trimester 3		0.90 (0.69-1.18)
	Pregnancy		0.75 (0.53-1.07)
	Childhood		0.80 (0.60-1.06)
O_3	Trimester 1	1 353 826 (885)	0.89 (0.71-1.12)
	Trimester 2		1.13 (1.00-1.29)
	Trimester 3		0.93 (0.75-1.15)
	Pregnancy		0.96 (0.61-1.50)
	Childhood		1.03 (0.92-1.14)
O_x	Trimester 1	468 914 (252)	0.93 (0.54-1.60)
	Trimester 2		1.41 (1.00-1.98)
	Trimester 3		0.96 (0.60-1.52)
	Pregnancy		1.29 (0.48-3.47)
	Childhood		1.14 (0.94-1.37)

Associations are for an interquartile range increase in each combination of pollutant and exposure period. All models were adjusted for the crossbasis representing weekly pregnancy exposures, the time-varying childhood exposure, sex, maternal IBD status, rural/urban residence, and neighborhood income quintile. HR: hazard ratio, CI: confidence interval.

- End of Chapter 3 -

Chapter 4: Greenspace and paediatric-onset inflammatory bowel disease

This chapter is based on the manuscript *Exposure to residential greenspace during childhood reduces the risk of paediatric-onset inflammatory bowel disease: a population-based cohort study*, which has been submitted to the journal *Gut*. This chapter addresses Objective 2 of this thesis, to investigate the association between greenspace and the occurrence of IBD <18 years. Ethics approval for this study was granted by the Research Ethics Boards of Health Canada and the Ottawa Health Science Network. The completed RECORD statement checklist for this manuscript and the confirmation of submission to *Gut* can be found in the Appendices.

Authors: Michael Elten[1,2], Eric I. Benchimol[1,3,4,5,6], Deshayne B. Fell[1,3,4], M. Ellen Kuenzig[1,3,4,5], Glenys Smith[4], Gilaad G. Kaplan[7], Hong Chen[4,8,9,10], Dan Crouse[11,12], Éric Lavigne[1,2,4]

1. School of Epidemiology and Public Health, University of Ottawa, Ottawa Canada

2. Air Health Sciences Division, Health Canada, Ottawa Canada

3. Children's Hospital of Eastern Ontario (CHEO) Research Institute, Ottawa Canada

4. ICES uOttawa, Ottawa, Canada

5. CHEO Inflammatory Bowel Disease Centre, Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Eastern Ontario, Ottawa Canada

6. Department of Pediatrics, University of Ottawa, Ottawa Canada

7. Departments of Medicine and Community Health Sciences, University of Calgary, Calgary Canada

8. Public Health Ontario, Toronto Canada

9. Dalla Lana School of Public Health, University of Toronto, Toronto Canada

10. Environmental Health Science and Research Bureau, Health Canada, Ottawa Canada

11. Department of Sociology, University of New Brunswick, Fredericton Canada

12. New Brunswick Institute for Research, Data, and Training, Fredericton Canada

Author contributions: ME, EIB, and ÉL conceived the study design. ME and GS prepared the data. ME conducted all analysis. All authors helped interpret the results, and provided critical feedback and revisions to the final manuscript

4.1 Link to previous manuscript

In Chapter 3, it was determined that exposures to one measure ambient air pollutants in childhood was associated with an increased risk of paediatric IBD. This gives further evidence to the theory that environmental exposures can modify one's risk of developing paediatric IBD, and that an exposure related to industrialization may increase ones risk. Similar to air pollution, there has been some preliminary evidence linking the gut microbiome to factors related to residential greenspace and vegetation, such as physical activity and biodiversity. In this analysis, the independent effects of residential greenness on the risk of paediatric-onset IBD were explored. We also adjusted for exposures to air pollution to see whether the possible effect of greenspace might be confounded by air pollution exposures.

4.2 Abstract

Objective: Environmental factors related to urbanization and industrialization are believed to be involved in inflammatory bowel disease (IBD) development, but no study has looked at the association between greenspace and IBD. This study investigates the potential association between residential greenspace and paediatric-onset IBD.

Design: We conducted a retrospective cohort study using linked population-based health administrative and environmental datasets. The study population comprised 2 715 318 mother-infant pairs from hospital births in Ontario, Canada between April 1st, 1991 and March 31st, 2014. We measured exposure to residential greenspace using

the normalized difference vegetation index (NDVI) derived using remote-sensing methods. Average greenness was estimated for the pregnancy and childhood periods. We used mixed-effects Cox proportional hazards models to assess potential associations between residential greenness and risk of developing IBD before 18 years, while adjusting for covariates including sex, maternal IBD, rural/urban residence at birth and neighbourhood income.

Results: There were 3 444 IBD diagnoses that occurred during follow-up. An increase in the interquartile range of residential greenness during the childhood period was associated with a lower risk of developing paediatric-onset IBD (hazard ratio [HR] 0.77, 95% confidence interval [CI] 0.74-0.81). This relationship was significant for both ulcerative colitis (HR 0.72, 95% CI 0.67-0.78) and Crohn's disease (HR 0.81, 95% CI 0.76-0.87). There was a linear dose-response across increasing quartiles of greenspace ($P < 0.0001$). No consistent association was detected between maternal intra-partum greenspace exposure and paediatric-onset IBD.

Conclusion: Residential greenspace during childhood was associated with a reduced risk of IBD.

Keywords: inflammatory bowel disease; Crohn's disease; ulcerative colitis; greenspace;

4.3 Summary

What is already known about this subject?

- Environmental exposures related to industrialization and urbanization are believed to be associated with IBD incidence, but the relationship between greenspace and IBD development is unknown.

What are the new findings?

- Residential greenspace during childhood was associated with a reduced risk of paediatric-onset IBD. This association was consistent for Crohn's disease and ulcerative colitis.

- This relationship showed a linear dose-response pattern and was strongest for very early onset IBD (diagnosed before 10 years of age).

How might it impact on clinical practice in the foreseeable future?

- Features related to increased residential greenspace during childhood may play an independent role in the development of IBD during childhood. Future focus on how these features impact the gut microbiome may lead to a new understanding of IBD etiology.

- Increased exposure to greenspace through better environmental health policy and urban planning could curb the increasing incidence of IBD among children.

4.4 Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are the two main subtypes of inflammatory bowel disease (IBD), a condition characterized by inflammation in the gastro-intestinal tract due to an inappropriate immune response.[1] The incidence of IBD increased in Western nations during the 20th century.[2] Newly industrialised countries have demonstrated a rapid increase in incidence more recently.[3–5] Further, rates of paediatric-onset IBD continue to rise globally.[6] One or more environmental factors related to Western culture and/or rapid socioeconomic development may be implicated in IBD pathogenesis.[3]

The health effects of exposure to features of natural vegetation, often referred to as greenspace is an emerging area of interest in environmental epidemiology. Urbanization has led to a decrease in the population's exposures to plants and the natural environment.[7] Despite this, greenspace has not been considered in any published analysis of IBD incidence. Currently, residential greenspace has been shown to protect against other immune-mediated conditions such as atopic sensitization and asthma.[8–10] Other environmental exposures that are typically negatively correlated with greenspace such as living in an urban residence and air pollution have been found to be positively associated with IBD risk, particularly among youth.[11–13]

Greenspace may improve health through many pathways, including the promotion of increased physical activity, psychosocial effects, decreased exposure to air pollution

and urban noise, or direct exposure to a wider range of biodiversity.[14,15] Several of these pathways may be implicated in modulating risk to IBD through effects on the gut microbiome. For instance, physical activity has been seen to lead to improvements in gut microbiome health.[16] The loss of greenspace inevitably leads to decreased exposure to microbes living in vegetation and soil, which may also contribute to alterations in the gut microbiome seen in IBD, given that regulation of the immune system depends in part on microbes found in the natural environment.[17].

The objective of this study was to investigate the potential independent effects of residential greenspace on the risk of developing paediatric-onset IBD.

4.5 Methods

4.5.1 Study setting and data sources

We conducted a retrospective cohort study using population-based health administrative data for the province of Ontario, Canada's most populous province (2014 population: 13.6 million).[18] We formed this cohort through deterministic linkage of various datasets at ICES, an independent non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyse health care and demographic data, without consent, for health system evaluation and improvement. Ontario health administrative data are housed at ICES through a comprehensive agreement with the Ontario Ministry of Health and Long-Term Care (MOHLTC) and collects information on all legal residents with a valid health card (>99% of the population). Authors had complete access to all uncleaned data used in

the study. Ethics approval for this study was granted by the Research Ethics Boards of Health Canada and the Ottawa Health Science Network.

The base of the study population was formed by the ICES-derived MOMBABY database, which links hospital admission records of mothers with their infants for all hospital births in Ontario using information from the Discharge Abstract Database which is provided by the Canadian Institute for Health Information. Linkage to the Registered Persons Database, a dataset provided by the MOHLTC that contains demographic information on those registered for universal provincial health insurance was also conducted. The data were then linked to the CENSUS dataset available at ICES to add information on area-level indicators from previous censuses of the population. Finally, the study population was further linked to the Ontario Crohn's and Colitis Cohort (OCCC) to determine IBD status. This ICES-derived cohort identifies persons with IBD according to algorithms that have been designed for both paediatric and adult populations, described below.[19,20] These datasets were linked using unique encoded identifiers and analysed at ICES.

4.5.2 Patient and public involvement

People with IBD were not directly involved in the conduct of this study. However, stakeholders and patients were involved in the 2018 Impact of Inflammatory Bowel Disease in Canada report (produced for Crohn's and Colitis Canada), which highlighted a need to better understand environmental factors involved in IBD development, particularly in children.[21,22]

4.5.3 Study population

We included all children born in an Ontario hospital between April 1st, 1991 and March 31st, 2014. We excluded mother-infant pairs who had invalid identification numbers, multiple-birth deliveries, children with an invalid or missing recorded biological sex, or who were ineligible for provincial health insurance at birth. Finally, children that had invalid or missing covariates or greenspace estimates during both pregnancy and childhood were excluded from the study population.

4.5.4 Outcome and exposure assessment

Our study outcome was defined as diagnosis with new-onset IBD before the age of 18 years, as identified by the OCCC validated algorithms. The algorithms were validated in Ontario and use IBD diagnostic codes from the International Classification of Diseases (ICD) -9 (555, 556) and ICD-10 (K50, K51). The optimal algorithm to capture paediatric-onset IBD cases required four physician contacts or two hospitalizations with IBD-related diagnostic codes within three years if a child had undergone a colonoscopy, and seven physician contacts or three hospitalizations with IBD-related diagnostic codes within three years if the child had not undergone a colonoscopy. This algorithm had the following properties when evaluated in the Ontario paediatric population: a sensitivity of 89.6%-91.1%, a specificity of 99.5%-100%, a positive predictive value of 59.2%-76.0%, and a negative predictive value of 99.9%-100%.[19] The diagnosis date for each child was the date with the first IBD diagnostic code. The diagnosis (CD or UC) associated with five of the last seven most recent outpatient visits was used to determine the disease subtype. This approach had a sensitivity of

95.1%, specificity of 86.0%, positive predictive value of 92.0%, and a negative predictive value of 91.2%. Cases where the subtype was unclassifiable were uncommon (8.5% of cases) and were included in the main analysis of overall IBD but were excluded from subtype analyses.

Exposure to greenspace was estimated using the satellite-derived normalized difference vegetation index (NDVI). A validated measure of neighbourhood greenspace,[23] NDVI is a continuous measure between -1 and 1, with negative values indicating water and positive values representing presence of green vegetation. For this study, we included only values greater than 0. We compiled NDVI readings from satellites Landsat 5 (1990-2011) and Landsat 8 (2013-2017) at a 30 metre (m) spatial resolution for each year of the study period, excluding 2012. These annual NDVI values were based on an average of readings taken approximately every 16 days based on the orbit pattern of the satellites.[24] We assigned the maximum of the annual NDVI values within a 250 m radius around the representative point (i.e., the corresponding geographic coordinates) of each individual's six-digit residential postal code during the growing season (May 1st to August 31st) for each year. In urban areas, the representative point for six-digit postal codes corresponds typically to one side of a street in a given block or the centre of an apartment building and has positional accuracy within approximately 100-160 m. Positional uncertainty is greater in rural areas. We then averaged these annual estimates over two exposure windows: the pregnancy period (when a pregnancy spanned 2 calendar years, the average was weighted by the number of weeks of gestation in each year) and exposures after birth

(weighted by weeks of life in each year when the birth or end of follow-up occur in the middle of a year) to get the pregnancy average and childhood average exposures, respectively. We chose to use a long-term exposure to greenspace during childhood rather than a time-varying method to get more stable estimates that were less prone to sampling variation, and to handle missing data (such as subjects whose follow-up included the year 2012 when no measures were taken due to the decommissioning of Landsat 5). Greenspace was modelled both as a continuous variable (range 0 to 1) and transformed into a categorical variable, with the study population split into quartiles of greenspace for both prenatal and postnatal exposures separately.

4.5.5 Statistical analysis

We used standard and mixed-effects Cox proportional hazards models to evaluate the relationship between exposure to residential greenspace and the risk of developing paediatric-onset inflammatory bowel disease, as well as CD and UC subtypes. Follow-up time in this study was measured as the child's age in years rounded to one decimal place, starting from birth and ending at the first of: diagnosis of IBD, age of 18 years, end of follow-up (March 31st, 2017), death, or ineligibility of provincial health insurance (primarily due to moving outside Ontario). We reported hazard ratios (HRs) and 95% confidence intervals (CI) for the increase in the interquartile range (IQR) of NDVI in the given exposure period.

We built models iteratively, starting with the univariate model (Model 1). Next, we built a fully adjusted model (Model 2) by adding potential confounders as covariates in the regression model. We selected the potential confounders *a priori* based on previous

published literature. Previous studies have shown that there is a male predominance among paediatric-onset IBD,[19] and that those living in urban areas,[12] those with family history of disease,[25] and those of higher socioeconomic status tended to have higher rates of IBD.[26] We therefore adjusted for the following variables: sex (male or female), maternal IBD status (yes or no, obtained from the OCCC),[19,20] rural or urban residence at birth (based on the Statistics Canada 2006 Census "rural area" definition for postal codes),[27] and median neighbourhood income quintile (based on the 2006 Canadian census of the population and calculated at the dissemination area level, spatial units of analysis that contain approximately 400-700 people), which is a validated proxy for individual-level income.[28]

As one's exposure to residential greenspace is heavily dependent on geographical location, there is a potential for covariance of greenspace with unmeasured risk factors of the population that tend to choose to live in these areas that could result in residual confounding. For instance, those that rely on active commuting may choose to live in certain areas that have better walking and biking infrastructure. To control for this, we fit a mixed-effects model (Model 3) to the data, with two levels of random intercepts corresponding to geographical variables representing spatial clusters. We used this two-level hierarchal model to fit the random effects of census divisions (large units, equivalent in size to counties) and then census tracts (smaller divisions of neighbourhoods available in cities) or census subdivisions (roughly representing municipalities and used when there was no census tract assigned). This spatial survival model may reduce the amount of confounding from unmeasured

characteristics,[29] and has been widely used in previous environmental epidemiological studies.[30,31]

We also explored further adjustments for exposure to ambient air pollution in Models 4-7. Average nitrogen dioxide (NO_2), fine particulate matter ($PM_{2.5}$), and ozone (O_3) exposures from birth to end of follow-up were separately added to the Models 4, 5, and 6 respectively to explore any potential confounding by air pollution. Finally, to control for all pollutants simultaneously, Model 7 was fitted with the previous adjustments and all three pollutants. More detailed information on how exposures to the ambient air pollutants were assessed and assigned to the study population are available in the 110.

As the gut microbiome is believed to be established early in life, we conducted a pre-planned sensitivity analysis to investigate the effects of very early-life exposures to residential greenspace by limiting the follow-up time to the first ten years of life. Here, all children were censored after ten years, their childhood greenspace estimates were recalculated for this period, and the iterative models were refit to this new data.

We explored how the effects of greenspace might differ by maternal IBD status, rural/urban residence at birth, and high/low concentrations of each of the air pollution measures (split at the median concentration). To do this, we stratified the population by each potential effect measure modifier, ran the fully adjusted mixed-effects models (Model 3) in each stratum, and then qualitatively compared the magnitude and direction of the HRs. We tested the statistical significance of the potential effect

measure modification using Wald's test of the interaction term derived from the cross product between the effect modifier and the main exposure.

To investigate possible dose-response relationships, we derived a categorical main exposure variable by splitting the study population into quartiles of greenspace. We reported HRs and 95% confidence intervals for the fully adjusted mixed-effects Cox proportional hazards model (Model 3) for each increasing quartile of greenspace, with the lowest quartile of greenspace acting as the reference group. We modelled exposure as a continuous variable (with values ranging from 1-4), and used Wald's test to investigate the statistical significance of the dose-response.

We conducted all analyses using R (version 3.1.4) packages survival (version: 2.42-3) and coxme (version: 2.2-5).[32–34]

4.6 Results

From a total of 2 994 625 initial mother-infant pairs, after exclusions there were 2 715 318 mother-infant pairs eligible for analysis (Supplementary Figure 3). Among them, 3 444 developed paediatric-onset IBD (1 906 CD, 1 246 UC, and 292 IBD type unclassifiable). The characteristics of the study population at birth are shown in Table 6. In general, those that developed IBD were more likely to be male, lived in higher income neighbourhoods, and were born in urban areas compared to those that did not develop IBD. The median exposure to greenspace for the study population during pregnancy and childhood were 0.69 and 0.72, respectively, while interquartile ranges were 0.10 and 0.08, respectively. Individuals were followed-up for a total of

33 115 706.8 years, for an average of about 12.2 years per person. The average follow-up time for those born before April 1st 1999 (those who would have been able to attain the maximum age of 18 years before the study cut off date) was 16.9 years.

Table 6. Demographic characteristics of the study population at birth stratified by disease status

Characteristic	IBD cases (n=3 444)	Non-IBD cases (n=2 711 874)
Sex		
Male	1 979 (57.5%)	1 390 409 (51.3%)
Female	1 465 (42.5%)	1 321 465 (48.7%)
Neighbourhood income quintile		
1st quintile (lowest)	566 (16.4%)	542 410 (20.0%)
2nd quintile	702 (20.4%)	542 361 (20.0%)
3rd quintile	677 (19.7%)	542 127 (20.0%)
4th quintile	692 (20.1%)	542 459 (20.0%)
5th quintile (highest)	807 (23.4%)	542 517 (20.0%)
Mother with IBD		
Yes	101 (2.9%)	11 950 (0.4%)
No	3 343 (97.1%)	2 699 924 (99.6%)
Residence at birth		
Urban	3 169 (92.0%)	2 417 502 (89.1%)
Rural	275 (8.0%)	294 372 (10.9%)
Maternal age (years)	30.0 (4.9)	29.5 (5.8)
Birth weight (grams)	3 444 (566)	3 411 (555)
Gestational age (weeks)	38.8 (1.5)	38.9 (1.6)

Results reported as counts (percentages) for categorical variables or means (standard deviations) for continuous variables. Percentages may not sum to 100 due to rounding.

The results of the iterative model building process are shown in Table 7. Residential greenspace during childhood was associated with a lower likelihood of developing IBD before the age of 18 based on the unadjusted model (Model 1 HR 0.79, CI 0.75-0.82). This association did not change substantially after adding additional covariates and random effects (Model 3 HR 0.77, 95% CI 0.74-0.81) or when also considering the effects of the three pollutants (Model 7 HR 0.82, 95% CI 0.77-0.87). These associations were consistent for CD (Model 3: HR 0.81, 95% CI 0.76-0.87) and UC (Model 3: HR 0.72, 95% CI 0.67-0.78). In contrast, there was no clear association between exposure to greenspace during pregnancy and risk of paediatric IBD, CD, or UC as the confidence intervals of most models crossed the null (Table 7).

Table 7. Associations between exposures to greenspace and paediatric-onset IBD and Crohn's disease and ulcerative colitis subtypes

	Overall IBD		Crohn's disease		Ulcerative colitis	
	NDVI during pregnancy	NDVI during childhood	NDVI during pregnancy	NDVI during childhood	NDVI during pregnancy	NDVI during childhood
Model 1: Unadjusted	0.95 (0.91-1.00)	0.79 (0.75-0.82)	1.02 (0.96-1.08)	0.84 (0.79-0.89)	0.89 (0.83-0.95)	0.72 (0.68-0.78)
Model 2: Adjusted*	0.96 (0.92-1.00)	0.78 (0.75-0.82)	1.03 (0.97-1.09)	0.84 (0.79-0.89)	0.89 (0.83-0.96)	0.72 (0.67-0.78)
Model 3: Model 2 + random effects	0.97 (0.93-1.02)	0.77 (0.74-0.81)	1.02 (0.96-1.09)	0.81 (0.76-0.87)	0.91 (0.85-0.98)	0.72 (0.67-0.78)
Model 4: Model 3 + NO_2	1.01 (0.96-1.07)	0.82 (0.78-0.87)	1.08 (1.00-1.16)	0.87 (0.81-0.94)	0.94 (0.87-1.03)	0.75 (0.69-0.82)
Model 5: Model 3 + $PM_{2.5}$	0.97 (0.92-1.02)	0.79 (0.75-0.83)	1.03 (0.96-1.11)	0.83 (0.77-0.89)	0.90 (0.83-0.98)	0.74 (0.68-0.80)
Model 6: Model 3 + O_3	0.97 (0.93-1.02)	0.78 (0.74-0.82)	1.03 (0.97-1.11)	0.82 (0.76-0.87)	0.91 (0.84-0.99)	0.74 (0.68-0.80)
Model 7: Model 3 + NO_2, $PM_{2.5}$, and O_3	1.01 (0.95-1.07)	0.82 (0.77-0.87)	1.10 (1.02-1.19)	0.87 (0.80-0.94)	0.92 (0.83-1.01)	0.75 (0.68-0.83)

Data are hazard ratios (95% confidence intervals) per interquartile range increase in NDVI for separate models evaluating the effects of greenspace during pregnancy (IQR=0.08) and during childhood (IQR=0.10). *covariates in Model 2 and all subsequent models were sex, rural/urban residence, maternal IBD status, and neighbourhood income quintile.

In the sensitivity analyses looking at only the first 10 years of exposures and follow-up times (Table 8), the effect of greenness was stronger for overall IBD (Model 3 HR 0.70, 95% CI 0.69-0.71), with very similar associations in CD (Model 3 HR 0.70, 95% CI 0.68-0.72) and UC (Model 3 HR 0.71, 95% CI 0.70-0.73). There were no statistically significant associations detected for exposures to greenspace during pregnancy.

Table 8. Associations between pregnancy and childhood exposures to greenspace and very early onset IBD, Crohn's disease and ulcerative colitis during the first 10 years of follow-up

	IBD		Crohn's disease		Ulcerative colitis	
	NDVI during pregnancy	NDVI during childhood	NDVI during pregnancy	NDVI during childhood	NDVI during pregnancy	NDVI during childhood
Model 1: Unadjusted	0.92 (0.84-1.00)	0.73 (0.72-0.74)	0.94 (0.83-1.07)	0.73 (0.71-0.74)	0.89 (0.79-1.01)	0.72 (0.71-0.74)
Model 2: Adjusted*	0.92 (0.84-1.00)	0.72 (0.71-0.73)	0.94 (0.82-1.07)	0.72 (0.71-0.74)	0.90 (0.79-1.03)	0.72 (0.71-0.74)
Model 3: Model 2 + random effects	0.92 (0.84-1.01)	0.70 (0.69-0.71)	0.92 (0.81-1.06)	0.70 (0.68-0.72)	0.92 (0.81-1.06)	0.71 (0.70-0.73)
Model 4: Model 3 + NO_2	1.02 (0.91-1.13)	0.71 (0.69-0.72)	1.00 (0.86-1.17)	0.71 (0.69-0.72)	1.00 (0.86-1.17)	0.72 (0.71-0.74)
Model 5: Model 3 + $PM_{2.5}$	0.93 (0.84-1.03)	0.70 (0.69-0.71)	0.93 (0.80-1.09)	0.70 (0.68-0.71)	0.94 (0.80-1.09)	0.71 (0.69-0.73)
Model 6: Model 3 + O_3	0.93 (0.84-1.02)	0.70 (0.69-0.71)	0.93 (0.80-1.07)	0.70 (0.68-0.72)	0.93 (0.81-1.08)	0.71 (0.70-0.73)
Model 7: Model 3 + NO_2, $PM_{2.5}$, and O_3	1.01 (0.89-1.14)	0.70 (0.69-0.72)	1.02 (0.85-1.23)	0.70 (0.68-0.72)	1.01 (0.84-1.20)	0.71 (0.69-0.73)

Data are hazard ratios (95% confidence intervals) per interquartile range increase in NDVI for separate models evaluating the effects of greenspace during pregnancy (IQR=0.08) and during childhood (IQR=0.10). *covariates in Model 2 and all subsequent models were sex, rural/urban residence, maternal IBD status, and neighbourhood income quintile

The results of the effect measure modification analyses are reported in Table 9. There were no major differences in the effect of childhood exposure to greenness on IBD by rurality of residence at birth (P=0.33), maternal IBD status (P=0.72) or by any of the air pollutants (P range: 0.33-0.93).

Table 9. Effect measure modification analysis for the association between childhood residential greenspace on overall paediatric-onset IBD

Effect modifier		Number of cases	Stratum-specific HR (95% CI)	P value for product interaction term
Rurality	Rural area	274	0.85 (0.69-1.03)	0.33
	Urban area	3168	0.77 (0.73-0.81)	
Maternal IBD	Yes	101	0.82 (0.63-1.06)	0.72
	No	3341	0.77 (0.74-0.81)	
NO_2	High	1703	0.81 (0.76-0.86)	0.93
	Low	1048	0.82 (0.74-0.90)	
$PM_{2.5}$	High	1706	0.79 (0.74-0.85)	0.33
	Low	1176	0.77 (0.71-0.83)	
O_3	High	1250	0.80 (0.74-0.87)	0.87
	Low	1788	0.77 (0.72-0.82)	

All models adjusted for sex, rural/urban residence, maternal IBD status, neighbourhood income quintile, and random effects.

Effect measure modification was also explored across disease subtypes (Supplementary Table 7 and Supplementary Table 8). Rurality did not modify the association between greenspace and CD (P=0.84), but the association between greenspace and UC seemed to be different for children in rural and urban areas (P=0.05). There was no association between greenspace and UC in rural areas (HR 1.00, 95% CI 0.71-1.42), but there was a protective effect in urban areas (HR 0.71, 95% CI 0.66-0.77). NO_2 did not modify the association of greenspace and IBD at all (P=0.93) but seemed to have divergent effects in CD and UC. For CD, those residing in areas high in NO_2 saw a slightly weaker protective effect than those living in areas lower in NO_2 (HR 0.89, 95% CI 0.81-0.97 and HR 0.79, 95% CI 0.69-0.89 respectively; P=0.05). For UC it was the opposite, where those in areas high in NO_2 had a somewhat stronger protective effect then those living in areas low in NO_2 (HR 0.71, 95% CI 0.64-0.79 and HR 0.87, 95% CI 0.73-1.03 respectively, P=0.05). There were no other notable differences across potential modifiers in the disease subtypes.

The results of the categorical analyses for childhood exposure to greenspace are reported in Table 10. When compared to the lowest quartile of greenness, each increasing quartile of greenness was associated with a decreasing risk of IBD, CD, and UC. The associations for overall IBD and being in quartile 2 (HR 0.77, 95% CI 0.71-0.84), quartile 3 (HR 0.72, 95% CI 0.66-0.79) and quartile 4 (HR 0.60, 95% CI 0.53-0.68) of greenness are indicative of a linear dose-response pattern (P value for linear trend <0.0001).

Table 10. Association between increasing quartiles of greenness during childhood and paediatric-onset IBD

Outcome	Quartile 1 NDVI: 0.106-0.677	Quartile 2 NDVI: 0.678-0.719	Quartile 3 NDVI: 0.720-0.754	Quartile 4 NDVI: 0.755-1.000	P value for linear trend
IBD	Reference	0.77 (0.71-0.84)	0.72 (0.66-0.79)	0.60 (0.53-0.68)	< 0.0001
Crohn's disease	Reference	0.79 (0.71-0.89)	0.78 (0.69-0.89)	0.64 (0.54-0.75)	< 0.0001
Ulcerative colitis	Reference	0.75 (0.66-0.87)	0.64 (0.55-0.75)	0.57 (0.46-0.69)	< 0.0001

Results are hazard ratios with 95% confidence intervals for the comparison of each quartile of greenspace with those in the lowest category of greenness. All models were adjusted for sex, rurality of residence at birth, maternal IBD status, neighbourhood income quintile, and two-level random effects

4.7 Discussion

In this retrospective population-based cohort study, we found that increasing levels of residential greenspace during childhood was associated with a reduced risk of paediatric-onset IBD. This association was robust, and persisted even after controlling for individual-level, geographical, and competing environmental risk factors (such as air pollution and rural/urban residence). These associations were observed for both subtypes and were stronger for early onset IBD, diagnosed before 10 years of age. A linear dose-response relationship was observed for all three outcomes when the main exposure was characterized as quartiles of greenspace. Greenspace during pregnancy did not show a consistent association with overall IBD, or CD and UC subtypes.

The observed protective association between residential greenspace and paediatric IBD may be explained by an increased exposure to microbes or physical activity. The possibility of exposure to biodiversity as a protective factor for IBD concurs with the widely supported “hygiene hypothesis”, which theorizes that the reduced microbial exposure and sterile environments that humans experience in industrialized nations may negatively impact the immune system. This reduced microbial load has been previously suggested to be responsible for the global increases in IBD incidence.[36] Bacteria and other organisms found in vegetation and soil may thus be the mediating factor linking increased greenspace to reduced risk of IBD among children. Helminths, for instance, are small parasitic worms found in soil that have infected and co-evolved with humans for centuries, and have an important role in immunoregulation of the

intestinal microbiome, and may reduce IBD risk.[37,38] When we assessed the effect of greenspace in urban and rural areas, the protective association was stronger in urban settings, and this effect measure modification was significant for UC. This may be because the increased overall microbial exposure in rural-dwellers results in them being less reliant on greenspace as a source of biodiversity.[35]

Physical activity in children promoted by proximity and access to greenspace such as by parks and playgrounds may also be exerting influence on IBD risk.[15,39] This may be due to changes in the composition of the gut microbiota caused by recreational exercise reported in several animal studies.[40] It is likely a combination of increased exposure to biodiversity, increased physical activity, and other factors associated with greenspace that work together to reduce the risk of developing IBD. Future studies should examine the composition and function of the microbiome in people exposed to different levels of greenspace.

This study is strengthened by its population-based design, which helps reduce possible selection biases, and ensures that the results of this study are generalizable to other regions with Canada, and to other countries that have similar levels of greenspace and population distributions. In addition, we were able to consider the effects of other environmental exposures that may be spatially correlated with greenspace, such as urban/rural residence and exposures to air pollutants as well as possible unmeasured confounding through the use of mixed-effects models. We were able to assess exposures throughout the entire life-course starting from the estimated date of conception. This allowed us to characterize both prenatal and postnatal effects

and we were able to determine that it is likely the childhood exposures to greenspace that are most relevant to paediatric-onset IBD development.

However, several limitations to this study should be considered. Firstly, exposure to greenspace tends to be spatially correlated with other environmental risk factors such as rurality and air pollution, so any protective effects may not be due to the primary action of the vegetation itself, but due to the rural environment, or the lack of pollution. We attempted to control for this in the study by directly adjusting for these factors in the regression models, but some residual confounding may still be present due to potential misclassification of these measures. Secondly, as this is an observational study, and areas of greenspace are not distributed randomly or equitably throughout the population, there may be important differences between the people who live in high and low greenspace areas that may also have affected their risk of developing IBD. For instance, those that reside in urban areas with high greenspace tend to be wealthier, are less likely to be ethnic minorities, and have better health literacy and behaviours.[41] To control for this possible confounding bias, neighbourhood income was added as a covariate in the models, and random effects representing geographical variables were also added to account for the clustering of unmeasured risk factors common to similar types of people that choose to live in the same areas. Another limitation of the study includes the use of secondary health-administrative data that were not collected for research purposes. This may mean that there are errors in diagnostic coding or changes in coding practice over the course of the study that are not accounted for, though as long-term estimates of greenspace were used, this would

be unlikely to greatly impact our results. Misclassification of outcome based on the algorithms is present as well, but this would be unlikely to differ greatly across greenspace levels, limiting the impact this would have on the observed associations. Residual confounding may be present, caused by confounders that were not available in the health administrative data sources, such as individual ethnicity. However, recent immigrants (that typically have lower rates of IBD) have lower average NDVI values in Canadian urban cities,[42] meaning that our associations may have been underestimated. Finally, the vast majority of Ontario's population live in urban areas, and the types of vegetation and other environmental exposures may be unique to this population, so caution should be exercised when interpreting the results of the present study and extrapolating to other geographic regions.

In summary, greater residential greenspace during childhood was associated with a reduced risk of IBD before the age of eighteen years. Future studies on environmental factors should consider exposures to greenspace and attempt to determine the specific features of greenspace responsible for this finding. Effort should also be made to investigate the interaction between greenspace and the gut microbiome, immune dysregulation, and genetic predisposition in IBD to identify possible causal mechanisms. In addition, intervention studies designed to target specific environmental risk factors to reduce the risk of IBD should include greenspace in study design. It is possible that by changing our immediate environment, we could prevent childhood-onset IBD.

4.8 Acknowledgements:

This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). This study also received funding from: Health Canada. Parts of this material are based on data and information compiled and provided by: MOHLTC, Canadian Institute for Health Information. The analyses, conclusions, opinions and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred.

NDVI metrics, indexed to DMTI Spatial Inc. postal codes, were provided by CANUE (Canadian Urban Environmental Health Research Consortium)

M. Ellen Kuenzig was supported by a Post-Doctoral Fellowship Award from the Canadian Institutes of Health Research (CIHR), Canadian Association of Gastroenterology (CAG), Crohn's and Colitis Canada, and Mitacs Canada. Eric Benchimol was supported by a New Investigator Award from the CIHR, CAG and Crohn's and Colitis Canada. Eric Benchimol was also supported by the Career Enhancement Program of the Canadian Child Health Clinician Scientist Program.

4.9 References for Chapter 4

1 Podolsky DK. Inflammatory bowel disease. N Engl J Med 1991;325:928–37.

2 Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 2012;142:46-54.e42. doi:10.1053/j.gastro.2011.10.001

3 Ng SC, Yun Shi H, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. www.thelancet.com 2017;390. doi:10.1016/S0140-6736(17)32448-0

4 Kotze PG, Underwood FE, Damião AOMC, et al. Progression of Inflammatory Bowel Diseases Throughout Latin America and the Caribbean: a Systematic Review. Clin Gastroenterol Hepatol 2019;:1–9. doi:10.1016/j.cgh.2019.06.030

5 Ng SC, Kaplan GG, Tang W, et al. Population Density and Risk of Inflammatory Bowel Disease: A Prospective Population-Based Study in 13 Countries or Regions in Asia-Pacific. Am J Gastroenterol 2018;:107–15. doi:10.1038/s41395-018-0233-2

6 Benchimol EI, Bernstein CN, Bitton A, et al. Trends in epidemiology of pediatric inflammatory bowel disease in Canada: Distributed network analysis of multiple population-based provincial health administrative databases. Am J Gastroenterol 2017;112:1120–34. doi:10.1038/ajg.2017.97

7 Pauleit S, Ennos R, Golding Y. Modeling the environmental impacts of urban land use and land cover change - A study in Merseyside, UK. Landsc Urban Plan 2005;71:295–310. doi:10.1016/S0169-2046(04)00083-0

8 Ruokolainen L, Von Hertzen L, Fyhrquist N, et al. Green areas around homes reduce atopic sensitization in children. Allergy Eur J Allergy Clin Immunol 2015;70:195–202. doi:10.1111/all.12545

9 Lovasi GS, Quinn JW, Neckerman KM, et al. Children living in areas with more street trees have lower prevalence of asthma. J Epidemiol Community Health 2008;62:647–9. doi:10.1136/jech.2007.071894

10 Sbihi H, Tamburic L, Koehoorn M, et al. Greenness and Incident Childhood Asthma: A 10-Year Follow-up in a Population-based Birth Cohort To. Am J Respir Crit Care Med 2015;192:1131–3.

11 Soon IS, Molodecky N, Rabi D, et al. The Relationship Between Urban Environment and Inflammatory Bowel Disease: A Systematic Review and Meta-

analysis. Am J Gastroenterol 2010;105:S465. doi:10.14309/00000434-201010001-01265

12 Benchimol EI, Kaplan GG, Otley AR, et al. Rural and Urban Residence during Early Life is Associated with a Lower Risk of Inflammatory Bowel Disease: A Population-Based Inception and Birth Cohort Study. Am J Gastroenterol 2017;112:1412–22. doi:10.1038/ajg.2017.208

13 Kaplan GG, Hubbard J, Korzenik J, et al. The inflammatory bowel diseases and ambient air pollution: a novel association. Am J Gastroenterol 2010;105:2412.

14 Markevych I, Schoierer J, Hartig T, et al. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. Environ Res 2017;158:301–17. doi:10.1016/j.envres.2017.06.028

15 Almanza E, Jerrett M, Dunton G, et al. A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. Heal Place 2012;18:46–54. doi:10.1016/j.healthplace.2011.09.003

16 Clarke SF, Murphy EF, O'Sullivan O, et al. Exercise and associated dietary extremes impact on gut microbial diversity. Gut 2014;63:1913–20. doi:10.1136/gutjnl-2013-306541

17 Rook GA. Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. Proc Natl Acad Sci 2013;110:18360–7. doi:10.1073/pnas.1313731110

18 Statistics Canada. Table 17-10-0005-01 Population estimates on July 1st, by age and sex. doi:https://doi.org/10.25318/1710000501-eng

19 Benchimol EI, Guttmann A, Griffiths AM, et al. Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. Gut 2009;58:1490–7. doi:10.1136/gut.2009.188383

20 Benchimol EI, Guttmann A, Mack DR, et al. Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. J Clin Epidemiol 2014;67:887–96. doi:10.1016/j.jclinepi.2014.02.019

21 Carroll MW, Kuenzig ME, Mack DR, et al. The Impact of Inflammatory Bowel Disease in Canada 2018: Children and Adolescents with IBD. J Can Assoc Gastroenterol 2019;2:S49–67. doi:10.1007/s005350300016

22 Benchimol EI, Bernstein CN, Bitton A, et al. The Impact of Inflammatory Bowel Disease in Canada 2018: A Scientific Report from the Canadian Gastro-Intestinal Epidemiology Consortium to Crohn's and Colitis Canada. J Can Assoc Gastroenterol 2019;2:S1–5. doi:10.1093/jcag/gwy052

23 Rhew IC, Vander Stoep A, Kearney A, et al. Validation of the normalized difference vegetation index as a measure of neighborhood greenness. Ann Epidemiol 2011;21:946–52.

24 Robinson NP, Allred BW, Jones MO, et al. A dynamic landsat derived normalized difference vegetation index (NDVI) product for the conterminous United States. Remote Sens 2017;9:1–14. doi:10.3390/rs9080863

25 Aujnarain A, Mack DR, Benchimol EI. The Role of the Environment in the Development of Pediatric Inflammatory Bowel Disease. Curr Gastroenterol Rep 2013;15. doi:10.1007/s11894-013-0326-4

26 Blanchard JF, Bernstein CN, Wajda A, et al. Small-area variations and sociodemographic correlates for the incidence of Crohn's disease and ulcerative colitis. Am J Epidemiol 2001;154:328–35. doi:10.1093/aje/154.4.328

27 du Plessis V, Beshiri R, Bollman RD, et al. Rural and Small Town Canada Analysis Bulletin. Rural Small T Canada Anal Bull 2001;3:1–17. doi:21-006-XIE

28 Glazier RH, Creatore MI, Agha MM, et al. Socioeconomic Misclassification in Ontario's Health Care Registry. Can J Public Heal 1996;94.

29 Sheppard L, Burnett RT, Szpiro AA, et al. Confounding and exposure measurement error in air pollution epidemiology. Air Qual Atmos Heal 2012;5:203–16. doi:10.1007/s11869-011-0140-9

30 Chen H, Burnett RT, Copes R, et al. Ambient fine particulate matter and mortality among survivors of myocardial infarction: Population-based cohort study. Environ Health Perspect 2016;124:1421–8. doi:10.1289/EHP185

31 Lavigne É, Bélair M-A, Do MT, et al. Maternal exposure to ambient air pollution and risk of early childhood cancers: A population-based study in Ontario, Canada. Environ Int 2017;100:139–47. doi:10.1016/j.envint.2017.01.004

32 R Core Team. R: A language and environment for statistical computing. 2019.https://www.r-project.org/

33 Therneau TM. A Package for Survival Analysis in S. 2015.https://cran.r-project.org/package=survival

34 Therneau TM. coxme: Mixed Effects Cox Models. 2018.https://cran.r-project.org/package=coxmeo

35 Ege MJ, Mayer M, Normand A, et al. Exposure to environmental microorganisms and childhood asthma. N Engl J Med 2011;364:701–9. doi:10.1056/NEJMoa1404595

36 Koloski NA, Bret L, Radford-Smith G. Hygiene hypothesis in inflammatory bowel disease: A critical review of the literature. World J Gastroenterol 2008;14:165–73. doi:10.3748/wjg.14.165

37 McSorley HJ, Maizels RM. Helminth infections and host immune regulation. Clin Microbiol Rev 2012;25:585–608. doi:10.1128/CMR.05040-11

38 Figueiredo CA, Barreto ML, Rodrigues LC, et al. Chronic intestinal helminth infections are associated with immune hyporesponsiveness and induction of a regulatory network. Infect Immun 2010;78:3160–7. doi:10.1128/IAI.01228-09

39 Wang Q, Xu KQ, Qin XR, et al. Association between physical activity and inflammatory bowel disease risk: A meta-analysis. Dig Liver Dis 2016;48:1425–31. doi:10.1016/j.dld.2016.08.129

40 Cook MD, Allen JM, Pence BD, et al. Exercise and gut immune function: Evidence of alterations in colon immune cell homeostasis and microbiome characteristics with exercise training. Immunol Cell Biol 2016;94:158–63. doi:10.1038/icb.2015.108

41 Pham TTH, Apparicio P, Séguin AM, et al. Spatial distribution of vegetation in Montreal: An uneven distribution or environmental inequity? Landsc Urban Plan 2012;107:214–24. doi:10.1016/j.landurbplan.2012.06.002

42 Villeneuve PJ, Jerrett M, G. Su J, et al. A cohort study relating urban green space with mortality in Ontario, Canada. Environ Res 2012;115:51–8. doi:10.1016/j.envres.2012.03.003

4.10 Supplementary material for Chapter 4

4.10.1 Air pollutants exposure assessment

Exposures to nitrogen dioxide (NO_2), particulate matter with diameter < 2.5 μm ($PM_{2.5}$), ozone (O_3) were assigned based on the centroid of the 6-digit postal for each member in the study population using surfaces created using the following methods.

Assessment of NO_2 exposures were based on a national land-use regression (LUR) model which used data from the National Air Pollution Surveillance (NAPS) monitoring network combined with satellite derived NO_2 estimates from 2005-2011, total distance of road lengths within 10 kilometres, area of industrial land use within 2 kilometres and mean summer rainfall.[1,2] To improve the model by capturing local variations in traffic, kernel density functions were used to describe the density of the roadways. This model explained 73% of the variation of NAPS measurements for NO_2 in 2006.

To estimate $PM_{2.5}$ exposures, satellite-derived estimates were combined with ground-level monitor data.[3] Total column aerosol optical depth measurements were generated at a 1 kilometre by 1 kilometre spatial resolution across North America from satellite retrievals and simulation. Next, a chemical transport model was used to relate these measurements to near-surface $PM_{2.5}$. Ground-based observations were then added to these geophysical estimates using geographically weighted regression. This provided an R^2 of 0.70 with cross-validated ground-based measurements over North America.

To calculate exposures to O_3, an average of daily 8-hour maximum concentrations during the warm seasons (May 1st to October 31st) was used to generate a surface at a 21 kilometre resolution using an optimal interpolation technique that was specially adapted for air pollutants.[4]

A spatiotemporal interpolation technique was then applied to the three pollutants above using monitors from the NAPS network to get exposure estimates on a weekly scale. First a scaling factor was derived for each NAPS ground monitor by taking the weekly mean concentration at that monitor and dividing it by the long-term modelled surface estimate at the monitor location. Inverse distance weighting was then used to assign the scaling factors to all postal codes with 25 kilometres of a NAPS monitor. To get final concentrations, these surfaces were multiplied back to the long-term estimates for each week.Finally, weekly estimates from birth to end of follow-up were averaged to get a single childhood average exposure to each of the selected pollutants.

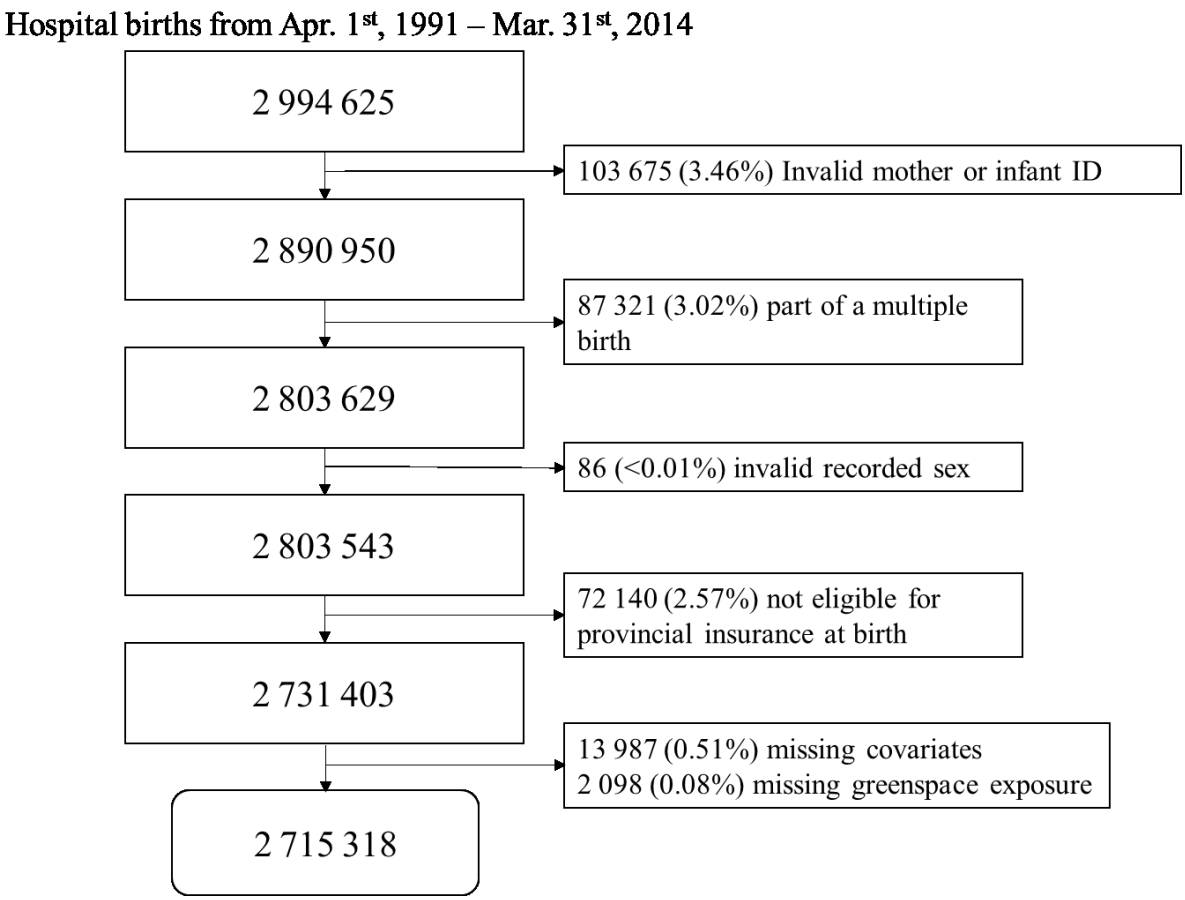
References:

1 Lamsal LN, Martin R V., van Donkelaar A, *et al.* Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. *J Geophys Res Atmos* 2008;**113**:1–15. doi:10.1029/2007JD009235

2 Hystad P, Setton E, Cervantes A, *et al.* Creating national air pollution models for population exposure assessment in Canada. *Environ Health Perspect* 2011;**119**:1123–9. doi:10.1289/ehp.1002976

3 Van Donkelaar A, Martin R V., Li C, *et al.* Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environ Sci Technol* 2019;**53**:2595–611. doi:10.1021/acs.est.8b06392

4 Robichaud A, Ménard R. Multi-year objective analyses of warm season ground-level ozone and PM2.5 over North America using real-time observations and Canadian operational air quality models. *Atmos Chem Phys* 2014;**14**:1769–800. doi:10.5194/acp-14-1769-2014



Supplementary Figure 3. (Greenspace) Study population flowchart.

Supplementary Table 7. Effect measure modification analysis for the association between childhood exposure to residential greenspace and paediatric-onset Crohn's disease

Effect modifier		Number of cases	Measure of association	P value for product interaction term
Rurality	Rural area	147	0.79 (0.61-1.02)	0.84
	Urban area	1758	0.82 (0.76-0.87)	
Maternal IBD	Yes	60	0.80 (0.57-1.13)	0.93
	No	1845	0.81 (0.76-0.87)	
NO_2	High	918	0.89 (0.81-0.97)	0.05
	Low	608	0.79 (0.69-0.89)	
$PM_{2.5}$	High	952	0.83 (0.76-0.91)	0.49
	Low	670	0.81 (0.73-0.90)	
O_3	High	710	0.83 (0.74-0.92)	0.86
	Low	995	0.81 (0.74-0.88)	

All models adjusted for sex, rural/urban residence at birth, maternal IBD status, median income quintile, and two-level random effects

Supplementary Table 8. Effect measure modification analysis for the association between childhood exposure to residential greenspace and paediatric-onset ulcerative colitis

Effect modifier		Number of cases	Measure of association	P value for product interaction term
Rurality	Rural area	99	1.00 (0.71-1.42)	0.0499
	Urban area	1146	0.71 (0.66-0.77)	
Maternal IBD	Yes	35	1.00 (0.62-1.61)	0.18
	No	1210	0.72 (0.67-0.77)	
NO_2	High	638	0.71 (0.64-0.79)	0.05
	Low	351	0.87 (0.73-1.03)	
$PM_{2.5}$	High	614	0.74 (0.66-0.82)	0.65
	Low	414	0.71 (0.63-0.81)	
O_3	High	434	0.73 (0.64-0.83)	0.54
	Low	649	0.75 (0.67-0.83)	

All models adjusted for sex, rural/urban residence at birth, maternal IBD status, median income quintile, and two-level random effects

Supplementary Table 9. Associations of greenspace and IBD after removal of those with estimated neighborhood income

	IBD		Crohn's disease		ulcerative colitis	
	NDVI during pregnancy	NDVI during childhood	NDVI during pregnancy	NDVI during childhood	NDVI during pregnancy	NDVI during childhood
Model 1: Unadjusted	0.95 (0.91-1.00)	0.79 (0.75-0.82)	1.02 (0.96-1.08)	0.84 (0.79-0.89)	0.89 (0.83-0.95)	0.72 (0.68-0.78)
Model 2: Adjusted*	0.96 (0.92-1.00)	0.78 (0.75-0.82)	1.03 (0.97-1.09)	0.83 (0.79-0.89)	0.89 (0.83-0.96)	0.72 (0.67-0.78)
Model 3: Model 2 + random effects	0.97 (0.92-1.02)	0.77 (0.74-0.81)	1.02 (0.96-1.09)	0.81 (0.76-0.86)	0.91 (0.84-0.98)	0.72 (0.67-0.78)

- End of Chapter 4 -

Chapter 5: Overall discussion and conclusions

Inflammatory bowel disease is a complex disorder that is becoming more common, especially among children and youth. Knowledge of IBD etiology is still evolving, but environmental exposures are a known contributor to this disease. As a cure for IBD does not exist, and treatment has varying effectiveness, primary prevention of IBD is a top priority. The overarching goal of this thesis was to address evidence gaps on how environmental exposures may be linked to IBD incidence among youth. This will hopefully lead to new knowledge of disease etiology and most importantly, identify potential targets for interventions to reduce the incidence of IBD among children.

In this thesis, we investigated the independent contributions of air pollution and greenspace to the risk of being diagnosed with IBD before the age of 18 years. Maternal and childhood exposures to air pollution were explored in Chapter 3, while maternal and childhood exposures to greenspace were investigated in Chapter 4.

5.1 Main findings

5.1.1 Objective 1: Air pollution and IBD

The results of the analysis looking at the effects of maternal and childhood air pollution exposures on risk of developing paediatric-onset IBD were presented in Chapter 3. We found a relationship between exposure to O_x during the childhood period and increased risk of childhood-onset IBD. When investigating in-utero exposures, a slight increased risk was detected during the mid-pregnancy period for exposure to O_x although when the whole pregnancy period was considered, there was no longer an

increased association. We did not find any significant associations between exposures to NO_2, $PM_{2.5}$, or O_3 and IBD.

5.1.2 Objective 2: Greenspace and IBD

The association between residential greenspace and paediatric-onset IBD was explored in Chapter 4. In this study, we found a protective association between increased greenspace during the childhood period, and risk of developing IBD <18 years. We also observed a linear concentration-response relationship and found that the relationships were stronger for very early onset cases of IBD <10 years of age. We did not find any association between residential greenspace during pregnancy and any of paediatric-onset IBD, CD, or UC.

5.2 Discussion of findings in context of previous knowledge

5.2.1 Previous literature of air pollution and IBD

No published study, to our knowledge, has evaluated maternal exposures to air pollution and the risk of IBD. Further, no study has looked at air pollution and the risk of paediatric-onset IBD. There have been two published studies conducted on air pollution and risk of adult IBD, one conducted by Kaplan et al in 2010,[71] and another by Opstelten et al in 2016.[72]

The Kaplan et al. study looked at 367 incident cases of CD and 591 incident cases of UC from the United Kingdom that were matched 1:5 with age and sex-matched controls. In this study, they did not find an association between exposure to NO_2, SO_2 and PM_{10} among the general population but when the population was split into

subgroups, they noted two hazardous associations among the youngest subgroups. The first reported association was found in those aged under 23 years at diagnosis, where the highest three quintiles of NO_2 exposure were associated with higher odds of IBD (OR 2.31, 95% CI of 1.25-4.28). There was also a linear relationship across quintiles of NO_2 detected (P=0.02). The second association was observed in people under 25 years of age, where the highest three quintiles of SO_2 was associated with the development of UC (OR 2.00, 95% CI of 1.08-3.72). Null associations were reported for all PM_{10} exposures, and for NO_2 and SO_2 exposures in all other age groups (except for those aged 44-57 where inverse associations with NO_2 and PM_{10} were identified).

The study by Opstelten et al. compared 38 cases of CD and 104 cases of UC with 568 controls of mainly middle-aged and elderly persons originating from several different European countries. They also observed a mix of positive, null and inverse associations. Traffic intensity on major roads within 100m was associated with increased odds of developing IBD (OR 1.60, 95% CI 1.04-2.46). However, PM_{10} and $PM_{2.5}$ were both inversely associated with odds of developing IBD (OR 0.25, 95% CI 0.08-0.78, and OR 0.24, 95% CI 0.07-0.81, respectively). NO_2, NO_x and alternative measures of traffic and particulate matter all displayed null associations.

In Chapter 3, we reported null associations for childhood exposures to NO_2, which is commonly used as a proxy for traffic-related air pollution. This null association for NO_2 was consistent with the null associations reported in the main analysis of the Kaplan et al. and the Opstelten et al. studies. However, this contradicts the association found in the younger cohort of Kaplan et al, and the increased association for nearby traffic

reported in Opstelten et al. The relevance of NO_2 and of the traffic-related air pollutants in IBD risk remains unclear, but based on the results presented in Chapter 3, they may not be as important as previously thought.

In both published studies and in our current results, particulate matter of any fraction was not associated with an increased risk of IBD. It should be mentioned, however, that recent studies on particulate matter toxicity suggest that the source of the particulates (i.e., anthropogenic or from nature), components of particulates and the oxidative potential of the particulates may influence the toxicity more than just the mass concentration.[102] These features of $PM_{2.5}$ were not considered in this, or any previous IBD study due to challenges in measuring these components, but this is something that should be taken into account in future studies, if possible.

Our study was the first to consider O_3 or O_x exposures. Given our findings, future studies on this topic should be sure to include measures of O_3 and O_x to see if our results can be replicated.

5.2.2 Previous literature of greenspace and IBD

There has not been any published research looking at the relationship between greenspace and IBD. However, our findings that greenspace was protective of paediatric IBD may complement the previous literature on the hygiene hypothesis. This hypothesis posits that reduced exposure to microbes and infections experienced by those in Western countries is a likely cause for many of the immune-mediated conditions that have risen in prevalence. Greenspace is a major source of biodiversity,

and the vegetation around one's home can contain many different species of bacteria and other microbes. Our results may also complement previously reported findings that rural living was protective of IBD, as rural areas also tend to contain greater greenspace. It is possible that accounting for greenspace may help explain the heterogeneity between studies found in a meta-analysis of rural/urban living and IBD risk.[65]

5.2.3 Differences in associations across disease subtypes

In both Chapter 3 and Chapter 4, while we did not formally test for differences in the associations across subtypes, we found only subtle changes in the measures of association when comparing the effects of environmental features on the risk of CD vs risk of UC. In stratified analyses in Chapter 3, exposures to NO_2 in the 2nd trimester had a statistically significant negative association with CD but no association with UC. The two statistically significant associations reported from the main analysis of Chapter 3, exposure to O_x in the 2nd trimester (HR_{CD} 1.18, 95% CI 0.95-1.47; HR_{UC} 1.26, 95% CI 0.97-1.64) and exposure to O_x during childhood (HR_{CD} 1.05, 95% CI 0.96-1.16; HR_{UC} 1.10, 95% CI 0.98-1.24) seemed to be marginally more strongly associated with UC than CD.

In Chapter 4, neither UC nor CD were associated with greenness during pregnancy, but both had an inverse association with increasing greenspace during childhood. In general, the protective associations seen for childhood greenness were modestly stronger for UC but they were nearly identical when only early onset cases were considered.

In the broader IBD literature, differences in risk across IBD subtypes are often noted, such as for appendectomy, antibiotics, Vitamin D, etc.[119–121] The most dramatic example is for current smoking status, where it is a risk factor for developing CD but inverse associations are commonly reported for UC.[49] We did not observed such dramatic differences in risk by IBD subtype in either study included in this thesis, and the direction of all the associations were consistent across CD and UC subtypes. This suggests that the potential mechanisms linking air pollution and greenness to IBD are common to both disease subtypes.

5.2.4 Effect modification analyses

In our analyses, effect modification was explored across rurality of residence at birth, and maternal IBD status. In Chapter 4, high and low concentrations of air pollution was also investigated as an additional potential effect modifier.

For rurality of residence at birth, associations for air pollution seemed to differ only slightly across rural and urban areas. First trimester exposure to NO_2 had a greater association with IBD amongst those living in urban spaces when compared to rural areas, as did second trimester exposures to O_3 and O_x. While it is possible that an increase in air pollution is more hazardous to those living in rural areas who tend to have lower levels of air pollution, a more likely explanation may be that those living in rural areas and experiencing higher air pollution may be living closer to urban areas, which is known to have higher IBD rates than rural areas (likely because of other, unrelated risk factors). For greenspace, the magnitude of association in rural areas (HR 0.85, 95% CI 0.69-1.03) was slightly weaker than in urban areas (HR 0.77, 95% CI

0.73-0.81), but the effect modification was not statistically significant (P=0.33). When we assessed the disease subtypes individually, we observed effect measure modification for the UC subtype, where the effect in rural areas was null (HR 1.00, 95% CI 0.71-1.42) but was protective in urban areas (HR 0.71, 95% CI 0.66-0.77; P=0.05). While this may also be a marker of a less urban area, another reason why the protective effect of greenspace might be stronger among those in urban areas is that greenspace might be a more important source of biodiversity in urban areas. It could also be that greenspace encourages physical activity in the urban setting. In rural areas, individuals tend to live lifestyles that lead to greater physical activity and exposure to biodiversity regardless of the level of greenspace, so greenness may not be as important of a factor in rural settings.[122]

Maternal IBD status reflects, in part, a child's genetic risk for IBD. Effect modification by maternal IBD might be suggestive of exposures that interact with heritable traits. A notable finding in the air pollution study was that $PM_{2.5}$ exposure during childhood was associated with IBD risk among those born to mothers with IBD. This was in contrast to the results from the unstratified analysis which showed no association with increase in $PM_{2.5}$ during childhood. This observed modification of effect suggests that perhaps there is some inherited trait that may predispose children to be more vulnerable to particulate matter exposures. It should be mentioned that this finding is based on only 25 cases of IBD among children born to mothers with IBD and should be interpreted cautiously. However, this is not the first time that such an association has been reported; previous research on smoking (which is also a considerable source of

particulate matter exposure)[123] and CD showed that the risk is inflated among those with a family history of IBD.[124] The effects of greenness were also compared across those born to mothers with IBD and those not born to mothers with IBD. The effect of greenspace on risk of IBD or CD was not modified by maternal IBD status, as the effect estimates were very similar (P=0.72, and 0.93 respectively). For UC, the effect seemed to be null in those with mothers that had IBD (HR 1.00, 95% CI 0.62-1.61) and protective in those born to non-affected mothers (HR 0.72, 95% CI 0.67-0.77), though it should be noted that this comparison was based on 35 cases of UC born to mothers with IBD and the effect estimate was found have a P value of 0.18. Together, these findings support the idea that genetic predisposition may be an important factor when evaluating the effects of some environmental exposures, and future research and guidelines should take this into consideration.

When greenness results were stratified across levels of air pollution, there was no significant interaction detected (P range: 0.33-0.93) and all estimates of association were qualitatively the same. When we assessed effect modification by air pollutants for specific disease subtypes, there was a slight difference in the effects of greenness across high and low levels of NO_2 exposures. For CD, the protective effect of greenness was somewhat stronger in areas of low NO_2 (HR 0.79, 95% CI 0.69-0.89) than in areas of high NO_2 (HR 0.89, 95% CI 0.81-0.97). However, for UC this comparison was reversed, and the effects of greenness were stronger in areas of high NO_2 (HR 0.71, 95% CI 0.64-0.79) compared to areas of low NO_2 (HR 0.87, 95% CI 0.73-1.03). This effect measure modification was marginally significant (P=0.05 for both

CD and UC). The results presented in previous studies on air pollution suggest that NO_2 might be more of a risk factor for CD than for UC.[71] This may help explain why the effects of greenspace on reducing risk of CD seemed to be attenuated in areas of high NO_2, especially if the two environmental factors share common pathways.

5.2.5 Differences found in early-onset IBD cases (age <10 years)

As the gut microbiome is believed to be colonized early after birth, environmental exposures may be most relevant in the early-life period. For this purpose, follow-up time in both studies was limited to the first 10 years of life in sensitivity analyses, to explore how the associations changed for very early onset cases of IBD, and if they were to be stronger as hypothesized.

When we assessed pollution exposures and early-onset paediatric IBD, the associations for the pregnancy period were largely unchanged compared with our primary results but had wider confidence intervals. This was likely caused by the loss of statistical power due to the reduction in number of events (IBD is much more rare in children <10 years old). Childhood exposures to O_x did have a larger effect size when compared to the association for the main analysis (HR 1.20, 95% CI 0.91-1.58, compared to HR 1.08, 95% CI 1.01-1.16). This could be because exposures to O_x earlier in childhood confer greater risk, or because exposure to O_x causes earlier clinical presentation of IBD in children.

When only the first ten years of follow-up were considered for greenspace, the protective effect for IBD also appeared to be even stronger. There was also a more

consistent association across the disease subtypes. This supports the hypothesis that exposures may be most important within the first ten years of life. This might be especially true for CD which saw a bigger increase in effect estimate once only the first 10 years of exposure were considered. This was similar to a previous study that looked at the risk of appendectomy on IBD incidence, that found that while appendectomy was associated with an increased risk of CD, an appendectomy before ten years of age was actually associated with a reduced risk.[125] A study looking at effects of rural/urban residence on IBD risk also found that rural residence had the greatest protective effect among those younger than 10 years.[66]

Together, these findings highlight the first decade of life as important for modulation of the environmental risk of IBD, and CD in particular.

5.2.6 Implications for disease etiology

The results from these studies may add to the current understanding of IBD etiology. The gut microbiome and other immunological changes represent popular pathways proposed to connect environmental exposures to the initiation of IBD.

There have been many studies linking dysfunction of the gut microbiome to IBD disease status.[126] Though the body of evidence on how air pollution affects the gut microbiome is still growing, preliminary studies show that exposure to air pollutants can increase the permeability of the gastrointestinal tract, leading to a pro-inflammatory response that results in systemic inflammation.[70] There may also be associated changes in the composition of the microbial community as a result of these changes.[70]

The lack of clear association for air pollution exposures during the in-utero period is also consistent with this theory, as it is known that the gut microbiome is not colonized in-utero.[43] Increased exposure to microbes and increased physical activity, which are correlated with greenspace,[114] may also be linked to the gut microbiome.[61,63,127,128] While the strong associations found in Chapter 4 provide compelling evidence for continued research into these two factors, it should be noted that NDVI provides only a crude estimate of the greenspace that one is exposed to, and has many drawbacks (discussed below) that limit its ability to point to specific mechanisms of disease.

Other popular theories on IBD etiology revolve around the immune system. For instance, there are many important developmental processes occurring during mid-pregnancy, such as development of the gut associated lymphoid tissue which can begin as early as 11 weeks of gestation, which develop into a mature ordered structure by week 18.[129] This coincides with the time period in which the increased association was detected for exposures to O_x in the study in Chapter 3. Cytokines may play an important role in IBD by determining the T cell differentiation of T-helper 1 (Th1), T-helper 2 (Th2), and regulatory T cells).[130] Th17 is also suspected to play an important role, particularly for Crohn's disease.[131,132] Exposure to air pollution may affect immunoregulation; a previous study showed that exposure to combustion based air pollutants raised the levels of tumor necrosis factor (TNFα), which may be regulated by the Th1 and Th2.[133,134] Greenspace may also be implicated in this immunoregulatory pathway through the hygiene hypothesis.[135] Exposure to greenspace may lead to

exposure to microbes that are essential in regulating inappropriate responses, which may help explain the protective effect seen in Chapter 4.

The fact that in both Chapter 3 and Chapter 4 the observed associations were stronger for very early onset IBD cases (<10 years) also provides further support for the theory that developmental processes in the early life period are susceptible to environmental factors.

It should be reiterated that research in the relationship between both air pollution and especially greenspace is very novel, and so the above-mentioned relationships are speculative. Future epidemiological and experimental research on the role of these environmental factors in IBD etiology is needed to support these mechanisms as theories for how IBD develops.

5.2.7 Generalizability

This population-based study was conducted in Ontario, the most populous, and diverse province in Canada. The results of this thesis may be fairly generalizable to other heterogenous areas in Canada. The results may also apply to other high-income nations, such as the US, western Europe and Australia, which are broadly similar in their environmental exposures and levels of urbanization, lifestyles, etc. Special attention should be paid to comparing those from different ethnic backgrounds, however, as people from different countries have a different background risk of IBD, which may be due to genetic factors, or how they respond to environmental stimuli.[136] In addition, patients of similar ethnic backgrounds tend to cluster in shared

neighbourhoods, particularly in cities. Therefore, ethnicity and genetics may have acted as confounding factors in our assessment of the association between greenness, air pollution and IBD incidence.

Exposures to air pollutants in Chapter 3 were relatively low, and relationships with IBD may be different at higher concentrations. The NDVI measure of greenness is a normalized measure, so an increase of 0.1 in NDVI may represent a different change in absolute levels of greenness in areas with a greater or lower distribution of greenness values. Caution should thus be taken when extrapolating the results of the studies enclosed in this thesis with other nations that do not share similar exposure profiles. Further, the objective of this thesis was to assess the relationship between environmental exposures and childhood-onset IBD at a population level, and the studies and conclusions presented are not intended to provide statements about individual risk.

5.3 Strengths

5.3.1 Available information on lifetime exposures

A strength of this thesis was our ability to assess environmental exposures since conception, something that would be extremely difficult to do with a different study design or in an adult population. Children have accumulated fewer exposures over their lifetimes than adults, so it is easier to characterize their complete environmental exposure history. This allowed us to more accurately assess their lifetime exposures and investigate associations with IBD status.

5.3.2 Statistical analyses

The use of a time-varying exposure to represent the childhood period in Chapter 3 clarified questions raised by previous air pollution and IBD studies, which either used a simple average or categorized exposures into quintiles. In using a time-varying estimate, representing the yearly cumulative exposure to air pollutants, we accounted for how changes in exposure over time may affect risk. For instance, when using a simple averaged exposure to represent pollutants, there is a forced assumption that an increase in exposure in early childhood is equivalent to an increase in exposure in late childhood, which is not likely to be true. Using a simple average to assign childhood exposures does not account for trends in air pollutants decreasing in concentration over time in a country like Canada. By not taking this trend into account, the effects of air pollution may be overestimated as those with longer follow-up in this study (who are also more likely to develop IBD) are more likely to have overestimated air pollution concentrations for their later years inflated when using a simple average.

In Chapter 3, the regression models were mutually adjusted for other time periods. This means that exposures to pollutants during pregnancy and during childhood were added in the same model. Since pollutant concentrations during pregnancy and childhood tend to correlate (especially if there is no residential mobility), it can sometimes be difficult to detect which exposure period is associated with the outcome. By mutually adjusting the regression models, we can more definitively point to one time period as more relevant to the investigated outcome. In the case of ambient air pollution and IBD, we concluded that exposures during childhood may be more

important than exposures during pregnancy. While collinearity is a concern when adding two covarying predictors in the model, initial diagnostics showed that the variance inflation factors (VIF) for the coefficients in the mutually adjusted models ranged up to 7.5 for NO_2, 3.2 for $PM_{2.5}$, 2.3 for O_3, and 2.3 for O_x, well below the accepted guideline of VIF ≥10 being suggestive of major multi-collinearity.[137]

A challenge of investigating environmental exposures such as air pollution and greenspace is that these exposures are correlated with a variety of other exposures because of the dependence on the physical geography of residences. In observational studies such as the ones presented in Chapters 3 and 4, exposures to air pollution and greenspace are not distributed randomly across the population. In fact, there are important trends to consider in the type of people that tend to live in areas of varying greenspace and air pollution; among those living in urban areas, people living in areas with high greenness and low pollution tend to be wealthier, are less likely to be immigrants / ethnic minorities, and have better health literacy and behaviours.[138,139] To account for this potential confounding by unmeasured spatially-correlated characteristics, mixed effects model with two levels of random effects were used in Chapter 4. While also relevant to the analyses presented in Chapter 3, they were initially employed there as well, but because there were minimal changes in the effect estimates, they were not kept in due to computational reasons. Such two-level models have been previously found to account for confounding by spatial contextual variables in air pollution studies,[140] and models very similar to ours have been increasingly employed in the literature.[141,142]

The distributed lag non-linear models (DLNM) are a statistical method borrowed from time series analysis that facilitate the investigation of exposures that are in a close temporal sequence. This technique was used in Chapter 3 to look simultaneously look at the effects of exposures during the different weeks of gestation. Using DLNM to investigate prenatal exposures is preferable to the conventional way of splitting exposures into trimesters, as a child's development does not occur along three arbitrarily defined phases. By considering all weeks of development, critical windows of exposure may be more clearly identified. This is important if a process that occurs between the trimesters is relevant to IBD, as it may be missed if a trimester-averaging approach is used. In our case, we were able to identify a potential association that started before the second trimester and continued afterwards into the third trimester.

5.3.3 Large population-based study

The use of large population-based cohort was important to this thesis. Though IBD is becoming more frequent, its incidence among those under 18 years of age is still uncommon. In fact, in our source cohort of 2 731 403 children, we only captured a total of 3 474 (0.13%) cases of IBD diagnosed before the age of 18 years during their follow-up. It certainly would not have been feasible to run a prospective cohort study of this magnitude. By capturing this many cases, we were able to run sophisticated analyses exploring subgroups while ensuring sufficient statistical power to detect important associations.

5.4 Limitations

5.4.1 Exposure misclassification

There are several sources of exposure misclassification that should be noted. Errors in how the exposures were estimated, the consideration of residential exposures only, and residential mobility may all have had some impact on the observed results.

The exposure assessment methods were based on NAPS monitors, which tend to be located in urban areas. This explains why more rural residents were excluded based on missing air pollution exposures than would otherwise be expected in Chapter 3. This may also have caused the rural estimates of pollution to be overestimated,[143] which would potentially lead us to underestimate the hazardous effects of air pollution in the entire population given that those in rural areas are known to be at a lower baseline risk of IBD. This bias may have explained why there were some differences in the air pollution associations after stratification by rural/urban residence at birth. The positional uncertainty of postal code assignment method is also likely to be greater in rural residence where six-digit postal codes have a greater area to cover, further contributing to exposure misclassification but in a more unpredictable manner.

Another source of exposure misclassification is the fact that only residential exposures to air pollution and greenspace were available. This means that we did not consider exposures at places where each individual worked, went to school, etc. This is mitigated by the fact that children tend to go to school and daycare close to their home so there are only small differences in exposure expected. This source of

misclassification is also unlikely to differ greatly by IBD status, and would bias the observed associations towards the null.

Residential mobility may have had contributed to exposure misclassification as well. For the assessment of maternal exposures, the six-digit postal code at birth was used. The rates of residential mobility during pregnancy in a similar population was estimated to be around 12%.[144] Further, pregnant women tend to move short distances resulting in small differences in exposures to air pollution.[145] Combined with the fact that previous unrecorded residential moves are likely to have been updated during one of several health care contacts for prenatal care that is typical of Canadian residents,[146] there is strong motivation for the use of six-digit postal code at delivery to represent maternal exposures.

There is no doubt greater residential mobility occurring during the potential 18 years of follow-up for each child. This is why the annual six-digit postal code (as of July 1st) was obtained through linkage with the RPDB and was used to assign annual exposures. In both cases, potential misclassification due to incorrect residence would likely be non-differential and would have biased the measures of association towards the null. This suggests that the true associations reported in this thesis might be stronger than we estimated.

There are several important limitations with the NDVI measure that should be considered in the interpretation of the results of Chapter 4. NDVI is a crude measure of how “green” the surface of Earth is at a given point of time. Though a child may be

living close to an area that is high in greenspace, it does not say anything about their access and interactions with that greenspace. Better indicators of greenspace looking into the types of vegetation, and assessing the level of interaction of greenspace should be used in order to further assess this relationship.

5.4.2 Use of routinely-collected data

An important limitation to acknowledge is that information on demographics and outcomes in this thesis are based on routinely-collected data that were not designed to be used for research purposes. Several inherent consequences of the use of such data include variation in coding practices across facilities and over time, and data quality issues such as missingness which was addressed somewhat, and miscoded data, which is more difficult to detect and address.

5.4.3 Missing data

The strategies used to deal with missing data were described in Section 2.3. Missingness of outcomes and of covariates were not a great concern to the validity of the results presented within this thesis. The missingness of covariates was small (less than 5% of the total cohort) and where possible, such as for neighborhood income levels, a reasonable substitute was used by taking the census subdivision area level income when dissemination area level income was missing. A sensitivity analysis excluding those with a substitution for income in Chapter 4 was conducted and no noticeable differences in the associations were noted (Supplementary Table 9).

Due, in part, to using highly resolved exposure assignment methods over a long follow-up period, there was a notable proportion of the population that had missing exposure values for the Chapter 3. There were 498,627, or about 18.3% of the source cohort that did not have a single valid exposure estimate for either the pregnancy or childhood periods for any of the pollutants, and were thus excluded. Further, there was a small proportion of the population that had to have at least 1 week or 1 year of exposure estimates extrapolated from other weeks or years. The proportion of children that had at least one exposure estimate extrapolated (ie. at least one week during pregnancy or one year of follow-up) was 25.4% for NO_2, 21.5% for $PM_{2.5}$, 19.0% for O_3, and 32.2% for O_x. In Chapter 4, there was considerably less missingness for the greenness exposures, so the 0.54% of the source cohort that were missing NDVI values were simply excluded. We chose a somewhat aggressive imputation technique to try and retain as many observations as possible, to keep sufficient statistical power. A drawback of such a method is the fact that the extrapolation method used likely did not capture the natural variability of exposures over time (especially for the air pollution analyses which had a more temporal component), which would have underestimated the standard errors of the associations resulting in narrower confidence intervals than would otherwise be expected. A sensitivity analysis excluding any extrapolated exposures was conducted to explore how the exposure estimation may have affected the results in Chapter 3 (Supplementary Table 6). The associations all had the same direction as for before, but the magnitudes were even greater and confidence intervals were much wider, as hypothesized.

5.4.4 Confounding

There are several risk factors that have the potential to confound the relationships between the environmental exposures of interest in this thesis and paediatric-onset IBD. This is best illustrated in the example DAG (Figure 2). Due to the observational study design used in Chapters 3 and 4, concerns about confounding are warranted. Several of the more important confounders such as rurality of residence, area-level income, maternal IBD status, and sex were available from the health administrative data, so we were able to control for them by adding them as independent variables in the regression models. Despite the fact that the measures of association were not significantly affected by the set of confounders employed (less than a 10% change in estimate after removal of any covariate) in Chapter 3 and Chapter 4, these covariates were retained in the models due to previous knowledge of their importance with IBD. A simulation study showed that performing covariate selection using change-in-estimate procedures was only effective when the set of covariates included potential nonconfounders.[147] In this thesis, minimally adjusted models containing only known confounders were used. Since we used administrative health data we were limited to only the variables that were available in the administrative data, and were unable to collect information on other potential confounders such as father's IBD status, dietary factors, and individual-level ethnicity.[91,148] We also had to use area-level income as opposed to individual-level income for the adjustment of socioeconomic status. However, area-level socioeconomic status indicators have been found to be valid proxies for individual-level socioeconomic status indicators in studies in the United States and Canada.[149,150] Area-level socioeconomic status indicators were also found to

be associated with disease outcomes, and are a meaningful indicator for population health outcomes.[151]

We could have linked to the Better Outcomes Registry & Network (BORN) dataset, a database comprising all births in Ontario from years 2006-2012 that has additional clinical and confounder information to try and capture other perinatal variables that may be related to gut microbiome and/or IBD risk such as intent to breastfeed and smoking during pregnancy. However, due to the limited availability of years (and their recency), any analyses with these covariates would not contain enough cases of paediatric IBD to be sufficiently powered to detect possible associations.

5.4.5 Changes in exposure assessment and outcome ascertainment over time

The study period of the analyses presented in this thesis spanned 26 years. In this time, there were undoubtedly some changes in the way that both the exposure was assessed, and in the way that the outcomes were ascertained. For air pollution exposures, NAPS monitor data were used for the temporal interpolation. Since 1991, additional NAPS monitors have been introduced. By having additional monitors in more recent years, we have less exposure misclassification (particularly among rural areas where the NAPS monitor networks are not particularly dense to begin with). In terms of NDVI values, changes in from Landsat 5 to Landsat 8 in 2012 may have had only a minimal impact on results as long-term averages were used to quantify exposures. In terms of outcome ascertainment, though there was an important shift in coding standards with the introduction of ICD-10 in the year 2002, the algorithm used to identify cases of IBD was validated in both the ICD-9 and ICD-10 time periods.[23]

Despite this, there may be changes in diagnostic patterns and greater awareness of IBD, especially among children, that would have caused cases to be diagnosed earlier in more recent years.

5.5 Conclusion

In this thesis, we demonstrated that environmental factors play an important role in the development of paediatric IBD. We were able to replicate some of the results of previous studies as well as investigate new unexplored relationships with risk of IBD. We showed that childhood exposures to certain measures of air pollution may confer an increased risk, and that important exposure windows may begin even sooner than previously hypothesized. The protective effects for greenspace had an unexpectedly strong association and may provide an avenue for reducing the risk of IBD among urban youth in particular, assuming a causal relationship. Our findings have already inspired future research, as a grant has been submitted to further explore and replicate the noted associations from this thesis, but in a national cohort. We have also shown that several environmental factors may act together, supporting future studies investigating the suite of exposures known as the “exposome” in relation to IBD risk.

As global trends in urbanization and the environment continue, limiting harmful environmental exposures while improving access to beneficial elements is an important priority. The studies included in this thesis provide further evidence that may be used by clinicians and policy makers to help prevent future lifelong cases of IBD not only among those predisposed to it, but to support better health for all.

5.6 References

1 Podolsky DK. Inflammatory bowel disease. *N Engl J Med* 1991; **325**: 928–37.

2 Knights D, Lassen KG, Xavier RJ. Advances in inflammatory bowel disease pathogenesis: Linking host genetics and the microbiome. *Gut* 2013; **62**: 1505–10.

3 Bernstein CN, Blanchard JF, Rawsthorne P, Yu N. The prevalence of extraintestinal diseases in inflammatory bowel disease: A population-based study. *Am J Gastroenterol* 2001; **96**: 1116–22.

4 Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of IBD. *Nature* 2011; **474**: 307–17.

5 Frolkis AD, Dykeman J, Negrón ME, *et al.* Risk of surgery for inflammatory bowel diseases has decreased over time: A systematic review and meta-analysis of population-based studies. *Gastroenterology* 2013; **145**: 996–1006.

6 Bernklev T, Jahnsen J, Lygren I, Henriksen M, Vatn M, Moum B. Health-related quality of life in patients with inflammatory bowel disease measured with the short form-36: Psychometric assessments and a comparison with general population norms. *Inflamm Bowel Dis* 2005; **11**: 909–18.

7 Pizzi LT, Weston CM, Goldfarb NI, *et al.* Impact of Chronic Conditions on Quality of Life in Patients With Inflammatory Bowel Disease. *Inflamm Bowel Dis* 2006; **12**: 46–52.

8 Lix LM, Graff LA, Walker JR, *et al.* Longitudinal study of quality of life and psychological functioning for active, fluctuating, and inactive disease patterns in inflammatory bowel disease. *Inflamm Bowel Dis* 2008; **14**: 1575–84.

9 Sajadinejad MS, Asgari K, Molavi H, Kalantari M, Adibi P. Psychological issues in inflammatory bowel disease: An overview. *Gastroenterol Res Pract* 2012; **2012**. DOI:10.1155/2012/106502.

10 Neuendorf R, Harding A, Stello N, Hanes D, Wahbeh H. Depression and anxiety in patients with Inflammatory Bowel Disease: A systematic review. *J Psychosom Res* 2016; **87**: 70–80.

11 Mackner LM, Sisson DP, Crandall W V. Review: Psychosocial issues in pediatric inflammatory bowel disease. *J Pediatr Psychol* 2004; **29**: 243–57.

12 Imielinski M, Baldassano RN, Griffiths A, *et al.* Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet* 2009; **41**: 1335–40.

13 Van Limbergen J, Russell RK, Drummond HE, *et al.* Definition of Phenotypic Characteristics of Childhood-Onset Inflammatory Bowel Disease. *Gastroenterology* 2008. DOI:10.1053/j.gastro.2008.06.081.

14 Motil KJ, Grand RJ, Davis-Kraft L, Ferlic LL, Smith EO. Growth failure in children with inflammatory bowel disease: A prospective study. *Gastroenterology* 1993; **105**: 681–91.

15 Kanof ME, Lake AM, Bayless TM. Decreased height velocity in children and adolescents before the diagnosis of Crohn's disease. *Gastroenterology* 1988; **95**: 1523–1527.

16 Mackner LM, Bickmeier RM, Crandall W V. Academic achievement, attendance, and school-related quality of life in pediatric inflammatory bowel disease. *J Dev Behav Pediatr* 2012; **33**: 106–11.
17 Ng SC, Yun Shi H, Hamidi N, *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *www.thelancet.com* 2017; **390**. DOI:10.1016/S0140-6736(17)32448-0.
18 Park SJ, Kim WH, Cheon JH. Clinical characteristics and treatment of inflammatory bowel disease: A comparison of Eastern and Western perspectives. *World J Gastroenterol* 2014; **20**: 11525–37.
19 Victoria CR, Sassaki LY, Nunes HRDC. Incidence and prevalence rates of inflammatory bowel diseases, in midwestern of São Paulo State, Brazil. *Arq Gastroenterol* 2009; **46**: 20–5.
20 Molodecky NA, Soon IS, Rabi DM, *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 2012; **142**: 46-54.e42.
21 Kaplan G, Benchimol E, Bernstein C, *et al.* Impact of Inflammatory Bowel Disease in Canada. 2018 https://www.crohnsandcolitis.ca/Crohns_and_Colitis/documents/reports/2018-Impact-Report-LR.pdf.
22 Benchimol EI, Bernstein CN, Bitton A, *et al.* Trends in epidemiology of pediatric inflammatory bowel disease in Canada: Distributed network analysis of multiple population-based provincial health administrative databases. *Am J Gastroenterol* 2017; **112**: 1120–34.
23 Benchimol EI, Guttmann A, Griffiths AM, *et al.* Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. *Gut* 2009; **58**: 1490–7.
24 Bernstein CN, Longobardi T, Finlayson G, Blanchard JF. Direct medical cost of managing IBD patients: A Canadian population-based study. *Inflamm Bowel Dis* 2012; **18**: 1498–508.
25 Coward S, Heitman SJ, Clement F, *et al.* Ulcerative colitis-associated hospitalization costs : A population-based study. 2015; **29**: 357–63.
26 Dan A, Boutros M, Nedjar H, *et al.* Cost of Ulcerative Colitis in Quebec, Canada: A Retrospective Cohort Study. *Inflamm Bowel Dis* 2017; **23**: 1262–71.
27 Kawalec P, Malinowski KP. Indirect health costs in ulcerative colitis and Crohn's disease: A systematic review and meta-analysis. *Expert Rev Pharmacoeconomics Outcomes Res* 2015; **15**: 253–66.
28 Büsch K, da Silva SA, Holton M, Rabacow FM, Khalili H, Ludvigsson JF. Sick leave and disability pension in inflammatory bowel disease: A systematic review. *J Crohn's Colitis* 2014; **8**: 1362–77.
29 Loftus E V., Skup M, Ozbay AB, *et al.* The impact of moderate-to-severe Crohn's disease on employees' salary growth. *Inflamm Bowel Dis* 2014; **20**: 1734–8.
30 Adler J, Raju S, Beveridge AS, Wang S, Zhu J, Zimmermann EM. College adjustment in University of Michigan students with Crohn's and colitis. *Inflamm*

Bowel Dis 2008; **14**: 1281–6.
31 Kahn SA, Lin CW, Ozbay B, Wang A, Chao J, Skup M. Indirect Costs and Family Burden of Pediatric Crohn's Disease in the United States. *Inflamm Bowel Dis* 2017; **23**: 2089–96.
32 Hugot J, Chamaillard M, Zouali H, Lesage S, Ce J, Macry J. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**: 599–603.
33 Hampe J, Cuthbert A, Croucher PJP, *et al.* Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001; **357**: 1925–8.
34 Jostins L, Ripke S, Weersma RK, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012; **491**: 119–24.
35 Franke A, Mcgovern DPB, Barrett JC, *et al.* Meta-Analysis Increases to 71 the Tally of Confirmed Crohn's Disease Susceptibility Loci. *Nat Genet* 2010; **42**: 1118–25.
36 Anderson CA, Boucher G, Lees CW, *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 2011; **43**: 246–52.
37 Nishida A, Inoue R, Inatomi O, Bamba S, Naito Y, Andoh A. Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clin J Gastroenterol* 2018; **11**: 1–10.
38 Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci* 2007; **104**: 13780–5.
39 Michail S, Durbin M, Turner D, *et al.* Alterations in the gut microbiome of children with severe ulcerative colitis. *Inflamm Bowel Dis* 2012; **18**: 1799–808.
40 Ott SJ, Musfeldt M, Wenderoth DF, *et al.* Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 2004; **53**: 685–93.
41 Manichanh C, Rigottier-Gois L, Bonnaud E, *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006; **55**: 205–11.
42 Manichanh C, Borruel N, Casellas F, Guarner F. The gut microbiota in IBD. *Nat Rev Gastroenterol Hepatol* 2012; **9**: 599.
43 Perez-Muñoz ME, Arrieta MC, Ramer-Tait AE, Walter J. A critical assessment of the 'sterile womb' and 'in utero colonization' hypotheses: Implications for research on the pioneer infant microbiome. *Microbiome* 2017; **5**: 1–19.
44 Spehlmann ME, Begun AZ, Burghardt J, Lepage P, Raedler A, Schreiber S. Epidemiology of inflammatory bowel disease in a German twin cohort: Results of a nationwide study. *Inflamm Bowel Dis* 2008; **14**: 968–76.
45 Aujnarain A, Mack DR, Benchimol EI. The Role of the Environment in the Development of Pediatric Inflammatory Bowel Disease. *Curr Gastroenterol Rep* 2013; **15**. DOI:10.1007/s11894-013-0326-4.

46 Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol* 2015; **12**: 205–17.
47 Burns DM, Lee L, Shen LZ, *et al.* Cigarette smoking behaviour in the United States. In: Changes in cigarette-related disease risks and their implication for prevention and control. Smoking and Tobacco Control Monograph 8. 1997: 13–42.
48 Allais L, Kerckhof FM, Verschuere S, *et al.* Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut. *Environ Microbiol* 2016; **18**: 1352–63.
49 Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: A meta-analysis. *Mayo Clin Proc* 2006; **81**: 1462–71.
50 Mahid SS, Minor KS, Stromberg AJ, Galandiuk S. Active and passive smoking in childhood is related to the development of inflammatory bowel disease. *Inflamm Bowel Dis* 2007; **13**: 431–8.
51 Jones DT, Osterman MT, Bewtra M, Lewis JD. Passive smoking and inflammatory bowel disease: A meta-analysis. *Am J Gastroenterol* 2008; **103**: 2382–93.
52 Rook GAW. Hygiene hypothesis and autoimmune diseases. *Clin Rev Allergy Immunol* 2012; **42**: 5–15.
53 Baron S, Turck D, Leplat C, *et al.* Environmental risk factors in paediatric inflammatory bowel diseases: A population based case control study. *Gut* 2005; **54**: 357–63.
54 Jakobsen C, Paerregaard A, Munkholm P, Wewer V. Environmental factors and risk of developing paediatric inflammatory bowel disease - A population based study 2007-2009. *J Crohn's Colitis* 2013; **7**: 79–88.
55 Amre DK, Lambrette P, Law L, *et al.* Investigating the hygiene hypothesis as a risk factor in pediatric onset Crohn's disease: A case-control study. *Am J Gastroenterol* 2006; **101**: 1005–11.
56 Bernstein CN, Rawsthorne P, Cheang M, Blanchard JF. A population-based case control study of potential risk factors for IBD. *Am J Gastroenterol* 2006; **101**: 993–1002.
57 Malmqvist E, Elding Larsson H, Jönsson I, *et al.* Maternal exposure to air pollution and type 1 diabetes - Accounting for genetic factors. *Environ Res* 2015; **140**: 268–74.
58 Cohen B. Urban growth in developing countries: A review of current trends and a caution regarding existing forecasts. *World Dev* 2004; **32**: 23–51.
59 UNITED NATIONS. World Urbanization Prospects 2018. 2018 https://population.un.org/wup/.
60 Zuo T, Kamm MA, Colombel JF, Ng SC. Urbanization and the gut microbiota in health and inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* 2018; **15**: 440–52.
61 Schnorr SL, Candela M, Rampelli S, *et al.* Gut microbiome of the Hadza hunter-gatherers. *Nat Commun* 2014; **5**. DOI:10.1038/ncomms4654.
62 De Filippo C, Cavalieri D, Di Paola M, *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural

Africa. *Proc Natl Acad Sci U S A* 2010; **107**: 14691–6.
63 Martínez I, Stegen JC, Maldonado-Gómez MX, *et al.* The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Rep* 2015; **11**: 527–38.
64 Yatsunenko T, Rey FE, Manary MJ, *et al.* Human gut microbiome viewed across age and geography. *Nature* 2012; **486**: 222–7.
65 Soon IS, Molodecky N, Rabi D, Ghali W, Kaplan G. The Relationship Between Urban Environment and Inflammatory Bowel Disease: A Systematic Review and Meta-analysis. *Am J Gastroenterol* 2010; **105**: S465.
66 Benchimol EI, Kaplan GG, Otley AR, *et al.* Rural and Urban Residence during Early Life is Associated with a Lower Risk of Inflammatory Bowel Disease: A Population-Based Inception and Birth Cohort Study. *Am J Gastroenterol* 2017; **112**: 1412–22.
67 Radon K, Windstetter D, Poluda AL, *et al.* Contact with farm animals in early life and juvenile inflammatory bowel disease: A case-control study. *Pediatrics* 2007; **120**: 354–61.
68 Van Donkelaar A, Martin R V., Brauer M, Boys BL. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ Health Perspect* 2015; **123**: 135–43.
69 Kish L, Hotte N, Kaplan GG, *et al.* Environmental Particulate Matter Induces Murine Intestinal Inflammatory Responses and Alters the Gut Microbiome. *PLoS One* 2013; **8**. DOI:10.1371/journal.pone.0062220.
70 Salim SY, Kaplan GG, Madsen KL. Air pollution effects on the gut microbiota: A link between exposure and inflammatory disease. *Gut Microbes* 2013; **5**. DOI:10.4161/gmic.27251.
71 Kaplan GG, Hubbard J, Korzenik J, *et al.* The inflammatory bowel diseases and ambient air pollution: a novel association. *Am J Gastroenterol* 2010; **105**: 2412.
72 Opstelten JL, Beelen RMJ, Leenders M, *et al.* Exposure to Ambient Air Pollution and the Risk of Inflammatory Bowel Disease: A European Nested Case–Control Study. *Dig Dis Sci* 2016; **61**: 2963–71.
73 Hathout EH, Beeson WL, Ischander M, Rao R, Mace JW. Air pollution and type 1 diabetes in children. *Pediatr Diabetes* 2006; : 81–7.
74 Morales E, Garcia-Esteban R, Asensio de la Cruz O, *et al.* Intrauterine and early postnatal exposure to outdoor air pollution and lung function at preschool age. *Thorax* 2015; **70**: 64–73.
75 Leon Hsu H-H, Mathilda Chiu Y-H, Coull BA, *et al.* Prenatal Particulate Air Pollution and Asthma Onset in Urban Children. Identifying Sensitive Windows and Sex Differences. *Am J Respir Crit Care Med* 2015; **192**: 1052–9.
76 Lavigne É, Bélair M-A, Rodriguez Duque D, *et al.* Effect modification of perinatal exposure to air pollution and childhood asthma incidence. *Eur Respir J* 2018; : 1701884.
77 Ritz SA. Air pollution as a potential contributor to the 'epidemic' of autoimmune disease. *Med Hypotheses* 2010; **74**: 110–7.
78 Aronson MFJ, La Sorte FA, Nilon CH, *et al.* A global analysis of the impacts of

urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proc R Soc B Biol Sci* 2014; **281**. DOI:10.1098/rspb.2013.3330.
79 Flies EJ, Skelly C, Negi SS, *et al.* Biodiverse green spaces: a prescription for global urban health. *Front Ecol Environ* 2017; **15**: 510–6.
80 Clarke SF, Murphy EF, O'Sullivan O, *et al.* Exercise and associated dietary extremes impact on gut microbial diversity. *Gut* 2014; **63**: 1913–20.
81 Almanza E, Jerrett M, Dunton G, Seto E, Ann Pentz M. A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. *Heal Place* 2012; **18**: 46–54.
82 Ruokolainen L, Von Hertzen L, Fyhrquist N, *et al.* Green areas around homes reduce atopic sensitization in children. *Allergy Eur J Allergy Clin Immunol* 2015; **70**: 195–202.
83 Lovasi GS, Quinn JW, Neckerman KM, Perzanowski MS, Rundle A. Children living in areas with more street trees have lower prevalence of asthma. *J Epidemiol Community Health* 2008; **62**: 647–9.
84 Sbihi H, Tamburic L, Koehoorn M, Brauer M. Greenness and Incident Childhood Asthma: A 10-Year Follow-up in a Population-based Birth Cohort To. *Am J Respir Crit Care Med* 2015; **192**: 1131–3.
85 Cordain L, Eaton SB, Sebastian A, *et al.* Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* 2005; **81**: 341–54.
86 Popkin BM. Urbanization, lifestyle changes and the nutrition transition. *World Dev* 1999; **27**: 1905–16.
87 Hansen TS, Jess T, Vind I, *et al.* Environmental factors in inflammatory bowel disease: A case-control study based on a Danish inception cohort. *J Crohn's Colitis* 2011; **5**: 577–84.
88 Sakamoto N, Kono S, Wakai K, *et al.* Dietary risk factors for inflammatory bowel disease: A multicenter case-control study in Japan. *Inflamm Bowel Dis* 2005; **11**: 154–63.
89 Hart A, Tjonneland A, Olsen A, *et al.* Linoleic acid, a dietary n-6 polyunsaturated fatty acid, and the aetiology of ulcerative colitis: A nested case-control study within a European prospective cohort study. *Gut* 2009; **58**: 1606–11.
90 Hou JK, Abraham B, El-Serag H. Dietary intake and risk of developing inflammatory bowel disease: A systematic review of the literature. *Am J Gastroenterol* 2011; **106**: 563–73.
91 Andersen V, Olsen A, Carbonnel F, Tjønneland A, Vogel U. Diet and risk of inflammatory bowel disease. *Dig Liver Dis* 2012; **44**: 185–94.
92 Pope CA, Dockery DW. Health effects of fine particulate air pollution: Lines that connect. *J Air Waste Manag Assoc* 2006; **56**: 709–42.
93 Pope CA, Burnett RT, Thurston GD, *et al.* Cardiovascular Mortality and Long-Term Exposure to Particulate Air Pollution: Epidemiological Evidence of General Pathophysiological Pathways of Disease. *Circulation* 2004; **109**: 71–7.
94 Hoek G, Krishnan RM, Beelen R, *et al.* Long-term air pollution exposure and cardio-respiratory mortality: A review. *Environ Heal A Glob Access Sci Source* 2013; **12**. DOI:10.1186/1476-069X-12-43.

95 Brook RD, Brook JR, Urch B, Vincent R, Rajagopalan S, Silverman F. Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation* 2002; **105**: 1534–6.
96 Vingarzan R, Taylor B. Trend analysis of ground level ozone in the greater Vancouver/Fraser Valley area of British Columbia. *Atmos Environ* 2003; **37**: 2159–71.
97 Zhang Q, Jiang X, Tong D, *et al.* Transboundary health impacts of transported global air pollution and international trade. *Nature* 2017; **543**: 705–9.
98 Health Canada. Health impacts of air pollution in Canada : estimates of morbidity and premature mortality outcomes, 2019 report. 2019.
99 Meng J, Martin R V., Li C, *et al.* Source Contributions to Ambient Fine Particulate Matter for Canada. *Environ Sci Technol* 2019; **53**: 10269–78.
100 Hand JL, Schichtel BA, Pitchford M, Malm WC, Frank NH. Seasonal composition of remote and urban fine particulate matter in the United States. *J Geophys Res Atmos* 2012; **117**: 1–22.
101 Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM. Spatial and temporal variation in PM2.5 chemical composition in the United States for health effects studies. *Environ Health Perspect* 2007; **115**: 989–95.
102 Ayres JG, Borm P, Cassee FR, *et al.* Evaluating the toxicity of airborne particulate matter and nanoparticles by measuring oxidative stress potential - A workshop report and consensus statement. *Inhal Toxicol* 2008; **20**: 75–99.
103 Weichenthal SA, Lavigne E, Evans GJ, Godri Pollitt KJ, Burnett RT. Fine particulate matter and emergency room visits for respiratory illness: Effect modification by oxidative potential. *Am J Respir Crit Care Med* 2016; **194**: 577–86.
104 Van Donkelaar A, Martin R V., Li C, Burnett RT. Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environ Sci Technol* 2019; **53**: 2595–611.
105 Robichaud A, Ménard R. Multi-year objective analyses of warm season ground-level ozone and PM2.5 over North America using real-time observations and Canadian operational air quality models. *Atmos Chem Phys* 2014; **14**: 1769–800.
106 Williams ML, Atkinson RW, Anderson HR, Kelly FJ. Associations between daily mortality in London and combined oxidant capacity, ozone and nitrogen dioxide. *Air Qual Atmos Heal* 2014; **7**: 407–14.
107 Weichenthal S, Lavigne É, Evans G, Pollitt K, Burnett RT. Ambient $PM_{2.5}$ and risk of emergency room visits for myocardial infarction: Impact of regional $PM_{2.5}$ oxidative potential: A case-crossover study. *Environ Heal* 2016; **15**: 46.
108 James P, Banay RF, Hart JE, Laden F. A Review of the Health Benefits of Greenness. *Curr Epidemiol Reports* 2015; **2**: 218–218.
109 Dadvand P, Villanueva CM, Font-Ribera L, *et al.* Risks and Benefits of Green Spaces for Children : *Environ Health Perspect* 2014; **122**: 1329–36.
110 De Vries S, van Dillen SME, Groenewegen PP, Spreeuwenberg P. Streetscape greenery and health: Stress, social cohesion and physical activity as mediators.

Soc Sci Med 2013; **94**: 26–33.
111 Hystad P, Davies HW, Frank L, *et al.* Residential Greenness and Birth Outcomes: Evaluating the Influence of Spatially Correlated Built-Environment Factors. *Environ Health Perspect* 2014; **122**: 1095–102.
112 Villeneuve PJ, Jerrett M, G. Su J, *et al.* A cohort study relating urban green space with mortality in Ontario, Canada. *Environ Res* 2012; **115**: 51–8.
113 Wilker EH, Wu C Da, McNeely E, *et al.* Green space and mortality following ischemic stroke. *Environ Res* 2014; **133**: 42–8.
114 Markevych I, Schoierer J, Hartig T, *et al.* Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environ Res* 2017; **158**: 301–17.
115 Canadian Urban Environmental Health Research Consortium. NDVI Landsat Metadata. 2018 https://canue.ca/wp-content/uploads/2018/11/CANUE-Metadata-NDVI-Landsat-Annual.pdf.
116 Benchimol EI, Guttmann A, Mack DR, *et al.* Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. *J Clin Epidemiol* 2014; **67**: 887–96.
117 Bechle MJ, Millet DB, Marshall JD. National Spatiotemporal Exposure Surface for NO2: Monthly Scaling of a Satellite-Derived Land-Use Regression, 2000-2010. *Environ Sci Technol* 2015; **49**: 12297–305.
118 Kramer MS, Platt RW, Wen SW, *et al.* A New and Improved Population-Based Canadian Reference for Birth Weight for Gestational Age. *Pediatrics* 2001; **108**: e35.
119 Andersson T, Alfredsson L, Källberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. *Eur J Epidemiol* 2005; **20**: 575–9.
120 Virta L, Auvinen A, Helenius H, Huovinen P, Kolho KL. Association of repeated exposure to antibiotics with the development of pediatric crohn's disease - A nationwide, register-based Finnish case-control study. *Am J Epidemiol* 2012; **175**: 775–84.
121 Ananthakrishnan AN, Khalili H, Higuchi LM, *et al.* Higher predicted vitamin D status is associated with reduced risk of crohn's disease. *Gastroenterology* 2012; **142**: 482–9.
122 Liu J, Bennett KJ, Harun N, Probst JC. Urban-rural differences in overweight status and physical inactivity among US children aged 10-17 years. *J Rural Heal* 2008; **24**: 407–15.
123 Invernizzi G, Ruprecht A, Mazza R, *et al.* Particulate matter from tobacco versus diesel car exhaust: An educational perspective. *Tob Control* 2004; **13**: 219–21.
124 Tuvlin JA, Raza SS, Bracamonte S, *et al.* Smoking and inflammatory bowel disease: Trends in familial and sporadic cohorts. *Inflamm Bowel Dis* 2007; **13**: 573–9.
125 Andersson RE, Olaison G, Tysk C, Ekbom A. Appendectomy is followed by increased risk of Crohn's disease. *Gastroenterology* 2003; **124**: 40–6.
126 Beamish LA, Osornio-Vargas AR, Wine E. Air pollution: An environmental factor contributing to intestinal disease. *J Crohn's Colitis* 2011; **5**: 279–86.

127 Cerdá B, Pérez M, Pérez-Santiago JD, Tornero-Aguilera JF, González-Soltero R, Larrosa M. Gut microbiota modification: Another piece in the puzzle of the benefits of physical exercise in health? *Front Physiol* 2016; **7**: 1–11.
128 Cook MD, Allen JM, Pence BD, *et al.* Exercise and gut immune function: Evidence of alterations in colon immune cell homeostasis and microbiome characteristics with exercise training. *Immunol Cell Biol* 2016; **94**: 158–63.
129 Spencer J, MacDonald TT, Finn T, Isaacson PG. The development of gut associated lymphoid tissue in the terminal ileum of fetal human intestine. *Clin Exp Immunol* 1986; **64**: 536–43.
130 Sanchez-Muñoz F, Dominguez-Lopez A, Yamamoto-Furusho JK. Role of cytokines in inflammatory bowel disease. *World J Gastroenterol* 2008; **14**: 4280–8.
131 Brand S. Crohn's disease: Th1, Th17 or both? The change of a paradigm: New immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. *Gut* 2009; **58**: 1152–67.
132 Ueno A, Jeffery L, Kobayashi T, Hibi T, Ghosh S, Jijon H. Th17 plasticity and its relevance to inflammatory bowel disease. *J Autoimmun* 2018; **87**: 38–49.
133 Törnqvist H, Mills NL, Gonzalez M, *et al.* Persistent endothelial dysfunction in humans after diesel exhaust inhalation. *Am J Respir Crit Care Med* 2007; **176**: 395–400.
134 Hernandez-Pando R, Rook GA. The role of TNF-alpha in T-cell-mediated inflammation depends on the Th1/Th2 cytokine balance. *Immunology* 1994; **82**: 591–5.
135 Rook GAW. Review series on helminths, immune modulation and the hygiene hypothesis: The broader implications of the hygiene hypothesis. *Immunology* 2009; **126**: 3–11.
136 Misra R, Faiz O, Munkholm P, Burisch J, Arebi N. Epidemiology of inflammatory bowel disease in racial and ethnic migrant groups. *World J Gastroenterol* 2018; **24**: 424–37.
137 O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant* 2007; **41**: 673–90.
138 Pham TTH, Apparicio P, Séguin AM, Landry S, Gagnon M. Spatial distribution of vegetation in Montreal: An uneven distribution or environmental inequity? *Landsc Urban Plan* 2012; **107**: 214–24.
139 Heynen N, Perkins HA, Roy P. The Political Ecology of Uneven Urban Green Space. *Urban Aff Rev* 2006; **42**: 3–25.
140 Sheppard L, Burnett RT, Szpiro AA, *et al.* Confounding and exposure measurement error in air pollution epidemiology. *Air Qual Atmos Heal* 2012; **5**: 203–16.
141 Chen H, Burnett RT, Copes R, *et al.* Ambient fine particulate matter and mortality among survivors of myocardial infarction: Population-based cohort study. *Environ Health Perspect* 2016; **124**: 1421–8.
142 Lavigne É, Bélair M-A, Do MT, *et al.* Maternal exposure to ambient air pollution and risk of early childhood cancers: A population-based study in Ontario,

Canada. *Environ Int* 2017; **100**: 139–47.
143 Hystad P, Setton E, Cervantes A, *et al.* Creating national air pollution models for population exposure assessment in Canada. *Environ Health Perspect* 2011; **119**: 1123–9.
144 Fell DB, Dodds L, King WD. Residential mobility during pregnancy. *Paediatr Perinat Epidemiol* 2004; **18**: 408–14.
145 Chen L, Bell EM, Caton AR, Druschel CM, Lin S. Residential mobility during pregnancy and the potential for ambient air pollution exposure misclassification. *Environ Res* 2010; **110**: 162–8.
146 Haertsch M, Campbell E, Sanson-Fisher R. What is recommended for healthy women during pregnancy? A comparison of seven prenatal clinical practice guideline documents. *Birth* 1999; **26**: 24–30.
147 Weng HY, Hsueh YH, Messam LLM V., Hertz-Picciotto I. Methods of covariate selection: Directed acyclic graphs and the change-in-estimate procedure. *Am J Epidemiol* 2009; **169**: 1182–90.
148 Basu D, Lopez I, Kulkarni A, Sellin JH. Impact of race and ethnicity on inflammatory bowel disease. *Am J Gastroenterol* 2005; **100**: 2254–61.
149 Krieger N. Overcoming the absence of socioeconomic data in medical records: Validation and application of a census-based methodology. *Am J Public Health* 1992; **82**: 703–10.
150 Glazier RH, Creatore MI, Agha MM, Steele LS. Socioeconomic Misclassification in Ontario's Health Care Registry. *Can J Public Heal* 1996; **94**.
151 Buajitti E, Chiodo S, Rosella LC. Agreement between area- and individual-level income measures in a population-based cohort: Implications for population health research. *SSM - Popul Heal* 2020; **10**: 100553.

- End of Chapter 5 -

Appendix A – RECORD statement checklist for air pollution manuscript

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	(a) Title: "a population-based cohort study" (b) Abstract "increased associations with pediatric-onset IBD were noted…"	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Abstract: "using linked population-based health administrative data Abstract: "all singleton live births in Ontario, Canada between 1991 and 2013" Abstract: "using linked population-based health administrative data"
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Intro (page 5): "air pollution is hypothesized to contribute to the risk of IBD through.."		
Objectives	3	State specific objectives, including any prespecified hypotheses	Intro (page 5): "to investigate the associations between		

			maternal and childhood exposures to NO_2, $PM_{2.5}$, O_3 and O_x with paediatric-onset IBD"		
Methods					
Study Design	4	Present key elements of study design early in the paper	Methods (page 6): "this study was a retrospective cohort comprised of…"		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods (page 6-9): "Ontario…April 1st 1991-March 31st 2014…followed until first of…."		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants	(a) Methods (page 6): "children were excluded from the analyses if…" Methods (page 10): "children were followed until…" (b) Not applicable here	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical	(6.1): methods (page 6): "all registered singleton live births" (6.2) methods (page 7): "using an algorithm that has been previously validated…" (6.3) Not included as no individuals were lost as a result of data linkage of different datasets

		(b) Cohort study - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Methods (pages 7-11): "Children diagnosed with IBD before 18 years of age…"	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	(7.1) Methods (page 7): "…with an IBD diagnostic code (ICD-9 codes 555, 556; ICD-10 code K50, K51)…"
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Methods (pages 8-9): "NO_2 was assessed…" "A spatial $PM_{2.5}$ surface was derived…" "O_3 was assessed…"		
Bias	9	Describe any efforts to address potential sources of bias	Methods (page 10): "estimates were then further adjusted for potential confounding factors…"		

Study size	10	Explain how the study size was arrived at	No pre-study calculation was done to determine sample size		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Methods (page 10): "and was divided into quintiles"		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy	(a) Methods (page 10): "Cox proportional hazards models were used…" (b) Methods (page 11): "sensitivity analyses restricted to children diagnosed <10 years of age was conducted" (c) Methods (page 7, 9-10): "missing gestational age variable" (d) Methods (page 11): "children were followed until…" (e) Methods (page 11): "a final sensitivity analysis that excluded observations that had extrapolated exposure estimates…"		

		(e) Describe any sensitivity analyses			
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	(12.1) methods (page 6): "…with full anonymized and unedited datasets were available to researchers for this study" (12.2) Methods (page 11): "For the 1.2% of the cohort that were missing…"
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	(12.3) Methods (page 6): "linked using unique-encoded identifiers"
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	(a) Supplementary Figure 1 (b) Supplementary Figure 1 (c) Supplementary Figure 1	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	(13.1) Supplementary Figure 1

Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	(a) Table 1 (b) Supplementary Table 1 (those missing covariates/exposures were excluded) (c) Results (page 13): "children were followed-up for a total…"		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	Results (page 13): "of these, 2491 children developed IBD…"		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval).	(a) Figure 1, Tables 2-3 (b) No categorization of main exposure occurred in this study (c) Not relevant here		

		Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	(Table 3, Supplementary Table 5, Supplementary Figure 1, Supplementary table 6)		
Discussion					
Key results	18	Summarise key results with reference to study objectives	Discussion (Page 15): "we observed an association between childhood exposures to oxidant capacity…"		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion (Page 17): "there are also some limitations to consider…"	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	(19.1) Discussion (page 17): "the health administrative data were not collected for the purpose of research…"

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion (page 17): "we found some evidence of an association between exposure to oxidant capacity during childhood…"		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion (page 16): "generalizable to other similar populations that have comparable exposures to air pollution"		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Acknowledgements (page 2): "this study also received some funding from Health Canada…"		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	(22.1): Instructions on third party access to ICES data to be submitted with journal

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

Appendix B – RECORD statement checklist for greenspace manuscript

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	(a) Title (page 1): "a population-based cohort study" (b) Abstract (page 2)	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Abstract (page 2): "using linked population-based health administrative data Abstract (page 2): "all singleton live births in Ontario, Canada between April 1st, 1991 and March 31st 2014" Abstract (page 2): "using linked population-based health administrative and environmental datasets"
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Intro (page 4): "urbanization has led to a decrease in the populations exposure to plants and the natural environment"		

Objectives	3	State specific objectives, including any prespecified hypotheses	Intro (page 5): "The objective of this study was to…"		
Methods					
Study Design	4	Present key elements of study design early in the paper	Methods (page 5) "we conducted a retrospective cohort study…"		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods (page 6) "from all children born in an Ontario hospital between…"		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed	a – Methods (page 6) "we excluded mother-infant pairs that had invalid identification numbers…" (b) – Not applicable	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the	(6.1): Methods (page 6): "We included all children born in an Ontario hospital…" (6.2) Not reported as no code or algorithm were used to select the study population (6.3) Not reported as no loss of individuals due to linkage occurred in this study

		Case-control study - For matched studies, give matching criteria and the number of controls per case		number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Methods (page 6-10) "exposure to greenspace was estimated using…",	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Methods (pages 6-7): "The optimal algorithm to capture paediatric-onset IBD cases required…"
Data sources/ measure ment	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Methods (pages 8-9): "We therefore adjusted for the following variables…"		
Bias	9	Describe any efforts to address potential sources of bias	Methods (page 9): "…we fit a mixed-effects model"		
Study size	10	Explain how the study size was arrived at	Results (page 11): "after exclusions there were…"		
Quantitati ve variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Methods (page 9): "we therefore adjusted for the following variables…"		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed	(a) Methods (pages 9-10) "we built models iteratively..." (b) Methods (page 10): "we explored how the effects of greenspace might differ…"		

		(d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	(c) Methods (page 8): "children that had invalid or missing greenspace estimates…were excluded" (d) Methods (page 8): "children were followed up until…" (e) Methods (page 10): "we conducted a pre-planned sensitivity analysis"		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	(12.1) Methods (page 7): "authors had complete access to all uncleaned data used in the study" (12.2) Not reported as the data cleaning steps taken in this study were straightforward
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	(12.3) Methods (page 7)

Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	(a) Supplementary Figure 1 (b) Supplementary Figure 1 (c) Supplementary Figure 1	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	(13.1) Supplementary Figure 1
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	(a) Table 1 (b) Table 1 (c) Results (Page 11): "Individuals were followed-up for a total of…"		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	Results (page 11): "among them, 3 449 developed IBD…"		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted	(a) Results (Table 2 + 3) (b) Results (Table 5)		

		estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	(c) Not relevant here		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Table 3 Supplementary Table 1 and 2		
Key results	18	Summarise key results with reference to study objectives	Discussion (page 19): "In this study, we found that increasing levels of residential greenspace during childhood was associated with reduced risk of IBD…"		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion (page 21-22): "However, several limitations to this study should be considered…"	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	(19.1) Discussion (page 22): "another limitation includes the use of secondary health-administrative data that were not collected for research purposes…"

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion (page 23): "in summary, greater residential greenspace during childhood was associated with a reduced risk of IBD…"		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion (page 21): "ensures that the results of this study are generalizable…"		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Acknowledgements (page 24): "This study was supported by ICES…also received funding from Health Canada…"		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	(22.1): added statement on accessing data in submission website

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

Appendix C – Receipt of submission to Gut

Submission Confirmation

Print

Thank you for your submission

Submitted to
Gut

Manuscript ID
gutjnl-2019-320212

Title
Exposure to residential greenspace during childhood reduces the risk of paediatric-onset inflammatory bowel disease: a population-based cohort study

Authors
Elten, Michael
Benchimol, Eric
Fell, Deshayne
Kuenzig, M Ellen
Smith, Glenys
Kaplan, Gilaad
Chen, Hong
Crouse, Dan
Lavigne, Eric

Date Submitted
31-Oct-2019