**Supplementary material**

Michaux et al

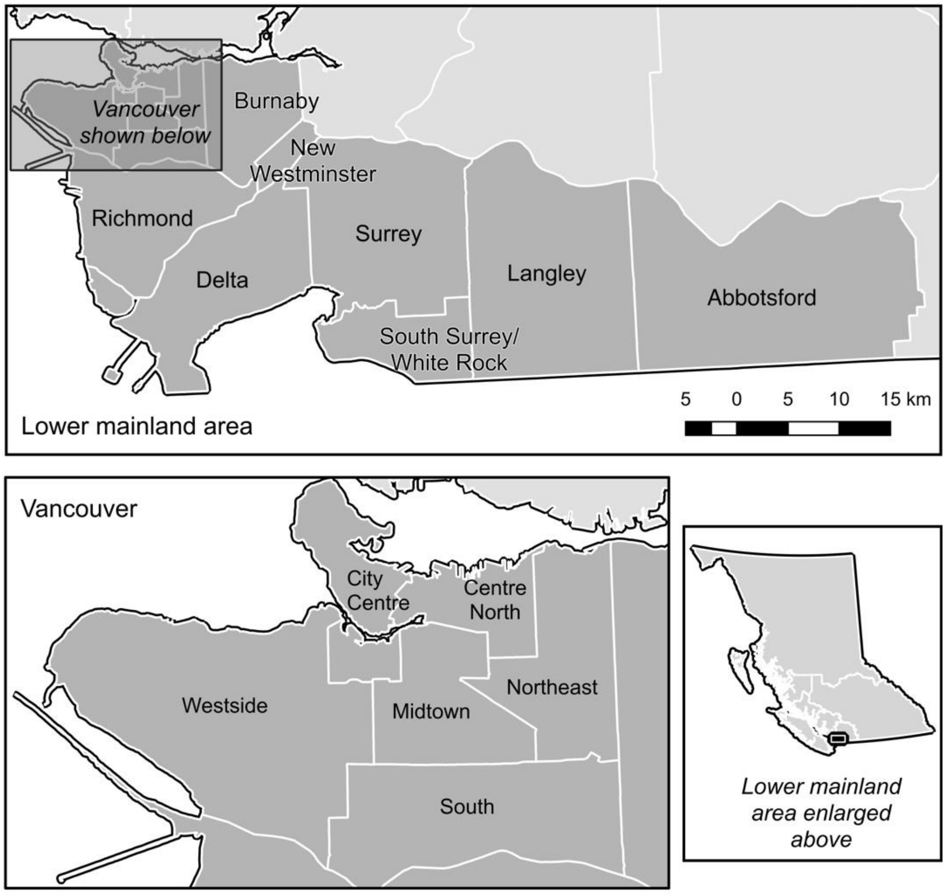
Table of Contents



## Description

This Supplement accompanies the article "Spatial Cluster Mapping and Environmental Modeling in Pediatric Inflammatory Bowel Disease" by Michaux et al. In this supplement we provide a detailed explanation of quantitative methods and data employed in the article. The next section shows how several datasets used were prepared, generated, or explored. The following section provides details about the population level risk modeling for Pediatric Inflammatory Bowel Disease (IBD).

**Data Preparation, Processing, and Exploration**




**Supplementary Figure 1 Local Health Areas of the City of Vancouver and Lower Mainland, British Columbia.**

**Missing exposure data**

Environmental exposure data was unavailable for Snow Country so it was excluded from exposure modeling. Exposure data was missing from 2001–2013 for Telegraph Creek. Greenness data was missing for Prince Rupert and Burns Lake for 2001 and 2002, and for Kitimat, Nisga'a, Smithers, Terrance, and Upper Skeena for 2001.

**Environmental exposure data descriptions**

Residential vegetation greenness

The Normalized Difference Vegetation Index (NDVI) is a commonly used measure of green vegetation cover calculated from near-infrared and visible red land surface reflection. NDVI values used in this study were derived from Landsat satellite data at a 30 m resolution and provided for six-digit postal codes by CANUE.

UV vitamin D dose

Mean daily vitamin D dose from solar UV radiation was calculated using solar radiation monitoring, ozone data, and dew point temperature, and adjusted for UV intensification from snow cover and latitude. Dose estimates were developed and produced by Environment Canada and Cancer Care Ontario at a roughly 100 km grid resolution and distributed for postal codes by CANUE. We averaged monthly values to produce average winter (December through February) and average summer (June through August) UV vitamin D.

Air pollution: $NO_2$, $O_3$, and $PM_{2.5}$

Annual average $NO_2$ concentrations in parts per billion received from CANUE were estimated for each postal code with land use regression models combining satellite-based approximations of $NO_2$ from 2005 - 2001, total length of roads within 10 km of the postal code, amount of land classified for industrial use within 2 km of the postal code, and quantity of summer precipitation. Environment and Climate Change Canada estimated hourly ground-level $O_3$ concentration for 2002 to 2009 with the Canadian Hemispherical Regional Ozone and NOx System model and for 2010 to 2015 with the Global Environmental Multi-scale Modeling Air Quality and Chemistry model. Ground-level $O_3$ measurements were integrated with model estimates. Annual warm season (May - September) average of the highest rolling 8-hour daily average concentration (parts per billion) for postal codes were produced and distributed by CANUE. Finally, annual average $PM_{2.5}$ (micrograms per meter$^3$) concentrations were produced by the Atmospheric Composition Analysis Group and provided at the six-digit postal code level by CANUE. The GEOS-Chem chemical transport model which associates aerosol optical depths to surface $PM_{2.5}$ concentration was used to produce surface-level $PM_{2.5}$ estimates from satellite-measured aerosol optical depths. To further refine the data, initial estimates were adjusted with surface monitor $PM_{2.5}$ data using geographically weighted regression.

Pesticides

We used data for common pesticides, including metam used for fruits and vegetables, petroleum oil employed for grapes and orchards, as well as the total glyphosate employed in major crops (wheat, corn, and others). These data were extracted from the Global Pesticide Grids, Version 1.01. We used estimates for 2015. The dataset has 5 arc-minute resolution, which yields cells of roughly 5 km by 9 km in BC. Pesticide applications are measured in kilograms per hectare per year and based on U.S. Geological Survey Pesticide National Synthesis Project and the Food and Agriculture Organization Corporate Statistical Database pesticide databases.

**Population weighting procedure**

To create environmental data at the appropriate scale of analysis, average population exposure to each environmental variable was approximated for each LHA. To accomplish this, postal code points were spatially joined with the nearest census Dissemination Area. Census population age 0-19 for each Dissemination Area was divided equally between all linked postal codes, and this population estimate was used to population-weight environmental exposures in each LHA. Multiple geographic areas can be represented by the same postal code; the data used in this study was derived for only the location that was most representative of population residential location. As a result, the population weighting process captured 92.6% of BC's youth population but excluded some populated areas where census population geography did not align with a postal code location. Postal and census geometry changes over time, so the population weighting process was repeated for each year of the study period using census population from the closest census year. Mean yearly population exposure was then averaged for the period of 2001 - 2016 for each LHA, resulting in a single exposure value at each location.

**Interregional distances and times**

Our studies of spatial clustering and econometric diagnostics both require a mathematical characterization of possible interconnections among IBD cases in space.

There are a variety of ways to characterize adjacency, or, in its non-binary generalization, proximity, between two regions. In the absence of a singular theoretical justification for choosing one over another in our analyses, we felt it was more robust to examine the results of our analyses using multiple such notions of proximity, which therefore involved calculating distances among LHAs in multiple ways. Here, we explain the methods we used to find Euclidean distances, driving distances, and driving times.

First, we characterize how we found representative points in LHAs with which to calculate interregional distances. LHAs can be large and their centroids are not necessarily characteristic of population distributions, nor even near the road network. As such, geometric centroids may not be ideal for calculating distances from one LHA to another, Euclidean or otherwise. We thus found the highest-density (non-sliver) 2016 Census Dissemination Area within a given LHA and measure or route to/from the centroid of that polygon as the point that characterizes the LHA for distances.

What we term Euclidean distances for the purposes of exposition are, more precisely, the spherical distances (calculated using the R package *sf* with the st_distance command implementing the *s2* geometry library) between the geometric centroids (calculated within the BC Albers / EPSG:3005 coordinate system) of the Census Dissemination Areas found as described above.

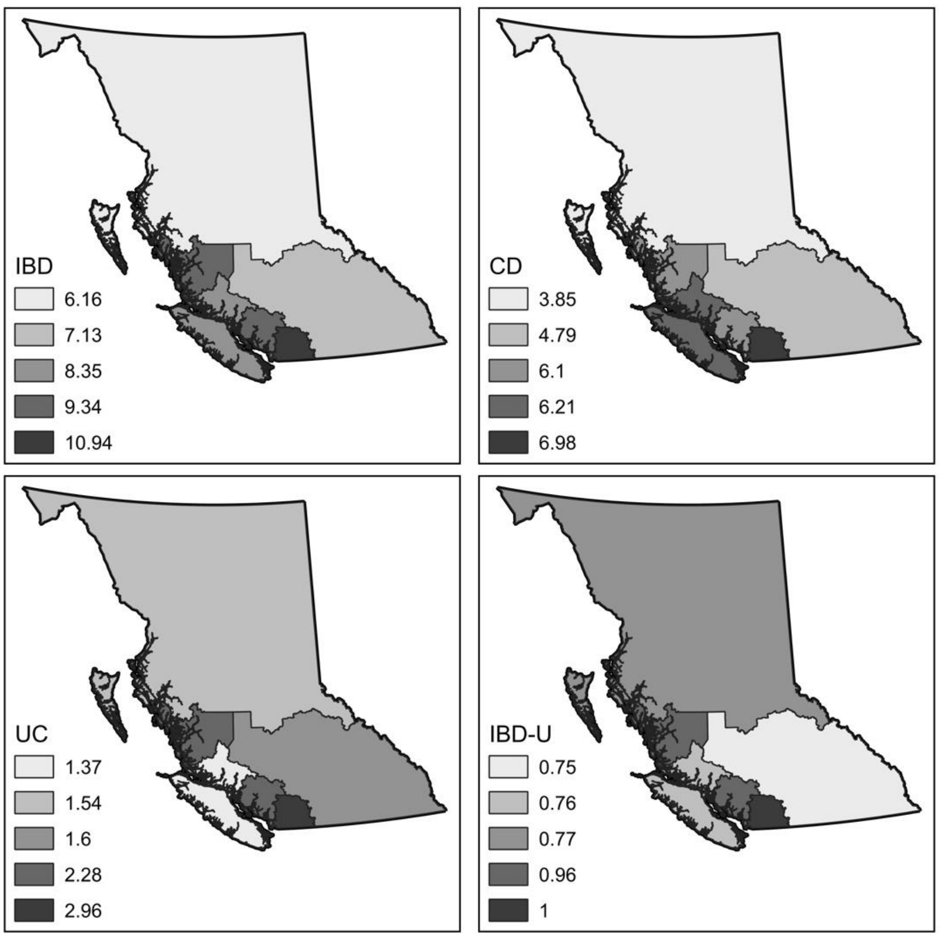
For the purposes of calculating driving distances (which may be better in some cases than Euclidean/spherical distances for characterizing human proximity), we chose and implemented the router *Valhalla* (https://github.com/valhalla/valhalla) using OpenStreetMap data from 09-2021. We adapted router query functions provided by the R wrapper *valhallr*.

The combination of the router, the data, and the LHA center point identification strategy described earlier yielded 87 of 89 LHAs as interconnected. The two that were not interconnected were LHA_CD 337 - Central Coast (in which our method selected Bella Bella as the key central point for the LHA but the router did not have an appropriate connection between Bella Bella and its long-distance ferry, separated as it is by a waterway) and LHA_CD 433 - Vancouver Island West (in which our approach selected

Kyuquot as the center of the LHA, but none of the boat routes serving that community were in the router's dataset.) Calculations involving driving distances or times therefore effectively dropped these two points from the dataset.

**Tables and visualization of IBD incidence**

IBD case counts for this study included patients with Crohn's Disease (CD), Ulcerative Colitis (UC), and IBD-unclassified (IBD-U). Supplementary Figure 2 and Supplementary Table 1 below provide a more detailed exploration of IBD incidence in British Columbia during the study period to accompany Table 1 in the main text.




**Supplementary Figure 2 British Columbia average pediatric incidence per 100,000 of IBD, CD, UC, and IBD-unclassified (IBD-U) by Health Authority, 2001–2016.**

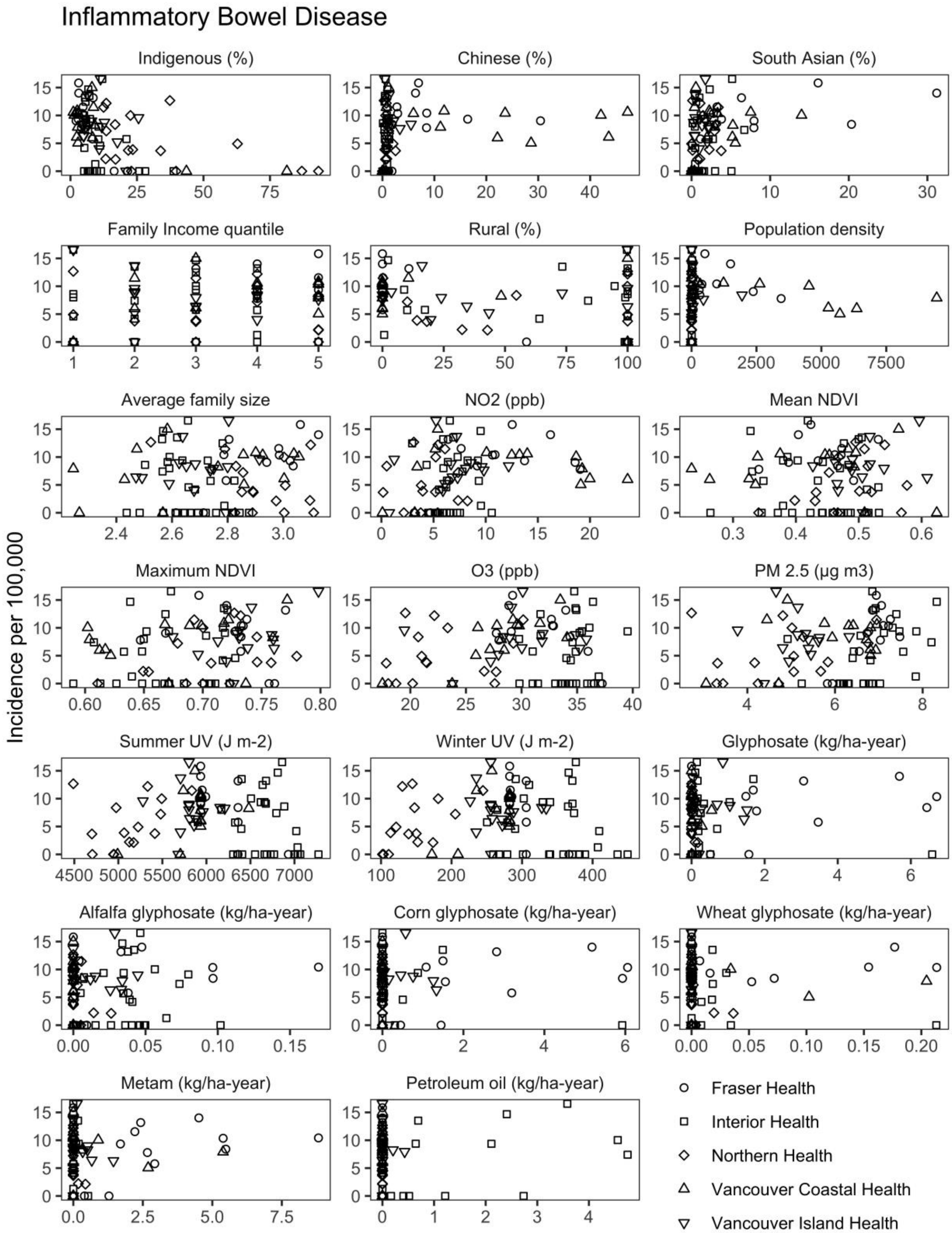
**Supplementary Table 1 British Columbia average pediatric incidence for IBD by age group, from 2001–2016, Values for CD, UC, and IBD-U suppressed due to small case numbers**

| Age | Health Authority | Cases | Incidence per 100,000 | 95% CI for incidence | |
|---|---|---|---|---|---|
| 0 to 4 | British Columbia | 67 | 1.94 | 1.5 | 2.46 |
| | Fraser | 34 | 2.44 | 1.69 | 3.4 |
| | Interior | 8 | 1.57 | 0.68 | 3.09 |
| | Northern | 7 | 2.46 | 0.99 | 5.07 |
| | Vancouver Coastal | 8 | 1.05 | 0.45 | 2.07 |
| | Vancouver Island | 10 | 1.96 | 0.94 | 3.6 |
| 5 to 9 | British Columbia | 229 | 6.27 | 5.48 | 7.14 |
| | Fraser | 106 | 7.24 | 5.93 | 8.75 |
| | Interior | 31 | 5.47 | 3.71 | 7.76 |
| | Northern | 7 | 2.35 | 0.94 | 4.84 |
| | Vancouver Coastal | 55 | 7.15 | 5.39 | 9.31 |
| | Vancouver Island | 30 | 5.41 | 3.65 | 7.72 |
| 10 to 14 | British Columbia | 644 | 16.14 | 14.92 | 17.44 |
| | Fraser | 299 | 19.07 | 16.97 | 21.36 |
| | Interior | 86 | 13.37 | 10.69 | 16.51 |
| | Northern | 34 | 10.62 | 7.35 | 14.84 |
| | Vancouver Coastal | 125 | 14.93 | 12.43 | 17.79 |
| | Vancouver Island | 100 | 16.08 | 13.08 | 19.56 |
| 15 to 16 | British Columbia | 243 | 14.09 | 12.37 | 15.98 |
| | Fraser | 119 | 17.85 | 14.79 | 21.36 |
| | Interior | 18 | 6.44 | 3.82 | 10.18 |
| | Northern | 16 | 11.92 | 6.81 | 19.36 |

| | | | | | |
|---|---|---|---|---|---|
| | Vancouver Coastal | 66 | 17.83 | 13.79 | 22.68 |
| | Vancouver Island | 24 | 8.75 | 5.61 | 13.02 |

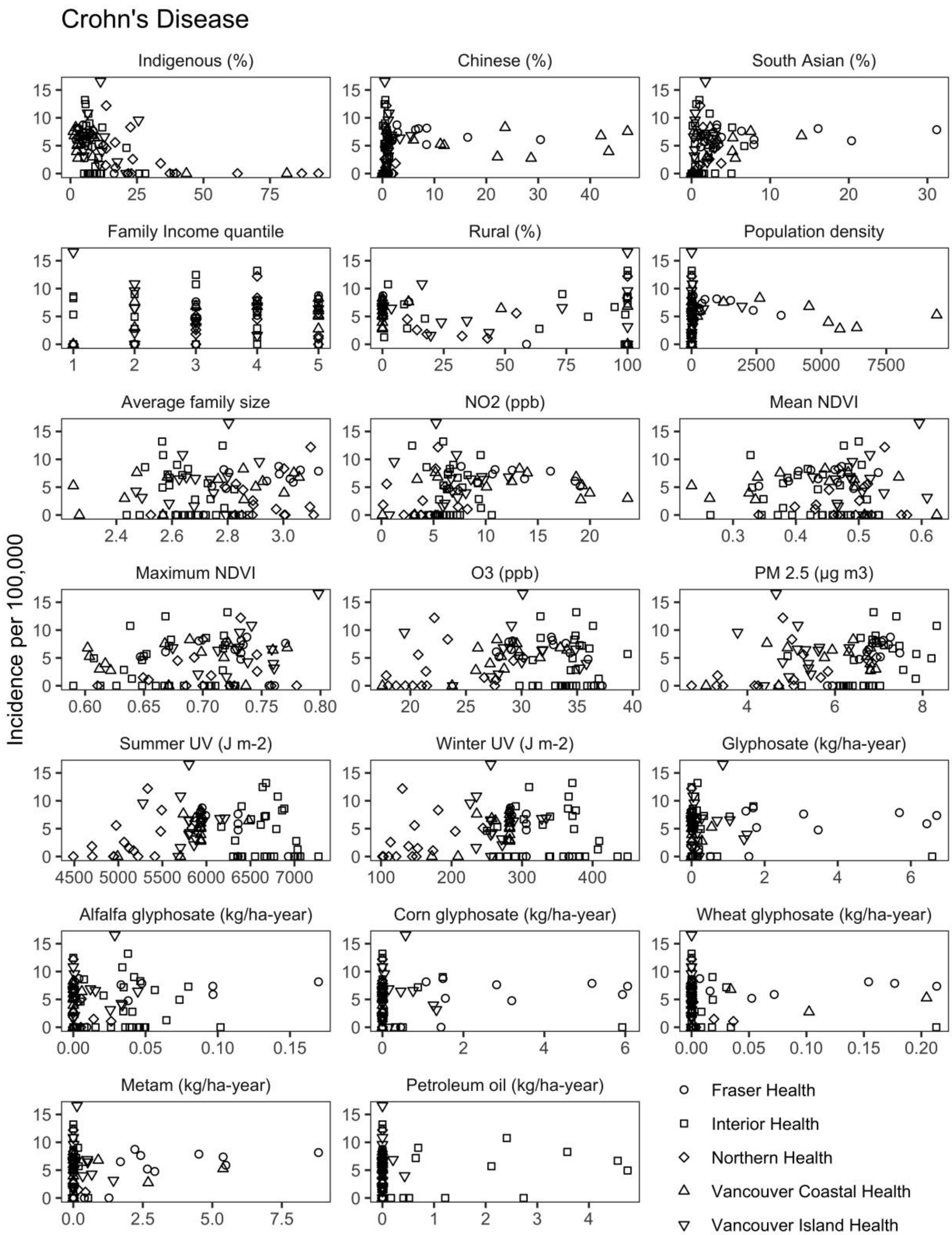
**Tables and visualization of modeling variables**

Supplementary Figures 3–5 and Supplementary Tables 2 below provide additional details about the variables used in environmental exposure modeling.

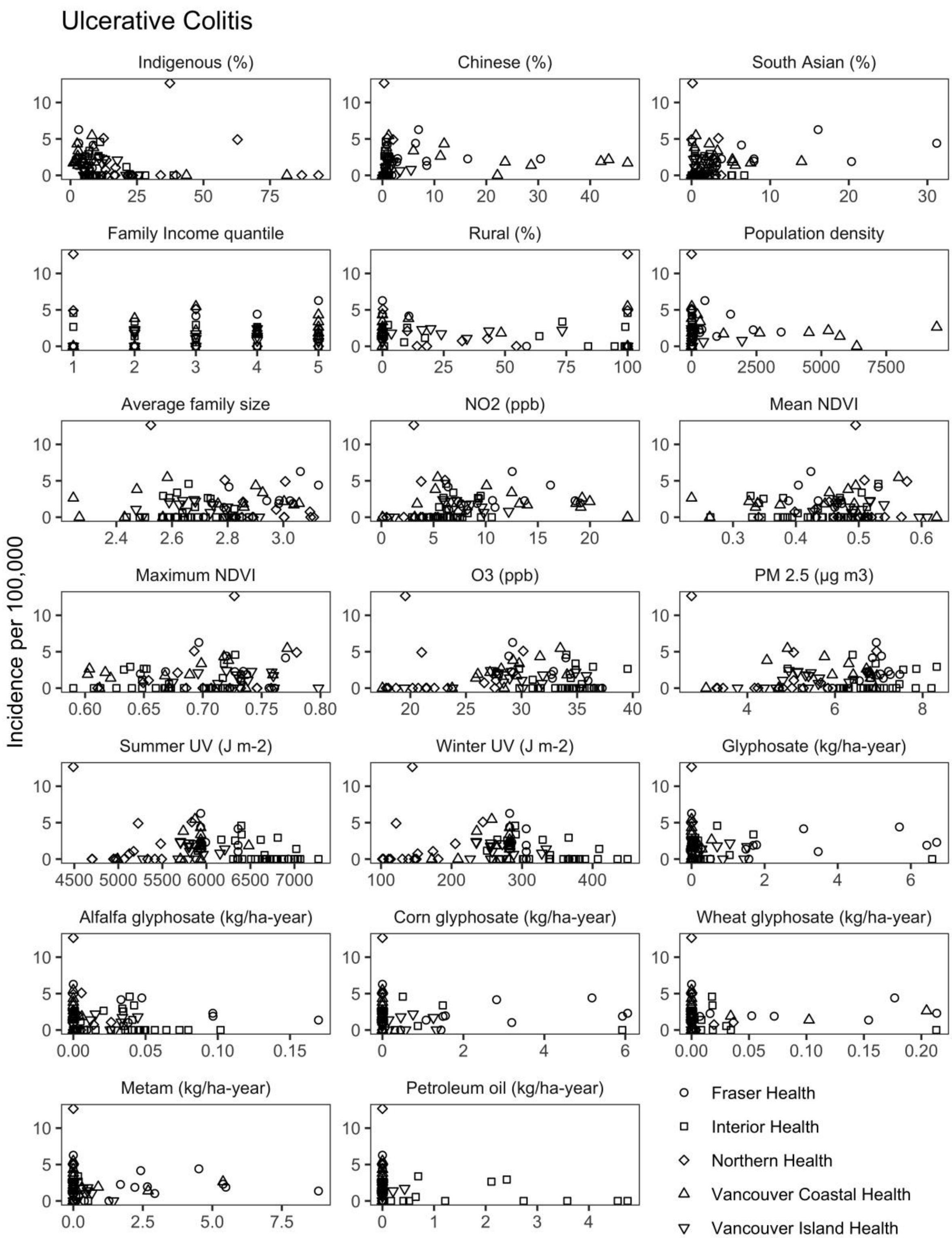



**Supplementary Figure 3 Scatterplots of modeling variables and pediatric inflammatory bowel disease incidence for British Columbia's Local Health Areas used**

**in modeling during 2001–2016.**



**Supplementary Figure 4 Scatterplots of modeling variables and pediatric Crohn's disease incidence for British Columbia's Local Health Areas used in modeling during 2001–2016.**




**Supplementary Figure 5 Scatterplots of modeling variables and pediatric Ulcerative Colitis incidence for British Columbia's Local Health Areas used in modeling during**

**2001–2016.**

**Supplementary Table 2 Mean values of Local Health Area explanatory variables for each Health Authority (Snow Country excluded)**

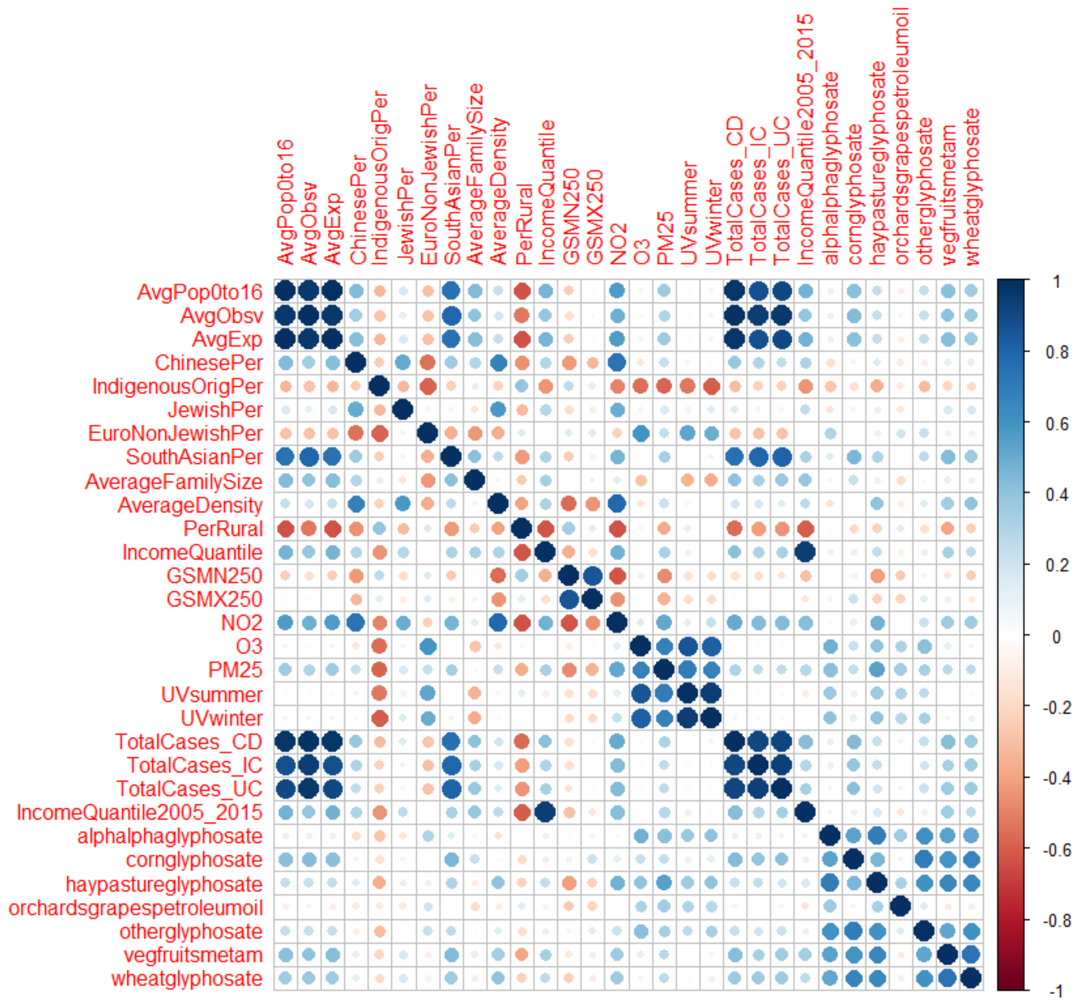
| Variable | Mean of LHA values for each Health Authority | | | | |
|---|---|---|---|---|---|
| | Fraser | Interior | Northern | Vancouver Coastal | Vancouver Island |
| Average population age 0–16.9 during 2001 - 2016 | 24,498.56 | 4,031.35 | 4,041.25 | 12,220.54 | 8,756.9 |
| Chinese ethnic origin (%) | 6.92 | 0.61 | 0.82 | 17.16 | 1.51 |
| Indigenous ethnic origin (%) | 6.63 | 10.32 | 33.18 | 12.64 | 10.84 |
| Jewish ethnic origin (%) | 0.41 | 0.35 | 0.22 | 1.00 | 0.41 |
| Non-Jewish European ethnic origin (%) | 65.52 | 82.33 | 60.44 | 53.71 | 79.77 |
| South Asian ethnic origin (%) | 8.14 | 1.38 | 1.28 | 3.69 | 1.27 |
| Average family income ($) | 984,53.11 | 83,289.72 | 91,301.50 | 107,766.93 | 89,598.74 |
| Family size | 2.89 | 2.67 | 2.91 | 2.71 | 2.69 |
| Population density (per square km) | 767.66 | 7.95 | 0.79 | 2,553.25 | 197.40 |
| Rural population (%) | 13.06 | 79.09 | 67.02 | 25.65 | 43.94 |
| NDVI maximum | 0.72 | 0.67 | 0.71 | 0.68 | 0.73 |
| NDVI mean | 0.45 | 0.43 | 0.48 | 0.42 | 0.51 |
| $NO_2$ (ppb) | 10.71 | 6.63 | 4.43 | 11.98 | 6.82 |
| $O_3$ (ppb) | 32.25 | 34.18 | 22.86 | 28.04 | 29.62 |
| $PM_{2.5}$ ($\mu g$ $m^3$) | 6.83 | 6.78 | 4.77 | 5.90 | 5.14 |

| | | | | | |
|---|---|---|---|---|---|
| UV vitamin D summer (J $m^{-2}$) | 6147.79 | 6685.32 | 5115.81 | 5871.51 | 5818.33 |
| UV vitamin D winter (J $m^{-2}$) | 296.38 | 342.58 | 145.15 | 262.67 | 266.1 |
| Glyphosate used in common crops (kg/ha-year) | 2.5 | 0.43 | 0.02 | 0.07 | 0.46 |
| Glyphosate used in alfalfa crops (kg/ha-year) | 0.04 | 0.03 | 0 | 0 | 0.01 |
| Glyphosate used in corn crops (kg/ha-year) | 2.25 | 0.29 | 0 | 0 | 0.34 |
| Glyphosate used in wheat crops (kg/ha-year) | 0.05 | 0.01 | 0 | 0.02 | 0 |
| Metam used in fruits and vegetables (kg/ha-year) | 2.91 | 0.03 | 0.04 | 0.66 | 0.26 |
| Petroleum oil used in orchards and grapes (kg/ha-year) | 0.00 | 0.77 | 0.00 | 0.00 | 0.05 |

**Population Level Risk Modeling for Pediatric Inflammatory Bowel Disease**

**Exploring collinearity**

We started by exploring the correlation between covariates as assessed by examining the corrplots that depict the strength and sign of the association between variables.




**Supplementary Figure 6 Corrplot of studied variables.**

Since data from ethnic groups are proportions, in linear models it is advisable to use a subset or combination of the proportional variables that do not add to 1 to avoid problems of collinearity in predictors for parameter estimation. In plain language, what happens is that matrix operations to estimate parameters become unfeasible, and parameters are unreliable because their values become sensitive to small changes in the approximations the computer nevertheless does to get the estimates.

**Model Fit**

For all pathologies combined, then for UC and for CD in particular, we started by fitting a full model that included all ethnic groups but European of non-Jewish origin, and all other environmental covariates described in the methods.

The model, as indicated in the methods section of the main manuscript, was a Poisson rate model. This type of models is described by the following equations:

$$Y_i \sim \ Poisson(\mu_i\ )$$

where $\eta = log(\mu)$ can be described by the following generalized linear model:

$$log(\mu_i\ ) = \alpha + \sum_i \beta_i X_i$$

Where $\alpha$ is an intercept, $\beta_i$ are parameter estimates for the $X_i$ covariates. Before the analysis we converted greenness variables to percentages to ease their interpretation. After fitting a series of initial models, we proceeded to select the best model using a process of mixed backward and forward elimination.

**Supplementary Table 3 Backward elimination model selection results for Pediatric Inflammatory Bowel Disease, all pathologies included. Round 1 is used to select demographic and environmental variables not related to pesticides. Round 2 compares the best fit model from round 1 with a zero inflated Poisson and negative binomial models Round 3 includes pesticides, contrasting models where Glyphosate is consolidated across all crops, and where it is separated by crop. The best models for each round are bolded, and the best model from all global selection is furtherly** *italicized*.

| **Distribution** | **Variables Initial Model** | **AIC** | **Variables Final Model** | **AIC** |
|---|---|---|---|---|
| Round 1 | | | | |
| Poisson | EuroNonJewishPer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | 426.41 | *GSMX250* + *NO2* + *PM25* + *UVsummer* | 416.62 |

| | | | | |
|---|---|---|---|---|
| Negative Binomial | EuroNonJewishPer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | *419.94* | *GSMX250 + NO2 + PM25 + UVsummer* | 409.23 |
| Poisson | ChinesePer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | 428.53 | *GSMX250 + NO2 + PM25 + UVsummer* | 416.62 |
| Negative Binomial | ChinesePer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | 419.51 | *GSMX250 + NO2 + PM25 + UVsummer* | 409.23 |
| Poisson | SouthAsianPer + AverageFamilySize +AverageDensity +PerRural | 419.30 | *SouthAsianPer + AverageFamilySize + AverageDensity + GSMN250 + NO2 + PM25 + UVsummer* | 412.96 |

| | | | | |
|---|---|---|---|---|
| | +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter | | | |
| Negative Binomial | SouthAsianPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter | 416.48 | *SouthAsianPer + AverageFamilySize + AverageDensity + GSMN250 + NO2 + PM25 + UVsummer* | 409.88 |
| Poisson | *IndigenousOrigPer +SouthAsianPer +JewishPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter* | *399.90* | ***IndigenousOrigPer + SouthAsianPer + AverageFamilySize + AverageDensity + GSMN250 + PM25 + UVsummer*** | **389.08** |
| Negative Binomial | *IndigenousOrigPer +SouthAsianPer +JewishPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250* | *401.90* | *IndigenousOrigPer + SouthAsianPer + AverageFamilySize + AverageDensity + GSMN250 + PM25 + UVsummer* | 391.08 |

| | | | | |
|---|---|---|---|---|
| | *+NO2 +O3 +PM25 +UVsummer +UVwinter* | | | |
| Round 2 | | | | |
| Zero Inflated Poisson | *IndigenousOrigPer + SouthAsianPer + AverageFamilySize + AverageDensity + GSMN250 + PM25 + UVsummer* | *393.08* | | |
| Zero Inflated Poisson conditioned on population size | *IndigenousOrigPer + SouthAsianPer + AverageFamilySize + AverageDensity + GSMN250 + PM25 + UVsummer* | *389.63* | | |
| Hurdle Poisson | *IndigenousOrigPer + SouthAsianPer + AverageFamilySize + AverageDensity + GSMN250 + PM25 + UVsummer* | *442.29* | | |
| Hurdle Poisson conditioned on | *IndigenousOrigPer + SouthAsianPer + AverageFamilySize + AverageDensity +* | *390.44* | | |

| | | | | |
|---|---|---|---|---|
| population size | *GSMN250 + PM25 + UVsummer* | | | |
| Poisson | *IndigenousOrigPer + SouthAsianPer + AverageFamilySize + AverageDensity + GSMN250 + PM25 + UVsummer* | ***389.08*** | | |
| Round 3 | | | | |
| Poisson – Glyphosate combined | *IndigenousOrigPer+ SouthAsianPer+JewishPer+ AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 + PM25 +UVsummer +UVwinter +combinedGlyphosate +vegfruitsmetam +orchardsgrapespetroleumoil* | *390.04* | *IndigenousOrigPer + SouthAsianPer + AverageFamilySize + GSMX250 + PM25 + UVsummer + vegfruitsmetam + orchardsgrapespetroleumoil* | 376.43 |
| Poisson – Glyphosate by crop | *IndigenousOrigPer+ SouthAsianPer+JewishPer+ AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250* | *394.27* | ***IndigenousOrigPer + SouthAsianPer + AverageFamilySize + GSMX250 + PM25 + UVsummer + vegfruitsmetam +*** | ***376.43*** |

| | | | | |
|---|---|---|---|---|
| | *+NO2 +O3 + PM25 +UVsummer +UVwinter +vegfruitsmetam +orchardsgrapespetroleumoil +alphalphaglyphosate +cornglyphosate +haypastureglyphosate +Otherglyphosate +wheatglyphosate* | | ***orchardsgrapespetroleumoil*** | |

**Supplementary Table 4 Backward elimination model selection results for Ulcerative Colitis. Round 1 is used to select demographic and environmental variables not related to pesticides. Round 2 compares the best fit model from round 1 with a zero inflated Poisson and negative binomial models Round 3 includes pesticides, contrasting models were Glyphosate is consolidated across all crops, and where it is separated by crop. The best models for each round are bolded, and the best model from all global selection is furtherly *italicized***

| **Distribution** | **Variables Initial Model** | **AIC** | **Variables Final Model** | **AIC** |
|---|---|---|---|---|
| Round 1 | | | | |
| Poisson | EuroNonJewishPer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | 270.00 | *AverageDensity + GSMX250 + NO2 + O3 + UVwinter* | 259.14 |

| | | | | |
|---|---|---|---|---|
| Negative Binomial | EuroNonJewishPer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | *264.86* | *AverageDensity + PerRural + GSMN250 + NO2 + O3 + UVwinter* | 256.87 |
| Poisson | ChinesePer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | 265.83 | *ChinesePer + AverageFamilySize + GSMN250 + NO2 + O3 + UVwinter* | 257.53 |
| Negative Binomial | ChinesePer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | 264.85 | *AverageDensity + PerRural + GSMN250 + NO2 + O3 + UVwinter* | 256.87 |
| Poisson | SouthAsianPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter | 257.52 | *SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter* | 250.63 |

| | | | | |
|---|---|---|---|---|
| Negative Binomial | SouthAsianPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter | 257.51 | *SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter* | 251.36 |
| Poisson | *IndigenousOrigPer +SouthAsianPer +JewishPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter* | 261.16 | ***SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter*** | **250.63** |
| Negative Binomial | *IndigenousOrigPer +SouthAsianPer +JewishPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter* | *261.16* | *SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter* | 251.36 |
| Round 2 | | | | |
| Zero Inflated Poisson | *SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter* | *256.02* | | |

| | | | | |
|---|---|---|---|---|
| Zero Inflated Poisson conditioned on population size | *SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter* | *253.32* | | |
| Hurdle Poisson | *SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter* | *321.84* | | |
| Hurdle Poisson conditioned on population size | *SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter* | *251.80* | | |
| Poisson | ***SouthAsianPer + PerRural + GSMN250 + NO2 + O3 + UVwinter*** | ***250.63*** | | |
| Round 3 | | | | |
| Poisson – Glyphosate combined | *IndigenousOrigPer+ SouthAsianPer+JewishPer+ AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 + PM25 +UVsummer +UVwinter +combinedGlyphosate* | 259.34 | *SouthAsianPer + PerRural + GSMN250 + O3 + UVsummer + combinedGlyphosate* | 244.38 |

| | | | | |
|---|---|---|---|---|
| | *+vegfruitsmetam +orchardsgrapespetroleumoil* | | | |
| Poisson – Glyphosate by crop | *IndigenousOrigPer+ SouthAsianPer+JewishPer+ AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 + PM25 +UVsummer +UVwinter +vegfruitsmetam +orchardsgrapespetroleumoil +alphalphaglyphosate +cornglyphosate +haypastureglyphosate +Otherglyphosate +wheatglyphosate* | *261.74* | ***SouthAsianPer + PerRural + GSMN250 + O3 + UVwinter+ alphalphaglyphosate + cornglyphosate + wheatglyphosate*** | ***240.44*** |

**Supplementary Table 5 Backward elimination model selection results for Crohn's Disease. Round 1 is used to select demographic and environmental variables not related to pesticides. Round 2 compares the best fit model from round 1 with a zero inflated Poisson and negative binomial models Round 3 includes pesticides, contrasting models where Glyphosate is consolidated across all crops, and where it is separated by crop. The best models for each round are bolded, and the best model from all global selection is furtherly *italicized***

| **Distribution** | **Variables Initial Model** | **AIC** | **Variables Final Model** | **AIC** |
|---|---|---|---|---|
| Round 1 | | | | |

| | | | | |
|---|---|---|---|---|
| Poisson | EuroNonJewishPer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | *374.53* | PerRural + GSMX250 + O3 + PM25 | 363.70 |
| Negative Binomial | EuroNonJewishPer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | *374.8* | *PerRural + GSMX250 + O3 + PM25* | 365.58 |
| Poisson | ChinesePer + AverageFamilySize + AverageDensity + PerRural + IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | *372.20* | ChinesePer + AverageFamilySize + PerRural + GSMX250 + O3 + PM25 | 361.92 |
| Negative Binomial | ChinesePer + AverageFamilySize + AverageDensity + PerRural + | 374.13 | *ChinesePer + AverageFamilySize + PerRural + GSMX250 + O3 + PM25* | *363.92* |

| | | | | |
|---|---|---|---|---|
| | IncomeQuantile2005_2015 + GSMN250 + GSMX250 + NO2 + O3 + PM25 + UVsummer + UVwinter | | | |
| Poisson | SouthAsianPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter | *374.81* | *PerRural + GSMX250 + O3 + PM25* | 363.70 |
| Negative Binomial | SouthAsianPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter | 376.81 | *PerRural + GSMX250 + O3 + PM25* | 365.58 |
| Poisson | *IndigenousOrigPer +SouthAsianPer +JewishPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter* | *349.41* | ***IndigenousOrigPer + IncomeQuantile2005_ 2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer*** | **338.24** |

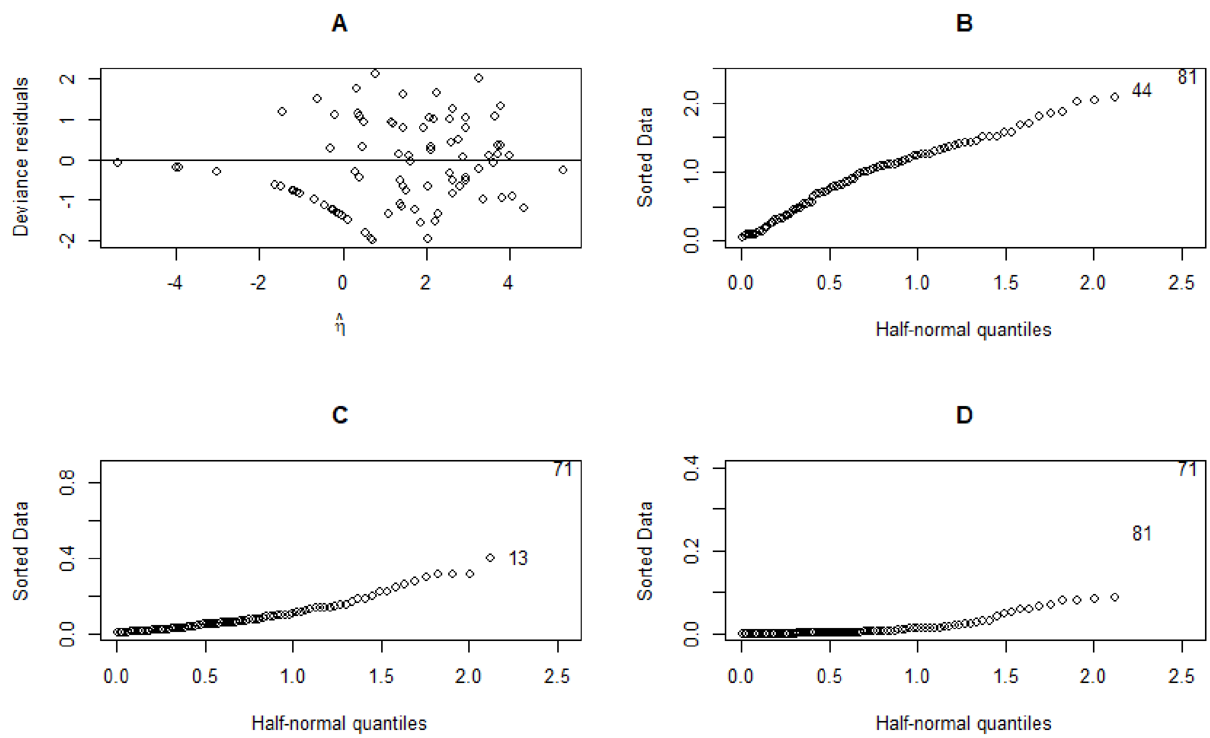
| | | | | |
|---|---|---|---|---|
| Negative Binomial | *IndigenousOrigPer +SouthAsianPer +JewishPer + AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 +PM25 +UVsummer +UVwinter* | *351.42* | *IndigenousOrigPer + IncomeQuantile2005_ 2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer* | 340.24 |
| Round 2 | | | | |
| Zero Inflated Poisson | *IndigenousOrigPer + IncomeQuantile2005_ 2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer* | *347.89* | | |
| Zero Inflated Poisson conditioned on population size | *IndigenousOrigPer + IncomeQuantile2005_ 2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer* | *344.75* | | |
| Hurdle Poisson | *IndigenousOrigPer + IncomeQuantile2005_ 2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer* | *412.71* | | |
| Hurdle Poisson conditioned on | *IndigenousOrigPer + IncomeQuantile2005_ 2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer* | *346.94* | | |

| | | | | |
|---|---|---|---|---|
| population size | | | | |
| Poisson | ***IndigenousOrigPer + IncomeQuantile2005_2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer*** | ***338.24*** | | |
| Round 3 | | | | |
| Poisson – Glyphosate combined | *IndigenousOrigPer+ SouthAsianPer+JewishPer+ AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 + PM25 +UVsummer +UVwinter +combinedGlyphosate +vegfruitsmetam +orchardsgrapespetroleumoil* | *349.94* | ***IndigenousOrigPer + IncomeQuantile2005_ 2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer + orchardsgrapespetroleumoil*** | ***336.27*** |
| Poisson – Glyphosate by crop | *IndigenousOrigPer+ SouthAsianPer+JewishPer+ AverageFamilySize +AverageDensity +PerRural +IncomeQuantile2005_2015 +GSMN250 +GSMX250 +NO2 +O3 + PM25 +UVsummer +UVwinter +vegfruitsmetam +orchardsgrapespetroleumoi* | *355.45* | ***IndigenousOrigPer + IncomeQuantile2005_ 2015 + GSMX250 + NO2 + O3 + PM25 + UVsummer + orchardsgrapespetroleumoil*** | ***336.27*** |

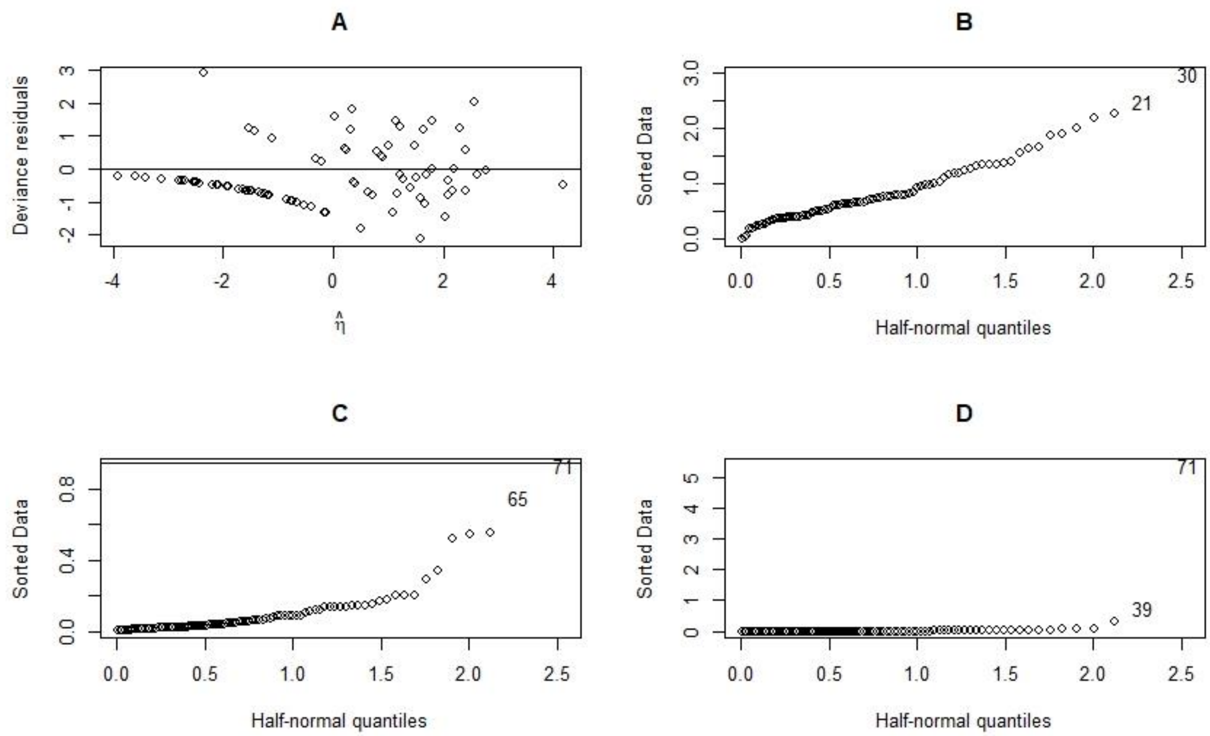
| | | | | |
|---|---|---|---|---|
| | *l* *+alphalphaglyphosate*<br>*+cornglyphosate*<br>*+haypastureglyphosate*<br>*+Otherglyphosate*<br>*+wheatglyphosate* | | | |

## Model Diagnostics

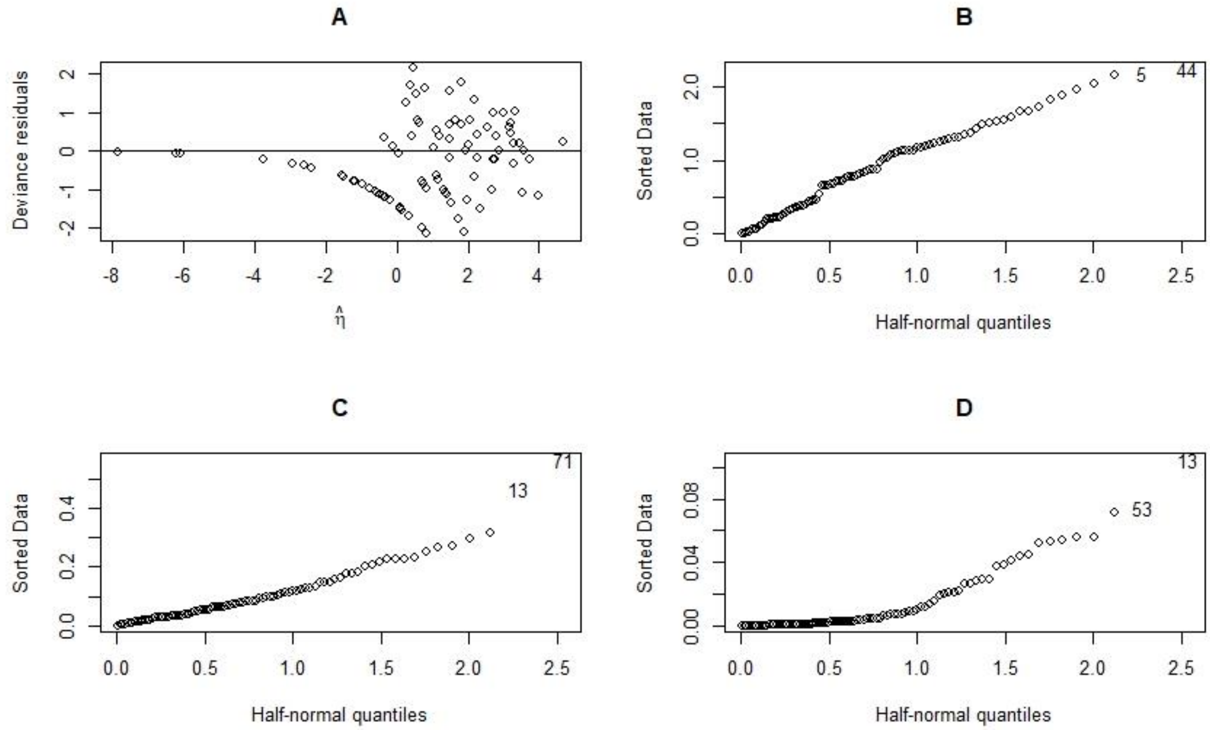
We then checked model diagnostics for the best model. We started by looking at the deviance residuals as function of the fitted values.




**Supplementary Figure 7 Model Diagnostics for the best Inflammatory Bowel Disease (all pathologies) model (A) Deviance residuals (B) Half normal plot of: residuals (C) Influence of points on fit (D) Cook's distance.**




**Supplementary Figure 8 Model Diagnostics for the best Ulcerative Colitis model (A) Deviance residuals (B) Half normal plot of: residuals (C) Influence of points on fit (D) Cook's distance.**



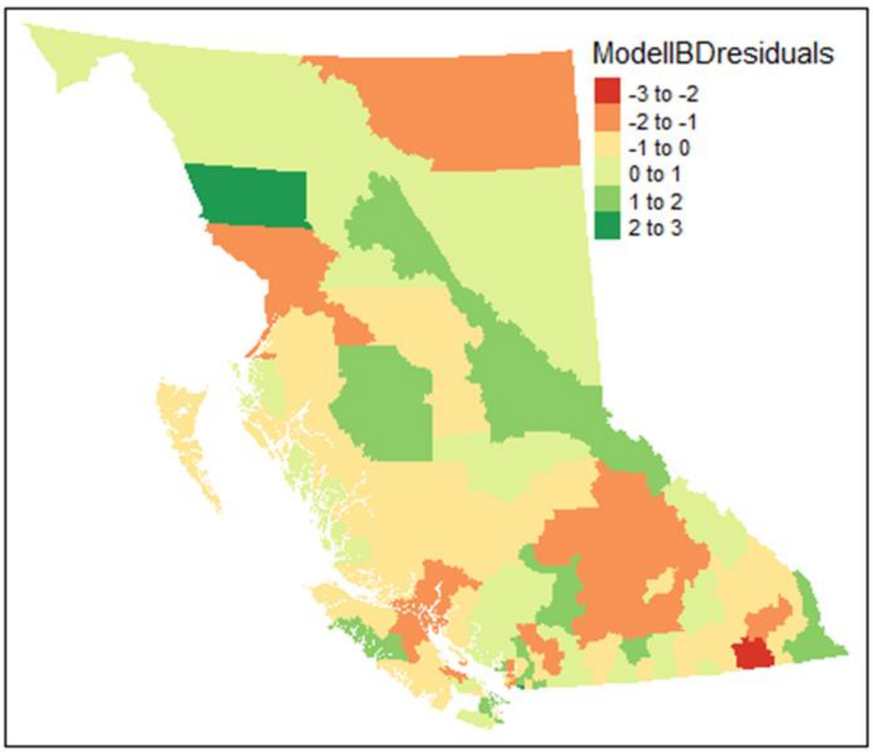

**Supplementary Figure 9 Model Diagnostics for the best Crohn's Disease model (A) Deviance residuals (B) Half normal plot of residuals (C) Influence of points on fit (D) Cook's distance.**

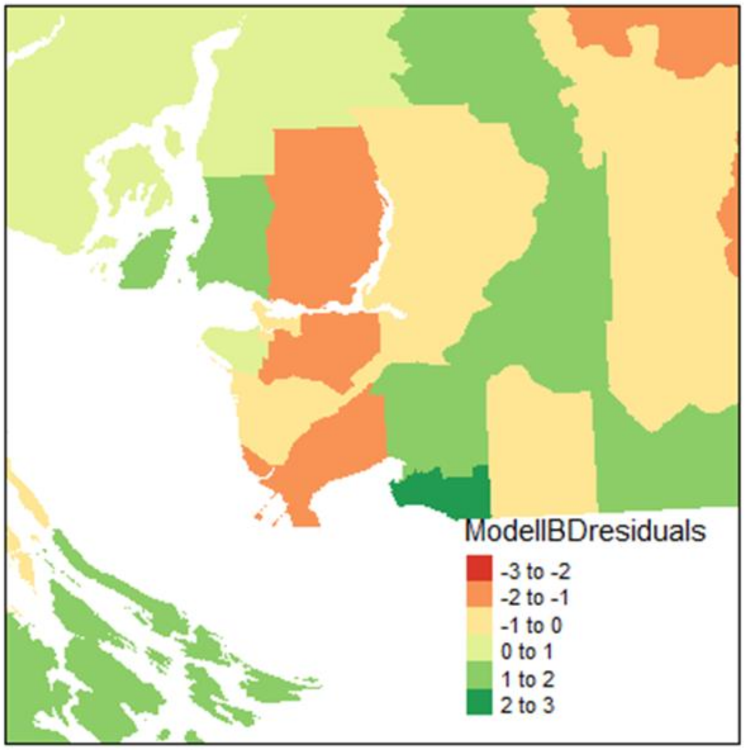
The diagnostics did not suggest a bias between residual magnitude and sign (Supplementary Figures 7A, 8A and 9A). We then explored a half-normal plot that compared the sorted absolute residuals and the quantiles of the half-normal distribution in order to detect outliers, which did not suggest the presence of outliers (Supplementary Figures 7B, 8B, and 9B). We then explored the influence of the observations. We also plotted the values using a half-normal plot (Supplementary Figures 7C, 8C and 9C). Finally, no anomalies were also observed when inspecting the estimated Cook's Distances using a half-normal plot for CD (Supplementary Figures 7D, 8D and 9D).

**Spatial independence diagnostics**

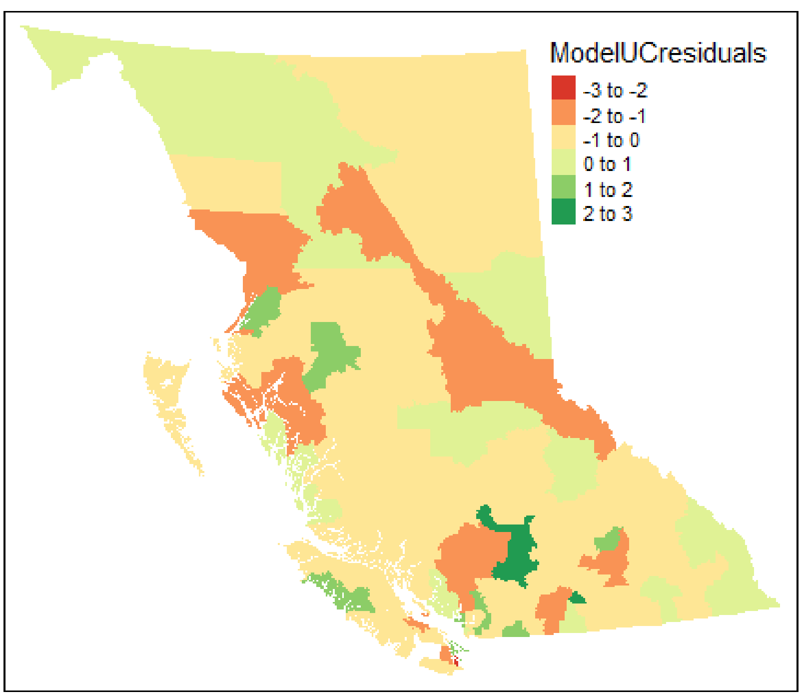
We examined our models' predictions to see if the residuals showed any forms of spatial dependence among them. To do so, we calculated Moran's I on the residuals for various notions of spatial proximity among the LHAs. As showed below, we found no strong evidence for unaccounted-for spatial dependency using any a number of (families of) adjacency metrics, which included Queen's adjacency, raising Euclidean distances among LHAs to inverse powers, raising driving distances to inverse powers, raising driving times to inverse powers, and adjacencies of population gravity models using driving times as distances. In other words, we confirmed the assumption of spatial independence that ensures valid inferences using Poisson rate models.

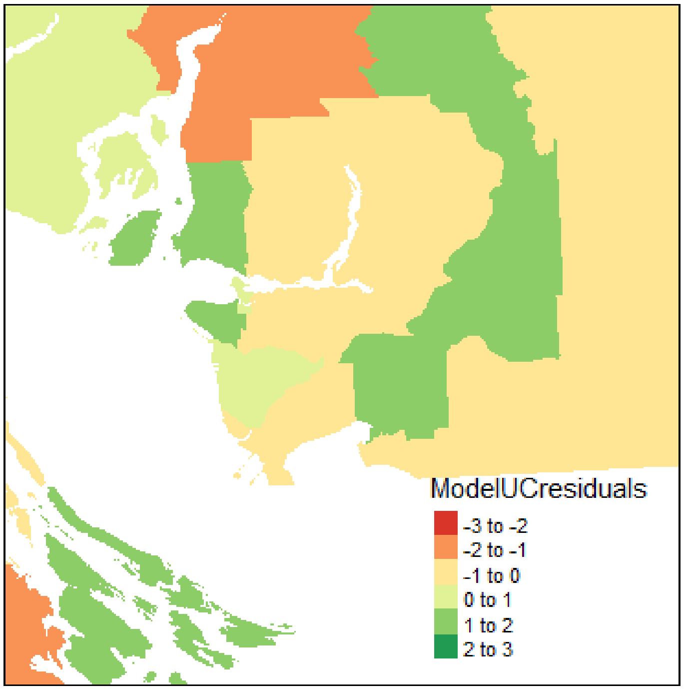
Maps of Model Residuals



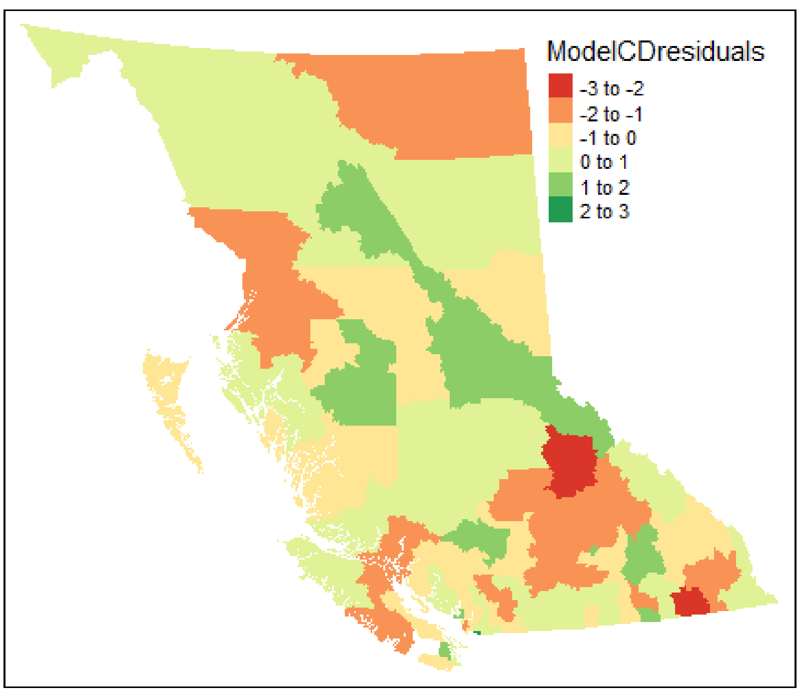


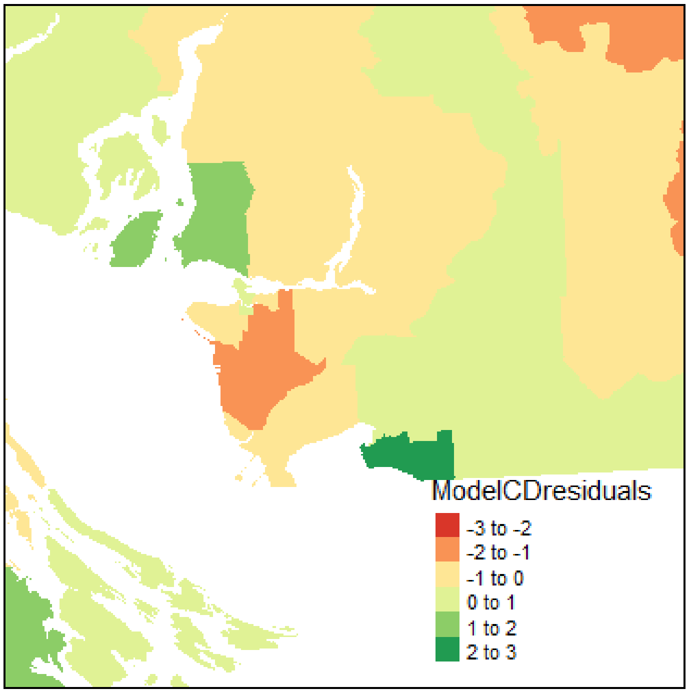

**Supplementary Figure 10 Inflammatory Bowel Disease (all pathologies) best model residuals.**






**Supplementary Figure 11 Ulcerative Colitis best model residuals.**






**Supplementary Figure 12 Crohn's Disease best model residuals.**

*Queen's adjacency*

We first tested for spatial dependence in the residuals using Queen's adjacency.

*Inverse powers of Euclidean distances adjacency*

We then tested for spatial dependence in the residuals using a range of inverse powers of the Euclidean distances between Local Health Areas (choices of points between which to measure distances are explained in the Supplement above under section on Data Preparation). We tested various possible adjacency matrices with exponents ranging from -0.5 to -3.0, stepping by 0.5.

*Inverse powers of driving distances adjacency*

Next, we tested for spatial dependence in the residuals using a range of inverse powers of the driving distances between Local Health Areas (both choices of points between which to measure distances as well as how driving distances are calculated are explained in the Supplement above under section on Data Preparation). We tested many possible adjacency matrices with exponents ranging from 0 to -3.0, stepping by 0.5.

*Inverse powers of driving time adjacency*

We also tested for spatial dependence in the residuals using a range of inverse powers of the driving times between Local Health Areas (both choices of points between which to measure distances as well as how driving times are calculated are explained in the Supplement above under section on Data Preparation). We tested many possible adjacency matrices with exponents ranging from -0.5 to -3.0, stepping by 0.5.

*Gravity models with distances of driving time adjacency*

Finally, we tested for spatial dependence in the residuals using a range of possible gravity models, which are commonly used to model spatial interactions. In our models, interaction between region $i$ and $j$ is proportional to the products of their total populations in 2016, $P_i$ and $P_j$, divided by the characteristic driving time $t_{ij}$ between them raised to a power $\alpha$, or in short: $(P_i P_j)/t_{ij}^{\alpha}$. As before, both our choices of points between which to measure distances as well as how we calculated driving times are explained in the Supplement above under section on Data Preparation). We tested many possible adjacency matrices with gravity model exponents $\alpha$ ranging from -0.5 to -5.0, stepping by 0.5.

**Supplementary Table 6 Results of the Moran's I Index tests for spatial independence of residuals assuming different spatial relations. In the leftmost column, Dis indicates disease, IBD indicates Inflammatory Bowel Disease, UC indicates Ulcerative Colitis, and, CD Crohn's Disease. Inferences are based on 1000 Monte-Carlo simulations. All results are based on the smallest exponent tested, as larger exponents presented had the same lack of statistical significance**

| Dis | Assumption | Statistic | Rank | P-value |
|---|---|---|---|---|
| IBD | Queen's adjacency | *0.031* | 743 | 0.257 |
| IBD | Inverse powers of Euclidean distances adjacency | *-0.016* | 322 | 0.678 |
| IBD | Inverse powers of driving distances adjacency | *-0.027* | 157 | 0.843 |
| IBD | Inverse powers of driving time adjacency | *-0.017* | 256 | 0.744 |
| IBD | Gravity models with distances of driving time adjacency | *-0.012* | 569 | 0.431 |
| UC | Queen's adjacency | *-0.145* | 16 | 0.984 |
| UC | Inverse powers of Euclidean distances adjacency | *-0.026* | 23 | 0.997 |
| UC | Inverse powers of driving distances adjacency | *-0.032* | 109 | *0.891* |
| UC | Inverse powers of driving time adjacency | *-0.019* | 157 | 0.843 |
| UC | Gravity models with distances of driving time adjacency | *-0.011* | 535 | 0.465 |
| CD | Queen's adjacency | *0.085* | 917 | 0.083 |
| CD | Inverse powers of Euclidean distances adjacency | -0.015 | 444 | 0.556 |
| CD | Inverse powers of driving distances adjacency | -0.026 | 198 | 0.802 |
| CD | Inverse powers of driving time adjacency | -0.018 | 202 | 0.798 |

| | | | | |
|---|---|---|---|---|
| CD | Gravity models with distances of driving time adjacency | -0.008 | 740 | 0.260 |